An Evolutionary Genome Scan for Longevity-Related Natural Selection in Mammals

Richard W. Jobson,*1 Benoit Nabholz,†1 and Nicolas Galtier1

1Institut des Sciences de l’Evolution, CNRS UMR 5554, Université Montpellier 2, Montpellier, France
†Present address: Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden
*Corresponding author: E-mail: rjobson@univ-montp2.fr.
Associate editor: John H. McDonald

Abstract

Aging is thought to occur through the accumulation of biochemical damage affecting DNA, proteins, and lipids. The major source of cellular damage involves the generation of reactive oxygen species produced during mitochondrial respiratory activity of the electron transport chain. Energetic metabolism, antioxidative processes, genome maintenance, and cell cycle are the cellular functions most commonly associated with aging, from experimental studies of model organisms. The significance of these experiments with respect to longevity-related selective constraints in nature remains unclear. Here we took a phylogenomic approach to identify the genetic targets of natural selection for elongated life span in mammals. By comparing the nonsynonymous and synonymous evolution of 5.7 million codon sites across 25 species, we identify codons and genes showing a stronger level of amino acid conservation in long-lived than in short-lived lineages. We show that genes involved in lipid composition and (collagen associated) vitamin C binding have collectively undergone increased selective pressure in long-lived species, whereas genes involved in DNA replication/repair or antioxidation have not. Most of the candidate genes experimentally associated with aging (e.g., FoxO, Sod, Fas) have played no detectable role in the evolution of longevity in mammals. A large body of current medical research aims at discovering how to increase longevity in human. In this study, we uncovered the way natural selection has completed this task during mammalian evolution. Cellular membrane and extracellular collagen composition, not genome integrity, have apparently been the optimized features.

Key words: phylogenomics, longevity, aging, mammals, natural selection.

Introduction

The life span of model species has been experimentally related to many physiological and cellular processes (Kujoth et al. 2005, 2007; Kuningas et al. 2008; Trifunovic and Larsson 2008). To attain long life, it is thought that organisms are required to maintain DNA replication fidelity (Kujoth et al. 2005, 2007) and limit oxidative damage via antioxidant and repair enzymes (Hsieh and Yamane 2008) or vitamins (Hulbert et al. 2007), particularly in the reactive oxygen species (ROS)—rich mitochondrial environment (Kujoth et al. 2005, 2007). Sugar-based energetic metabolism is controlled by the insulin/insulin growth factor signaling pathway, which was associated with aging in several model organisms (Kuningas et al. 2008). Mitochondria, furthermore, play a central role in apoptosis, and mitochondrial dysfunction due to oxidative damage could therefore deregulate the cell cycle and contribute to aging either through tumorigenesis (Lee and Wei 2007) or through the premature initiation of cell death (Kujoth et al. 2005). The lipid composition of cellular membranes, finally, has been connected to longevity and aging (Hulbert 2008; Puca et al. 2008). Membrane fatty acid saturation decreases sensitivity to peroxidation and reduces oxidative damage caused elsewhere by the byproducts of lipoxidation (Hulbert 2008; Pamplona 2008).

It is unclear whether experimental manipulation of these processes in model organisms reflects the selective forces really applying to aging and longevity in nature (Gems and Doonan 2009). In particular, we do not know if some of these physiological and cellular functions, and which ones, have been involved in the control of life span variations during the evolution of species. In an attempt to address these problems, we took a phylogenomic approach to identify the functions and genes that have been targets of natural selection for elongated life span in mammals.

The functional importance of a protein-coding gene is reflected by the level of amino acid conservation through evolutionary time, classically measured by dN/dS (Yang 1994). In this study, we reasoned that genes preventing premature aging should be under stronger selective pressure in long-lived species, relative to short-lived ones. We made use of fully sequenced genomes to scan the mammalian proteome in search of codons, genes, and functional categories of genes showing a lower dN/dS ratio in long-lived than in short-lived lineages—a signature of longevity-related selection.

Materials and Methods

Alignments

We retrieved 11,370 alignments of mammalian orthologous protein-coding sequences from the OrthoMaM version 4 database (Ranwez et al. 2007), which is built on EnsEMBL (Birney et al. 2004) version 49 (March 2008).
Alignments including at least 15 distinct species were retained. Sequences including 50% or more of gaps and undetermined characters, or less than 50 codons, were discarded. Alignments showing extreme levels of amino acid variability were manually assessed and discarded if nonorthologous sequences were detected. Our main results utilized 10,762 alignments that were finally analyzed (available on request). Only sites (codons) determined in 15 species or more were considered for evolutionary analysis. For comparison, we also analyzed a data set of genes with limited thresholds of 75% gaps and undetermined characters, and sites determined in at least 20 species. Application of these thresholds provided 6,917 alignments for analysis and had little effect on levels of significance.

**Longevity Data and Tree Annotation**

The maximal longevity of the 25 mammals shown in figure 1 was obtained from the AnAge database (de Magalhães et al. 2009). Much of the age-related data provided in this database were obtained from records of animals in captivity, and the question of whether these ages are representative of those attained in nature was examined by Ricklefs and Scheuerlein (2001), who showed that mortality rate and rate of aging are well correlated between captive and wild population (Speakman 2005). For each species of our data set, we calculated the average maximal longevity of congeneric species available in AnAge. This genus average was taken to decrease the effect of very recent changes in maximal longevity, especially in domesticated animals. Terminal branches of the placental subtree were labeled as “short lived” (species whose maximal longevity is less than 12 years), “medium,” or “long lived” (species whose maximal longevity is more than 26 years). This annotation was propagated to internal branches such that all the branches within a clade made of only short-lived (or medium, or long lived) species are considered short lived (or medium, or long lived, respectively; see fig. 1). We also trialed an analysis using two groups, with the split between short- and long-lived lineages set at 24 years. In addition, we also performed trials that lumped short- with medium-lived species, and medium- with long-lived species. For these additional trials, we obtained similar results across Gene Ontology (GO) categories (data not shown).

**GO Categories**

The GO database (Carbon et al. 2009; http://amigo.geneontology.org/cgi-bin/amigo/go.cgi, release: 3 April 2009) was used to build groups of genes sharing common biological functions or processes, built from a list of GO terms most frequently mentioned in the aging literature (Speakman et al. 2004; Kijima et al. 2005; Lee and Wei 2007; Hsieh and Yamane 2008; Hulbert 2008). Among these, GO terms associated with at least 10 genes of our mammalian data set were retained as a functional category, to be assessed for longevity-dependent selection. Some of the retained terms are hierarchically nested in GO, child terms (e.g., mitochondrial DNA [mtDNA] replication) being specialized instances of parent terms (e.g., DNA replication). Corresponding groups of genes are obviously not independent in this case. We assessed the overlap between groups of genes associated with terms from the same hierarchical level (e.g., in the first column of table 1—see supplementary table S1, Supplementary Material online, for GO terms) and found it to be limited: below 5% for the majority of pairs of terms, and below 30% for all but one. Terms (GO:0006281 and GO:0006974) shared strong overlap (69%), as indicated in table 1 and supplementary table S1 (Supplementary Material online). Distinct groups of equal hierarchical level can therefore be considered as roughly independent, with one exception.

![Fig. 1. Substitution mapping on the mammalian phylogenetic tree.](https://academic.oup.com/mbe/article-abstract/27/4/840/1748306/1748306)
Substitution Mapping

For each alignment, substitutions were mapped onto the mammalian tree using the probabilistic method introduced by Dutheil et al. (2005). This method is based on the maximum likelihood (ML)–based estimate of the number of substitutions for each site in each branch of the tree, similar to the probabilistic ancestral sequence reconstruction method (Yang et al. 1995). The method was applied to amino acid sequences, under the JTT + Gamma substitution model (Yang 1994), and then to nucleotide sequences, using the HKY + Gamma substitution model (Hasegawa et al. 1985; Jones et al. 1992) assuming Gamma-distributed rates across sites. The use of alternative models did not qualitatively change the results. For each branch and codon, the number of synonymous substitutions was inferred by subtracting the estimated number of amino acid substitutions from the estimated number of nucleotide substitutions.

For each gene g, the total (over all sites) number of synonymous substitutions having occurred in short-lived branches, $S_{SL_{tot}(g)}$, and the total number of synonymous substitutions having occurred in long-lived branches, $S_{LL_{tot}(g)}$, were calculated. Then, for each codon (say, codon c) of gene g, the number of nonsynonymous changes having occurred in short lived, $N_{SL}(gc)$, and long lived, $N_{LL}(gc)$, were computed. Codons for which the total estimated number of nonsynonymous changes in both short-lived and long-lived branches was below three were not considered further.

Significance Tests

Codons. For each codon c of gene g, depletion of nonsynonymous changes in long-lived branches, knowing the
synonymous substitution pattern, was assessed through a binomial test in which:

- the number of trials was \( n = N_{LL}(g,c) + N_{SL}(g,c) \), the total number of nonsynonymous changes in long-lived + short-lived branches
- the number of successes was \( k = N_{LL}(g,c) \), the number of nonsynonymous changes in long-lived branches
- the expected probability was \( p = \frac{S_{LL\_tot}(g)}{(S_{LL\_tot}(g) + S_{SL\_tot}(g))} \), the relative proportion of synonymous changes in long-lived branches.

The \( P \) value for amino acid conservation at codon \( c \) in long-lived branches was defined as \( P_{\text{val}_{LS}} = \Pr(X \leq k) \), when \( X \) follows a binomial distribution \( B(n,p) \). Codons for which \( P_{\text{val}_{LS}} \) was below 0.35 were called “longevity selected” (LS). Similarly, a \( P \) value for excess of nonsynonymous changes at codon \( c \) in long-lived branches was defined as \( P_{\text{val}_{LR}} = \Pr(X' \leq k') \), where \( k' \) is \( N_{SL}(g,c) \) and \( X' \) follows a binomial distribution \( B(n, 1 - p) \). Codons for which \( P_{\text{val}_{LR}} \) was below 0.35 were called “longevity relaxed” (LR). Rounded values of \( n \) and \( k \) were used (probabilistic substitution mapping returns noninteger estimates).

For comparison, we also analyzed the data using \( P \)-val thresholds of 0.25 and 0.49. Application of these thresholds tended to decrease or increase the numbers of LR and LS sites equally, and therefore had little effect on overall estimates for both site-specific and gene-specific (see below) analyses.

For the main analysis, codon-specific nonsynonymous branch lengths were purposely compared with total synonymous branch lengths, not codon-specific synonymous branch lengths, in order to decrease the stochastic effect of codon-specific synonymous evolution. This was empirically found more stable, for this data set, than methods comparing codon-specific synonymous and nonsynonymous rates, such as the FITMODEL software (Guindon et al. 2004). Our method is similar in spirit to the “branch” model of PAML (Yang 2007), with the additional advantage of identifying target-selected codons and requiring shorter running times. Note that we do not need to account for the gene-specific proportion of synonymous versus nonsynonymous sites: It cancels out when taking the ratio of long-lived and short-lived lineages. Codon-specific \( P \) values, finally, do not account for phylogenetic nonindependence of the observations. This, we think, is not a major problem because these \( P \) values are not interpreted in terms of statistical significance of the effect. Rather, they are used to identify, and count, the most extreme codons with respect to longevity-related selection (see below).

**Genes.** To provide a gene-specific measure of conservation in short-lived versus long-lived species, we counted, for each gene \( g \), the number \( n_{LS}(g) \) of LS codons (\( P_{\text{val}_{LS}} < 0.35 \)) and the number \( n_{LR}(g) \) of LR codons (\( P_{\text{val}_{LR}} < 0.35 \)). Note that these are only small fractions of total codons because only codons for which the total number of nonsynonymous changes in short-lived and long-lived branches was above three were considered (see above). The strength of longevity-dependent selection for gene \( g \) was measured by \( H(g) = \frac{[n_{LS}(g) - n_{LR}(g)]/l(g)}{l(g)} \), where \( l(g) \) is the number of analyzed codons for gene \( g \). A gene, therefore, was said to be long-lived when it included a large proportion of LS codons, as compared with LR codons.

The \( H(g) \) statistic, however, is mechanically dependent on the nonsynonymous variability of gene \( g \) across species. A highly conserved gene can only yield low \( n_{LS} \) and \( n_{LR} \) and therefore a low \( n_{LS} - n_{LR} \) difference. We assessed the significance of \( H \) for a candidate gene \( g \) by comparing with random genes, accounting for the nuisance effect of gene variability.

The variability of any gene \( x \) was measured by the per-site–estimated number of nonsynonymous changes, \( N_{\text{tot}}(x) \) (all codons, all branches). Genes \( x \) such that the absolute value of the difference between \( N_{\text{tot}}(x) \) and \( N_{\text{tot}}(g) \) was below 0.5, were said to be comparable to \( g \). Assume that there are \( m \) such genes in the data set (including \( g \)). The \( m \) genes were ranked in decreasing order of \( H \), and the \( P \) value for gene \( g \) was defined as \( \text{rank}(g)/m \) (see Dutheil and Galtier 2007). Genes for which \( m \) was lower than 100 (extreme \( N_{\text{tot}} \) ) were not considered. A candidate gene, therefore, is said to be under longevity-dependent selection when it shows an exceptional proportion of LS codons (relative to LR codons) when compared with genes of similar nonsynonymous variability.

**Groups of Genes.** Two approaches were used to assess whether a group \( G \) of genes (typically, genes sharing a given GO annotation) were collectively under longevity-dependent selective pressure. The first one involved counting the total number of LS, \( n_{LS}(G) \), and LR, \( n_{LR}(G) \), codons for all genes in the group. Then we asked whether group \( G \) was enriched in LS sites, as compared with random groups of genes. We calculated the \( H \) statistic for the group, defined as \( H(G) = \frac{[n_{LS}(G) + n_{LR}(G)]/l(G)}{l(G)} \), where \( l(G) \) is the total number of analyzed codons for genes in \( G \). We generated the null distribution of this statistic by randomly sampling 10,000 equally sized groups from the full list of genes. The \( P \) value for group \( G \) was defined as the proportion of random groups \( X \), such that \( H(X) \geq H(G) \).

The above method assigns equal weights to codons, so that long genes matter more than short genes. Alternatively, we performed a test giving equal weight to genes. We counted the number of genes within the group showing significant LS signal at the 10% level. By definition, the expectation of this proportion is 10% under the null hypothesis of no longevity-dependent selection. We tested for enrichment in significant genes through a binomial test (number of trials: number of genes in the group; number of success: number of significant genes in the group; probability: 10%). The two approaches yielded qualitatively similar results (table 1).

**Controls**

**Tree Topology.** We reconducted the analysis using alternative topologies, in which branching orders within Laurasiatheria, or between the four superorders, were modified—other nodes in the tree shown in figure 1 are
well-supported. We got nearly identical results (not shown). Here are the additional topologies we tried:

1. (Monodelphis, ((((((Mus, Rattus), Cavia), Spermophilus), (Ochotona, Oryctolagus), (((Otolemur, Microcebus), (Macaca, (Pongo, (Pan, Homo)))), Tupaia)), (Sorex, Erinaceus), (((Felis, Canis), Equus), (Myotis, Bos))), (Loxodonta, Echinops), Dasypus)), Ornithorhynchus);
2. (Monodelphis, ((Loxodonta, Echinops), ((((((Mus, Rattus), Cavia), Spermophilus), (Ochotona, Oryctolagus), (((Otolemur, Microcebus), (Macaca, (Pongo, (Pan, Homo)))), Tupaia)), (Sorex, Erinaceus), (((Felis, Canis), Equus), (Myotis, Bos))), (Loxodonta, Echinops), Dasypus)), Ornithorhynchus);
3. (Monodelphis, ((Loxodonta, Echinops), ((((((Mus, Rattus), Cavia), Spermophilus), (Ochotona, Oryctolagus), (((Otolemur, Microcebus), (Macaca, (Pongo, (Pan, Homo)))), Tupaia)), (Sorex, Erinaceus), (((Felis, Canis), Equus), (Myotis, Bos))), (Ochotona, Oryctolagus)), Dasypus)), Ornithorhynchus).

Tree Annotation. We replicated the analysis without propagating the annotations short, medium, and long lived to internal branches, so that synonymous versus nonsynonymous evolution in external branches only was analyzed. This did not qualitatively affect the results (see supplementary table S1, Supplementary Material online).

Weighted Mapping. We applied an alternative substitution mapping procedure in which nonsynonymous changes are weighted according to Grantham’s (1974) biochemical distance between the initial and final amino acid states (Duthel and Galtier 2007), thus conferring higher weights to radical changes and lower weights to conservative changes. Codon-specific counts of synonymous and amino acid changes and the subsequent statistical analyses were applied separately to the unweighted and Grantham mapping (see supplementary table S2, Supplementary Material online).

Taxonomic Sampling. For the candidate LS genes listed in Table 2, 5–16 additional eutherian species available from public databases were manually added to the alignment and the analyses were reconducted. With the addition of the new taxa, we found a similar ratio of LS to LR sites (see supplementary table S3, Supplementary Material online).

Codon-Based Models. To confirm our substitution mapping–based analyses, we applied classical codon models to the 18 genes listed in Table 2, with additional taxa (see above), using the PAML software (Yang 2007). We used a model involving four distinct dN/dS (ω) ratios across branches: one for short-lived lineages (ω0), one for medium-lived lineages (ωM), one for long-lived lineages (ωL), and one for unlabeled lineages (ωU), using the same procedure as above to label branches of the tree. Estimated ω values are given in supplementary table S3 (Supplementary Material online). These results are generally consistent with our main analysis.

Results and Discussion

We analyzed 10,762 nuclear protein-coding genes available in at least 15 of the 25 mammalian species shown in figure 1. For each codon in each alignment, we mapped nonsynonymous and synonymous changes onto the (supposedly known) mammalian phylogenetic tree in a ML framework (Duthel and Galtier 2007). From the maximal recorded life span of extant species (de Magalhães et al. 2009), we classified the various lineages of the tree as short lived, long lived, or intermediate/unknown. We then estimated, for each codon, the total number of nonsynonymous and synonymous substitutions having occurred in the long-lived versus the short-lived branches. As a comparison, nonsynonymous change counts were either unweighted or weighted according to the biochemical difference between the ancestral and derived amino acid states (see Materials and Methods).

The literature suggests that the general trend in mammals is a higher codon-specific dN/dS ratio in large species (Nikolaev et al. 2007; Popadin et al. 2007), a pattern interpreted as reflecting the smaller average population sizes, and hence less effective purifying selection (Wright 1931), of large animals. We sought codons showing the opposite patterns, that is, a stronger level of amino acid conservation in long-lived lineages. Such sites were called LS, whereas sites showing the majority trend (more nonsynonymous variability in long lived) were called LR. The evolutionary involvement of a gene in longevity-related processes was measured by the relative difference between numbers of LS and LR sites. Consistent with previous analyses (Nikolaev et al. 2007), we found a strong excess of LR over LS sites when analyzing the whole 10,762 genes (table 1, last row), confirming that protein-coding genes, on average, are under weaker effective selective constraint in long-lived mammals.

Target Cellular Functions

To assess the influence of the various processes involved in mammalian life span evolution, we constructed a list of mammalian genes involved in DNA replication and repair, antioxidation, insulin signaling, apoptosis, vitamin binding, and cellular lipid metabolism, transport, and binding, categorized using the GO database (Carbon et al. 2009). For each candidate functional category, we calculated the numbers of LS and LR sites and measured enrichment in LS sites by comparing with random genes (see Materials and Methods). Significance was assessed using two distinct randomization procedures, one giving equal weight to sites and one giving equal weight to genes (see Materials and Methods). The analysis was essentially robust to a battery of control analyses we performed: usage of alternative phylogenetic trees, exclusion of internal branches of the tree in substitution count (supplementary table S1, Supplementary Material online), and biochemical weighting of amino acid changes (supplementary table S2, Supplementary Material online), did not significantly affect the results.

We found that genes involved in cellular lipid metabolism, especially that of fatty acid biosynthesis, are significantly enriched in LS sites (table 1). This is consistent with recent reports highlighting the importance of lipid composition in longevity and aging: Long-lived mammals have more saturated or monounsaturated lipids than...
short-lived species (Hulbert 2008). The peroxidation of lipids by ROS produces reactive carbonyl species, which, being more stable than ROS, can diffuse and oxidize other cellular or extracellular components (Pamplona 2008). Fatty acids are also known to stimulate uncoupling proteins, thereby reducing ROS generation, particularly in calorie-restricted organisms (Kua 2006). Such a link implicates fatty acid biosynthesis, not only in the reduction of peroxidative damage (via increased saturation) but also in the control of mitochondrial ROS production through reduced membrane potential and efficient uncoupling of the electron transport chain (Kua 2006).

The other strongly significant function was the category of genes involved in the binding of ascorbic acid (vitamin C). Vitamins are a disparate group of compounds necessary for normal metabolic function and have sometimes been linked to longevity in mammals (Hulbert et al. 2007). Categories of genes involved in the metabolism of vitamins exhibit biosynthetic and lineage-specific variation and loss and were therefore not examined. Instead, we investigated genes that interact selectively with vitamins through functional binding processes and found a significant excess of LS sites in ascorbic acid–binding genes (table 1). Of the ascorbic acid–binding enzymes that selectively utilize ascorbic acid for biosynthesis and stabilization of collagen’s triple helix structure, a process that is essential for the integrity of connective tissues (Sell et al. 1996). This would appear again in line with Pamplona’s (2008) conclusion that resistance of cellular and extracellular components to oxidative damage is a key trait of longevity.

To a lesser extent, genes of the insulin/insulin growth factor signaling and insulin receptor signaling pathways, and negative regulation of apoptosis also appeared more constrained in long-lived mammals (table 1). Genes controlling ROS production and management and genes ensuring genome integrity, surprisingly, were not enriched in LS sites and even tended to behave the opposite way: Categories “DNA repair,” “response to DNA damage,” and “antioxidant activity” were significantly enriched in LR sites, that is, more constrained during the evolution of short-lived mammals (table 1).

### Aging Genes

Then we examined the numbers and significance of LS and LR sites for individual genes. Many genes classically associated with aging did not show any enrichment in LS sites, or even showed the reverse pattern—increased amino acid variability in long-lived species (table 2, bottom). This list includes PolG, which encodes the mtDNA polymerase, and its companion gene PolG2. PolG was recently highlighted as a candidate aging gene when two independent groups reported accelerated senescence in a transgenic mouse carrying a defective, low-fidelity version of the enzyme (Trifunovic et al. 2004; Kujoth et al. 2005). However, the relevance of this transgenic mouse to life in natural conditions has been questioned (Vermulst et al. 2007). Our analysis suggests that the evolution of this gene was not affected by changes in species longevity. This result was unchanged when we analyzed the various domains of the protein separately—none did respond.

This statement also applies to superoxide dismutases, catalase, and coupling/uncoupling proteins, which have long been associated with aging because of their ability to control ROS production and prevent oxidative damage (Tommaso et al. 1980; Sun et al. 2002; Speakman et al. 2004). As suggested by Hulbert et al. (2007), the enzymes directly responsible for antioxidant defense can be

### Table 2. List of Potential Longevity-Related Genes.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Process/Function</th>
<th>n Taxa</th>
<th>nLS</th>
<th>nLR</th>
<th>p</th>
<th>wSL</th>
<th>wML</th>
<th>wLL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elovl5</td>
<td>FAS—elongase</td>
<td>16</td>
<td>171</td>
<td>4</td>
<td>2</td>
<td>0.016</td>
<td>0.08</td>
<td>0.073</td>
</tr>
<tr>
<td>Scd5</td>
<td>FAS—desaturase</td>
<td>20</td>
<td>253</td>
<td>3</td>
<td>2</td>
<td>0.063</td>
<td>0.162</td>
<td>0.121</td>
</tr>
<tr>
<td>Oxsm</td>
<td>FAS—synthase</td>
<td>20</td>
<td>459</td>
<td>13</td>
<td>11</td>
<td>0.001</td>
<td>0.126</td>
<td>0.14</td>
</tr>
<tr>
<td>MerC</td>
<td>FAS—malonyltransferase</td>
<td>22</td>
<td>373</td>
<td>9</td>
<td>5</td>
<td>0.049</td>
<td>0.176</td>
<td>0.17</td>
</tr>
<tr>
<td>Ptgese</td>
<td>FAS—prostaglandin</td>
<td>20</td>
<td>149</td>
<td>3</td>
<td>1</td>
<td>0.027</td>
<td>0.117</td>
<td>0.113</td>
</tr>
<tr>
<td>Ogfod2</td>
<td>AAB—dioxygenase</td>
<td>23</td>
<td>350</td>
<td>10</td>
<td>3</td>
<td>0.017</td>
<td>0.124</td>
<td>0.112</td>
</tr>
<tr>
<td>Alkbh3</td>
<td>AAB—lipoxidation repair</td>
<td>22</td>
<td>286</td>
<td>12</td>
<td>5</td>
<td>0.005</td>
<td>0.182</td>
<td>0.209</td>
</tr>
</tbody>
</table>

**Mean** | 7.7              | 2.7    | 0.14 | 0.16 |
|----------|------------------|--------|-----|-----|

**Sod1** | ANT/RDD/ROS      | 21     | 154 | 0   | 9   | 0.251 | 0.188 | 0.307 | 0.557 |
| Sod2    | ANT/ROS          | 18     | 197 | 1   | 4   | 0.818 | 0.083 | 0.117 | 0.187 |
| Cat     | ANT/ROS/APO      | 23     | 527 | 1   | 17  | 0.997 | 0.12  | 0.118 | 0.169 |
| Peo1    | DNA replication  | 23     | 684 | 1   | 17  | 0.913 | 0.113 | 0.146 | 0.207 |
| Tfam    | DNA replication  | 21     | 244 | 0   | 11  | 0.84  | 0.293 | 0.514 | 0.432 |
| Ant1    | DNA replication  | 21     | 296 | 0   | 4   | 0.76  | 0.048 | 0.061 | 0.066 |
| PolG    | DNA repair/replication | 20     | 1,164 | 19 | 6.7  | 0.131 | 0.1  | 0.128 |
| PolG2   | DNA repair/replication | 23     | 483 | 0   | 11  | 0.642 | 0.235 | 0.33  | 0.295 |
| Ucp1    | ROS—uncouple ETC | 22     | 307 | 0   | 10  | 0.854 | 0.122 | 0.199 | 0.294 |
| Ucp2    | ROS—uncouple ETC | 24     | 309 | 0   | 4   | 0.914 | 0.06  | 0.06  | 0.11 |
| Ucp3    | ROS—uncouple ETC | 24     | 311 | 0   | 9   | 0.959 | 0.075 | 0.149 | 0.175 |

**Mean** | 0.36 | 10.5 | 0.13 | 0.24 |

**Note.**—Top section: Example genes revealed as LS by our analysis. Bottom section: Genes classically associated with aging, not showing any excess LS sites. Nonsynonymous/synonymous substitution ratio (v) values estimated from ML analyses using PAML software (Yang 2007) are also given for short-lived lines (vSL), medium-lived lines (vML), and long-lived lines (vLL). ENSMBL alignment ID and GO reference numbers are shown in supplementary table S3 (Supplementary Material online). Respective P values shown for the ratio of nL and nS with significance <5% in green, <10% in blue, and >95% in red. AAB, ascorbic acid binding; ANT, antioxidant; APO, apoptosis; ETC, electron transport chain; FAS, fatty acid synthesis; RDD, response to DNA damage.
upregulated in line with ROS production and damage (but see Pérez et al. 2009), and perhaps this is the reason their primary sequences are not targets for strong selection in long-lived mammals (Gems and Doonan 2009).

From an evolutionary viewpoint, our inability to detect longevity-dependent selection on mtDNA replication and repair genes, including PolG, is a bit of a paradox: Several lines of evidence suggest that the mtDNA mutation rate is under selective pressure in long-lived mammals (Samuels 2004; Nabholz et al. 2008; Welch et al. 2008), in agreement with the mitochondrial theory of aging (Galtier et al. 2009). There are, however, good evolutionary “reasons” to optimize the fidelity of DNA replication and repair irrespective of longevity. This is because deleterious mutations in DNA are transmitted to the next generation. Degradation of lipids and collagen, in contrast, only affects the current generation and is plausibly under weaker counterselection in short-lived organisms, for which long-term preservation of a functional soma is not a major component of fitness.

Table 2 (top) provides a list of example genes showing a significant excess of LS sites, which we propose as candidate genes for future functional research in human aging. The addition of extra taxa did not change the relative level of LS enrichment for these genes (supplementary table S3, Supplementary Material online), and usage of the classical ML analysis approach (Yang 2007) yielded results essentially congruent with our substitution mapping–based analysis.

This list includes the mitochondrial fatty acid synthase (FAS II) pathway genes MeCr and Oxsm. The mitochondrial FAS II pathway is primarily involved in the endogenous synthesis of the antioxidant compound lipoic acid, along with the reconstruction of membrane cardiolipins (Chen et al. 2008). MeCr encodes the dimeric reduced form of nicotinamide adenine dinucleotide phosphate–binding 2-enoyl thioester reductase enzyme that catalyzes the final step of the fatty acid elongation process (Chen et al. 2008). Oxsm encodes the dimeric mitochondrial β-ketoacyl synthase moiety that provides octanoyl acyl-carrier protein for the synthesis of short- and medium-chain fatty acids and may also play a role in the biosynthesis of long-chain fatty acids (Zhang et al. 2005). These genes could therefore be directly involved in the specific lipid composition observed in long-lived mammals (Hulbert et al. 2007). This could also be the case of the delta-9-desaturase (Sdc5) gene, which converts saturated fatty acids to monounsaturated fatty acids, and the ElOv5 gene, which encodes essential proteins for the synthesis of very long chain fatty acids during ceramide and sphingolipid formation (Wang et al. 2005). In addition, ElOv5 may also control the synthesis of long-chain polyunsaturated fatty acids (Wang et al. 2005).

Finally, we illustrate the potential of our substitution mapping–based approach by locating the detected LS sites on the crystal structure of the α-ketoglutarate–dependent dioxygenase 3 gene (Alkhh3), which encodes an ascorbic acid–activated dioxygenase that repairs alkylated DNA (Sundheim et al. 2006; supplementary fig. 1A, Supplementary Material online). The Pro221-Pro222 (Pro223 not present in structure) motif may play a role in DNA damage sensing (Sundheim et al. 2006), and/or possibly ascorbic acid binding. Pro222 is selectively relaxed to Leu/Ser in short-lived species, which could alter the integrity of the motif (supplementary fig. 1B, Supplementary Material online), if in fact it is functionally constrained.

In short-lived lineages, Glm128 is substituted with Leu/Pro (supplementary fig. 1A, Supplementary Material online). This residue is positioned on the β5 strand and forms polar contact with Thr119 on β4, forming a groove containing Pro130 and Tyr127 (supplementary fig. 1C, Supplementary Material online). These two residues are in close contact with the 3′-phosphate of 1-methyladenine–damaged DNA (1-meA), and although Gln128 does not make direct contact with 1-meA, selective relaxation in short-lived species may interrupt the Pro130–Tyr127 interaction (supplementary fig. 1C, Supplementary material online), thereby altering contact with DNA via water (Sundheim et al. 2006).

Conclusions

A large body of current medical research aims at discovering how to increase longevity in human. In this study, we uncovered the way natural selection has completed this task during mammalian evolution. Cellular membrane and extracellular collagen composition, not genome integrity, have apparently been the optimized features. We identified specific genes and amino acids showing evidence for longevity-dependent selection, of which the classical PolG, Tfam, Sod, Cat, and Ucp’s, and genes with antioxidant activity as a whole, are consistently excluded. The functional relevance of the detected LS genes with respect to longevity is, however, not demonstrated. Life span in mammals is strongly correlated to a number of life history traits, including body mass, metabolic and development rate, and fecundity, for instance. Some of the LS genes could actually show the observed pattern for reasons not directly connected to longevity. Experimental confirmation is obviously required before drawing definitive conclusions.

Supplementary Material

Supplementary tables S1–S3, supplementary figure 1, and supplementary data file 1 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the Centre National de la Recherche Scientifique and by Agence Nationale pour la Recherche project MITOSYS. We thank Pierre Blier, Pierre-Henri Fabre, Sylvain Glémin, Joao Lourenco, and Vincent Ranwez for helpful contributions. This is contribution number ISEM 2009-143.

References

access to ontology and annotation data. Bioinformatics 25: 288–289.


847

Downloaded from https://academic.oup.com/mbe/article-abstract/27/4/840/1748306 by guest on 17 December 2018