Nucleotide repeats in mitochondrial genome determine human lifespan

Konstantin Popadin[1], Georgii Bazykin[1].

1 Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Bolshoi Karetny pereulok 19, Moscow 127994, Russia.

Direct nucleotide repeats can facilitate deletions of segments of mitochondrial genome[1], leading to a wide range of neuromuscular disorders[1,2] as well as aging[2,3] in humans. We hypothesized that the number of the direct perfect repeats in human mitochondrial genomes influences longevity through the formation of harmful mtDNA deletions in the somatic cells. The analysis of the complete mitochondrial genomes of 762 unrelated Japanese individuals[4-6] reveals a negative correlation between the abundance of the direct perfect repeats and the expected longevity. This association is largely due to the disruption of the common repeat (8470,13447) by a point mutation 8473C which occurred at the origin of the D4a haplogroup characterized by extreme longevity in Japan[7]. Our results provide the first evidence for correlation between the number of nucleotide repeats and the lifespan on intraspecific level.

Phenotypic effects of aging are largely determined by the expression of hundreds of alleles which are deleterious in late life, and either neutral, slightly deleterious, or slightly beneficial in early life[8]. Germline mutations giving rise to these alleles can't be eliminated effectively through purifying selection and accumulate due to genetic drift. Some of these alleles do not have a direct phenotypic effect of their own; instead, they facilitate the somatic mutations that do have phenotypic consequences. As the mitochondrial DNA replicates intensively during the whole life cycle, mitochondrial theory of aging[9] postulates a vicious cycle whereby somatic mtDNA mutations generate excessive reactive oxygen species and these, in turn, further damage mtDNA. Since somatic mutations (predominantly deletions[10]) in mitochondrial genome are particularly likely to contribute to phenotypes associated with aging the germline mutations that increase the somatic mutability of the mtDNA are likely to have a strong effect late in life. For example, the defected nucleus-encoded mtDNA polymerase causes somatic accumulation of point mutations[11] and deletions[12] in mtDNA, which leads to reduced lifespan and premature onset of aging-specific phenotypes in mice.

Similarly, the mtDNA repeats promote occurrence of harmful deletions in the mitochondrial genomes of somatic cells. Such mutated mitochondrial genomes are favored by intracellular selection, which amplifies the deleterious effect of the repeats for the organism. Since the probability of deletions relates to mtDNA replication events, and intracellular selection takes some time the deleterious effect of mtDNA repeats might be delayed until late in life[2,3,13]. Therefore, selection against direct repeats can be expected to be stronger in long-lived animals than in short-lived ones. Indeed, a negative correlation between the number of the direct mitochondrial repeats and the maximal lifespan of mammalian species has been shown recently[14,15].

Here we hypothesized that the number of the direct perfect repeats in human mitochondrial genomes influences longevity through the formation of harmful mtDNA deletions in the somatic cells. To test this hypothesis, we analyzed the fully sequenced mitochondrial genomes of 762 unrelated Japanese individuals with known ages of sampling of blood for mtDNA analysis[5,6].

In each genome, we identified all direct perfect repeats of length 13 bp or more in which both copies of the repeat are located within the same (major or minor) arc of mtDNA[16] and are at least 30 nucleotides from each other (**Supplementary methods** online). A total of 17 repeats match these constraints. Of these, only five repeats were frequent: cccataccccgaa (535, 4430), acctcccctcacca (8470, 13447), caacataaaaccc (6410, 12031), accaacaaactta (13858, 16278) and acccccctcccca (955, 3565), while all others were only observed once or twice in our dataset (**Supplementary Table 1** online).

The number of the direct repeats has a significant negative correlation with age (Kendall's tau = -0.049, P = 0.0225, N =762). To analyze this effect in more detail, we split the population into two parts depending on whether the patient was younger ("young", N=366) or older ("seniors", N=376) than the median age of 64 years in our population. The negative correlation between the age and the number of the repeats was highly significant in the seniors (Kendall tau = -0.08733382, p = 0.005729) but not in the young (Kendall tau = -0.03316608, p = 0.1718), suggesting that the effect of the repeats is constrained to late in life.

We determined the effect of each of the 17 individual repeats separately by multiple linear regression. Two repeats had a significant effect on age: acctcccctcacca(8470,13447) (P = 0.0122) and atacttcctattc(11200,15562) (P = 0.0390), both on the major arc. The former is a frequent and well known common repeat[1,8], while the latter is a rare repeat observed only in two individuals in our sample (**Supplementary Table 1, Supplementary Figure 1** online) .

Since these two repeats are the major determinants of age, we studied their effect on longevity in more detail. The subpopulation without either of these repeats reveals a significant

excess (P = 0.01, two-sided Fisher's exact test) of centenarians and super centenarians (25 of 61, 41.0%), compared to the rest of the population (178 of 701, 25.4%). The survival curves between the two subpopulations were significantly different (P < 0.01, Cochran-Mantel-Haenszel Chi-Squared Test; **Fig. 1**), with the average expected lifespan of the individuals without these repeats (92.2 years) being 4 years longer than in individuals with either of these repeats (88.3 years).

To understand the evolutionary dynamics of these two repeats, we reconstructed the phylogenetic tree (**Supplementary Figure 2** online) and assessed their phylogenetic distribution. The common repeat acctccctcacca(8470,13447) was present in the last common ancestor of the sampled population and in most individual samples. Any single-nucleotide mutation in either copy of the repeated sequence is sufficient to destroy the repeat. Such substitutions occurred a total of four times independently on the phylogenetic tree (**Supplementary Figure 2** online). All four substitutions were observed in the first arm of the repeat (i.e. the arm starting in position 8470), in the synonymous positions of the ATP8 gene; the second arm of the repeat remained invariant. The ancestral sequence of the repeat was disrupted by a 8473C mutation at the origin of a large monophyletic group consisting of 58 sampled individuals representing the D4a haplogroup. The other three mutations – 8479G before the origin of sequence NDsq0210 and 8470G before the origin of sequences HNsq0181 and PDsq0068 – occurred on the external branches. The second repeat atacttcctattc(11200,15562) arose by a single-nucleotide mutation only once, giving rise to two individuals from our sample (**Supplementary Figure 2** online).

The analyzed Japanese population was stratified by several disease-specific groups: individuals with Parkinson's disease, individuals with Alzheimer's disease, young obese males and type-2 diabetes individuals with or without severe vascular involvement. To assess the possible contribution of the diseases to our analysis, we performed a multiple logistic regression, using patient's age, sex and disease group to predict the presence or absence of these two repeats. Again,

only age emerged as a significant factor (P = 0.0363), while neither sex nor belonging to any of the disease groups were associated with presence or absence of the repeats.

D4a is a Japanese haplogroup well known for its unusual longevity. Several works[4-7] analyzed the mtSNPs characteristic of this haplogroup without accounting for the possible link between the repeat abundance and longevity. The most recent and extensive study[7] of the association between the mtSNPs and longevity of the D4A haplogroup revealed four mtSNPs of strong effect, including the repeat-disrupting 8473C mtSNP, although the authors conservatively denied them function based on the strong phylogenetic clustering of the corresponding phenotypes evident of phylogenetic inertia. However, phylogenetic signal *per se* is not an evidence for the lack of functional significance. Indeed, all rare advantageous mutations in maternally inherited mitochondrial genome will have a strong phylogenetic signal.

Instead, our observations of the association of the two mitochondrial repeats with longevity and the few independent events of loss of the common repeat, together with the known deleterious effect of the repeats[1-3,13,16], indicate that the 8473C mtSNP caused by a mutation at the synonymous position of ATP8 gene is the main factor contributing to the longer lifespan of the D4a haplogroup, and point to the disruption of the nucleotide repeat as the likely mechanism of its effect.

Most deletions in human mtDNA appear to be caused by the (8470,13447) repeat[16]. It has been suggested that this common repeat is the principal factor behind the formation of most deletions, irrespective of the precise deletion breakpoint, and that the disruption of this repeat will substantially reduce the chance of forming most mtDNA deletions[16]. The effect of the loss of the repeat in the human population is consistent with this hypothesis: even a single nucleotide substitution in the repeated segment is sufficient to significantly increase the lifespan, probably by reducing the total deletion score. The deleterious somatic effect of the nucleotide repeats makes it important to consider the effect of mtDNA mutations on abundance of nucleotide repeats in searching and explaining the associations between mtSNPs and phenotypes.

REFERENCES

1. Schon, E. A., Rizzuto, R., Moraes, C. T., Nakase, H., Zeviani, M. & DiMauro, S. *Science* **244**, 346–349 (1989).
2. Bender, A., Krishnan, K.J., Morris, C.M., Taylor, G.A., Reeve, A.K., Perry, R.H., Jaros, E., Hersheson, J.S., Betts, J., Klopstock, T., Taylor, R.W. &Turnbull, D.M. *Nature genetics* **38**, 515-517 (2006).
3. Kraytsberg, Y., Kudryavtseva E., McKee, A.C., Geula, C., Kowall N.W. & Khrapko, K. *Nature genetics* **38**, 518-520 (2006).
4. Alexe, G., Fuku, N., Bilal, E., Ueno, H., Nishigaki, Y., Fujita, Y., Ito, M., Arai, M., Hirose, N., Bhanot, G. & Tanaka, M. *Hum Genet.* **121**, 347–356 (2007).
5. Tanaka, M., Takeyasu, T., Fuku, N., Li-Jun, G. & Kurata M. *Ann. N.Y. Acad. Sci.* **1011**, 7–20 (2004).
6. Tanaka, M. and 27 coauthors. *Genome Research* **14**, 1832–1850 (2004).
7. Bilal, E., Rabadan, R., Alexe, G., Fuku, N., Ueno, H., Nishigaki, Y., Fujita, Y., Ito, M., Arai, Y., Hirose, N., Ruckenstein, A., Bhanot, G. & Tanaka, M. *Plos One* **3**, e2421(2008).
8. Cortopassi, G.A. *Free Radic Biol Med.* **33**, 605–610 (2002).
9. Harman D. *J. Am. Geriat. Soc.* **20**, 145–47 (1972).
10. Harding, A.E. *Nat Genet.* **2**, 251–52 (1992).
11. Trifunovic, A., Wredenberg, A., Falkenberg, M., Spelbrink, J.N., Rovio, A.T., Bruder, C.E., Bohlooly-Y, M., Gidlof, S., Oldfors, A., Wibom, R., Tornell, J., Jacobs, H.T., Larsson N-G. *Nature* **429**, 417-423 (2004).
12. Vermulst, M., Wanagat, J., Kujoth, G.C., Bielas, J.H., Rabinovitch, P.S., Prolla, T.A. & Loeb, L.A. *Nat Genet.* **40**, 392-394 (2008).
13. Cortopassi, G.A. & Arnheim, N. 1990. *Nucleic Acids Res.* **18**, 6927-6933 (1990).
14. Samuels, D.C. *Trends Genet.* **20**, 226-229 (2004).
15. Khaidakov, M., Siegel, E.R. & Shmookler, R.J. *Mech Ageing Dev.* **127**, 808–812 (2006).
16. Samuels D.C., Schon, E.A. & Chinnery, P.F. *Trends Genet.* **20**, 393-398 (2004).

ACKNOWLEDGMENTS

We thank Masashi Tanaka for discussion and providing the data of sex and age of each patient, Alexey Kondrashov, Mikhail Gelfand, Shamil Sunyaev, Anatolii Teriokhin and Leila Mamirova for comments and helpful discussions. This research was supported by the Molecular and Cellular Biology Program of the Russian Academy of Sciences, by Russian Foundation for Basic Research Grants 08-04-09680 and 08-04-01394, and by the Russian Science Support Foundation Grants to K.P. and G.B.

Figure 1. Survival curves for individuals with (dashed line) and without (solid line) either of the repeats acctccctcacca(8470,13447) and atacttcctattc(11200,15562).