Experimental Gerontology xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Experimental Gerontology

journal homepage: www.elsevier.com/locate/expgero

Identification of amino acids in mitochondrially encoded proteins that correlate with lifespan

Mahendra Mariadassou [a], François-Xavier Pellay [b,c,*]

[a] INRA, UR 1077 Mathématique, Informatique et Génome, 78350 Jouy-en-Josas, France
[b] INSERM Unit 1001, Université Paris-Descartes, Sorbonne Paris Cité, Faculté de Médecine Paris Descartes, 75014 Paris, France
[c] NAOS group/Jean-Noël Thorel, 13855 Aix-en-Provence, France

ARTICLE INFO

Article history:
Received 13 November 2013
Received in revised form 3 March 2014
Accepted 9 March 2014
Available online xxxx

Section Editor: Natascia Ventura

Keywords:
Lifespan
Longevity
Mitochondria
Evolution
mtDNA
Mitochondrially-encoded protein

ABSTRACT

Animals show a huge diversity in their lifespan that can vary from a few weeks to over a hundred years in vertebrates. Size is a key element in this variation and the positive correlation between size and maximum lifespan can be observed in each class of vertebrate. Some groups and species clearly stand out in this size–lifespan relationship and the ones with exceptionally long lifespan have been studied to understand the biological causes of their low aging rate. Among the potential explanations of animals' lifespan variations, mitochondria and mitochondrially encoded genes have drawn attention because of their importance in the aging process. To understand both the extent of lifespan variations and their dependence to genes and amino acid variations in mitochondrial genes and DNA (mtDNA), we analyze in a systematic way all 13 proteins encoded by mitochondria in all vertebrates for which we had information on weight, maximum lifespan and mtDNA sequence. This comparison allows us to visualize positions, and even specific amino acids, in these sequences that correlate with lifespan. With this approach, we draw a map of 356 amino acid residues, at 296 positions within the sequence, that correlate with longer or shorter lifespan. We also compared this map with the human mitochondrial polymorphism to determine its potential as a predictive tool.

1. Introduction

1.1. Mitochondria and its role in ATP and ROS production

Mitochondria are essential organelles of eukaryotic cells and have been linked to aging and longevity in numerous publications (Balaban et al., 2005; Harman, 1972; Loeb et al., 2005). They produce the energy of the cell by oxidative phosphorylation of ADP into ATP through the transfer of electron in the respiratory chain. An incomplete coupling in this process, mostly at the complex I of the electron transfer chain (Barja and Herrero, 1998), leads to electron leakage that can interact with oxygen to create superoxide radicals(Brand, 2000). Superoxide dismutase can then turn superoxide into hydrogen peroxide, which can lead to the production of hydroxyl radicals (Murphy, 2009). All these reactive oxygen species (ROS) are involved in the oxidation of all types of macromolecules in the cell and are believed to be one of the main causes of aging (Barja, 2013; Beckman and Ames, 1998; Cadenas and Davies, 2000; Harman, 1972). They also damage mitochondrial proteins and membrane, leading to an increase in ROS production (Remmen and Richardson, 2001). The implication of mitochondria in the regulation of apoptosis is also referred to play a key role in the aging process of animals (Green and Reed, 1998).

1.2. Mitochondria, lifespan and comparative biology

Comparative studies of various vertebrates show that the efficiency of the electron transfer chain (ETC) and the ROS production are tightly linked with lifespan (Pamplona and Costantini, 2011). Longer living species show a lower production of ROS, a lower activity of antioxidant enzymes (Perez-Campo et al., 1998), a lower degree of fatty acid unsaturation (Barja, 2002) and a lower amount of mtDNA oxidation (Barja and Herrero, 2000). Considering that all the proteins encoded by the mitochondrial genome are part of the ETC, the correlation of lifespan with mitochondrial genome has been measured in various aspects. The composition of mtDNA and mitochondrially encoded proteins (MEPs) show the most significant correlation with longevity, captured whether by the proportion of nucleotides in mtDNA (Jobson et al., 2010; Lehmann et al., 2008; Min and Hickey, 2008) or the proportion of some amino acids (Lehmann et al., 2006; Moosmann, 2011; Schindeldecker and Moosmann, 2013), especially methionine (Aledo et al., 2012) and cysteine (Moosmann and Behl, 2008). These findings show the strong role of MEPs in longevity but reduce mtDNA and

Abbreviations: mtDNA, Mitochondrial DNA; MEPs, Mitochondrially encoded proteins; ROS, Reactive oxygen species; ACL (+ or −), Amino acid correlated with (longer or shorter) lifespan.

* Corresponding author.
E-mail address: francois-xavier.pellay@inserm.fr (F.-X. Pellay).

http://dx.doi.org/10.1016/j.exger.2014.03.009
0531-5565/

MEPs to their composition in nucleotides and amino acids and are oblivious to the primary and secondary structure of these MEPs.

1.3. Role of mitochondrial proteins' sequence in lifespan

Because of the importance of these proteins in ATP and ROS production, the amino acid sequence of the proteins encoded by mtDNA and their variations must play a role in the determination of lifespan. Such a relationship was shown by Rottenberg to depend on the rate of evolution for placental mammals (Rottenberg, 2007).

In our study, we have studied the evolution of MEPs and its correlation with longevity by comparing their amino acid sequences in vertebrates. Our goal was to identify the key positions and amino acids within the sequence of MEPs that play a role in aging and longevity and to chart a map of these amino acids.

2. Material and methods

2.1. Selection of species

We used the AnAge database (De Magalhães and Costa, 2009) (*genomics.senescence.info/species/*), containing over 4000 species with information on both maximum lifespan (MLSP) and body mass (BM), and the PubMed database on mitochondrial genome sequences with over 2000 vertebrate mtDNA sequences (www.ncbi.nlm.nih.gov/genomes/OrganelleResource.cgi?opt=organelle&taxid=7742), in order to obtain 324 species with all three information. These species were tested to be significantly representative in terms of MLSP/BM distribution compared with the 4000 vertebrate species present in AnAge (data not shown).

2.2. Phylogenetic analyses

The 13 protein-coding genes (A6, A8, C1, C2, C3, Cb, N1, N2, N3, N4, N4L, N5, N6) of the mitochondrial genomes of the 324 species were aligned separately using the muscle (version 3.7, default parameters) (Edgar, 2004). Most genes were present in all species (N2 was missing from Oncorhynchus kisutch, N5 from *Sphenodon punctatus* and N6 from Castor canadensis). Alignments were both analyzed separately and concatenated in a supermatrix resulting in a total of 14 alignments. The supermatrix had 4005 columns, corresponding to as many amino acids. The best evolution model was selected for each alignment using Prottest v3.0 (Darriba et al., 2011). The mitochondrial replacement matrix MtMAM + I + G was strongly favored for all genes but N6 for all criteria (AIC, BIC and variations). The posterior probability was even close to 1.00 for 9 of the 13 (A6, C1, C2, C3, Cb, N1, N2, N4, N5). In each case, the second best model was MtMAM + G. For N6, the favored model was JTT + I + G, with only contender JTT + G (respective posterior probabilities: 0.85/0.15 for AIC, 0.53/0.47 for BIC). Fourteen phylogenetic trees were reconstructed using PhyML v3.0 (Guindon et al., 2010) with best of SPR and NNI moves, 5 random starting trees and 100 bootstrap replicates. We performed further analyses with the tree reconstructed from the concatenated alignment named *T* hereafter, as it minimized the sum of distances to the other trees.

2.3. Phylogenetic regression

To investigate the association between particular alignment sites and lifespan while correcting for body mass, we performed a phylogenetic regression of Lifespan against Body mass and amino acid distribution at that site. Namely, for each amino acid *aa* at each site *i*, we constructed a presence/absence vector of $X_{aa,i}$ of that amino acid at that site and estimated the linear model:

$$\log(\text{Lifespan}) \sim \log(\text{Body mass}) + X_{aa,i} + \varepsilon$$

where ε is the centered Gaussian with covariance matrix $\sigma^2\,\Sigma(T)$ where σ^2 is unknown and $\Sigma(T)$ is completely specified by *T*, as the covariance matrix of a Brownian motion running on *T* (Felsenstein, 1985). Lifespan and Body Mass were log-transformed to account for well-described allometric relations between the two.

This procedure resulted in $4005 * 20 = 80{,}100$ linear models and as many coefficients $a_{aa,i}$ to estimate. The coefficients were estimated using the compar.gee function of the R ape package (version 2.8) (Paradis et al., 2004). Visual inspection of the p-values did not show signs of significant correlations between the tests and we therefore used FDR correction (Benjamini and Hochberg, 1995) to correct for multiple testing, resulting in a total of 356 significant coefficients with a p-value under 0.01 [Supplementary Table 1].

3. Results

3.1. Map of ACL

As previously described (De Magalhães et al., 2007; Speakman, 2005), the maximum lifespan of vertebrates, with or without correction for their body mass, has a log-normal distribution. Each class of vertebrates shows a similar distribution with a class-specific mean and variance. Based on this observation, it is possible to compare the maximum lifespan distribution of vertebrates for every amino acid at every position in the 13 MEPs. We could therefore estimate the extent to which these amino acid residues are related to lifespan and draw a map of the Amino acids Correlated with maximum Lifespan (ACL) whether positively (ACL+) or negatively (ACL−). In Fig. 1, we illustrate by a hypothetic example the way we have selected these ACL. This illustration doesn't take into account the correction for phylogeny that defines the p-value and correlation coefficient of each ACL.

The multiple alignments of the concatenated MEP sequences of 324 vertebrates give a matrix of 4005 positions [Supplementary Table 1]. Among those, 296 positions have at least one amino acid that correlates, positively or negatively, with maximum lifespan for a total of 356 ACL, 206 ACL+ and 150 ACL−. There are 242 positions in the alignment sequence with a single ACL, 48 with 2 ACL and 6 with 3 ACL at a single position. Taken together, these ACL draw a map [Fig. 2] of the influent of

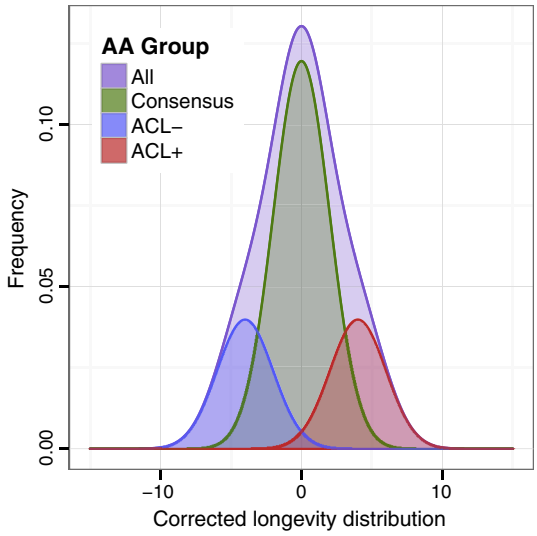

Fig. 1. Longevity distribution centered around 0 after correction for body mass. In this fictitious example, species are grouped by amino acid at a given position. The distribution of longevity for all species is featured in purple. The corresponding distribution for species exhibiting the consensus amino acid (green) is also centered around 0. By contrast, species with the ACL+ amino acid (red) have a right-shifted distribution and thus longer lifespan. Similarly, species with the ACL− amino acid (blue) have a left-shifted distribution and thus longer lifespan. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: Mariadassou, M., Pellay, F.-X., Identification of amino acids in mitochondrially encoded proteins that correlate with lifespan, Exp. Gerontol. (2014), http://dx.doi.org/10.1016/j.exger.2014.03.009