**Comparative analysis of molecular basis underlying mammalian lifespan**

Dongming Xu1, Shaogang Qu1

1 Central Laboratory and Department of Neurology, Shunde Hospital, Southern Medical University (The First People’s Hospital of Shunde Foshan), Foshan, Guangdong, China

**\*Corresponding author:** Shaogang Qu (shuifeng.paper@aliyun.com)

**Abstract**

The maximum lifespan varies dramatically across mammals. However, the molecular mechanisms underlying such variation are poorly understood. Here, we genomic-wide explored the correlation between protein evolutionary rate and lifespan in 13381 one-to-one orthologue genes from 74 mammals. We totally identified 370 negatively correlated genes and 300 positively correlated genes. Evolutionary analyses further suggested that most of the positively correlated genes were driven by relaxed selection in long-lived mammals. Enrichment analyses showed that the positively correlated genes and the negatively correlated genes were not only uniquely enriched in many canonical pathways that regulation of lifespan and aging respectively, but also were overrepresented in several groups of categories with adverse or closely functions simultaneously, such as oxygen utilization, purine nucleotide biosynthesis and purinergic nucleotide receptor signaling pathway. Finally, combing the results of correlative and evolutionary driving forces analyses, our signed functional interaction network analyses uncovered several modules with different patterns of evolutionary constraints and highlighted the core genes that might play central role in regulation of mammalian lifespan. Our genomic-wide analyses uncovered the common molecular mechanism underlying lifespan determination, which may provide new routes for delaying human aging and decreasing aging-associated diseases.

**Highlights**

**Largest genomic data ever used to uncover the protein-coding genes underlying regulation of mammalian lifespan, including 13381 one-to-one orthologous genes across 74 mammals.**

**Most positively correlated genes were driven by relaxed selection in long-lived mammals.**

**Positively correlated genes and negatively correlated genes enriched in several groups of categories with adverse or closely relative functions simultaneously, such as oxygen utilization, purine nucleotide biosynthesis and purinergic nucleotide receptor signaling pathway.**

**Comprehensive network analyses uncovered several modules with different patterns of evolutionary constraints and highlighted the core genes that might play central role in regulation of mammalian lifespan.**

**Introduction**

Aging and aging-associated diseases present a major threat to human health and are affected by inherently complex process1, 2. However, the genetic mechanisms of human aging determination are still poorly understood3. Most of our knowledge about the genetic mechanisms that govern aging were obtained by studying genetic manipulations of short-lived laboratory animal models2, 4. It is unclear if insights from such studies can be transferred to long-lived mammals like humans. Moreover, insights from lab studies of model animals may not represent the universal mechanisms of longevity regulation, instead, it could be due to species-specific adaptation or coincidental neutral changes5-7.

Fortunately, the rate of aging varies dramatically across wild mammals8. Maximum lifespan (MLS), which can reflect the inherent longevity and “rate of aging” in organisms9, is positively correlated with their body size10. However, many mammals are known to deviate their expected MLS estimated from body mass11. For example, the naked-mole rat and Brandt’s bat (Myotis) can live at least 31 and 41 years, over 4 and 10 times longer with respect to the predicated lifespan based on their body mass, respectively12. On the contrary, the rat, mouse and shrew can only live about half of their expected MLS even in the well-cared laboratory13. More importantly, these long-lived or short-lived species exhibit delayed or accelerated age-associated physiological declines. For example, the long-lived naked-mole rat exhibit little age-specific hazard of mortality and fertility14 and the longest-lived genus (*Myotis*) maintains the length of their telomeres with aging and dose not develop cancer15, while cancer-related mortality could be up to 90% in short-lived mice2, 16. Therefore, the repeated changes of MLS in mammals could provide a good opportunity for exploring the molecular mechanisms underlying regulation of lifespan and aging.

The evolution of protein-coding sequences plays central role in regulation of lifespan, for example, DNA repair and p53 signaling pathway associated genes were often showed positive selection and convergent evolution in long-lived mammals3, 7, 17, 18. However, most of previous studies only focused on a few long-lived lineages3, 17, 19, 20, thus cannot differentiate whether the uncovering molecular mechanisms is lineage-specific or universal across mammals6. Moreover, most of genomic-wide comparative studies lie emphasis on the contributions of positive selection or convergent evolution in regulation of lifespan3, 17, little concerned about the roles of relaxation of evolutionary constraints. In spite of that relaxed selection has been documented to be a dominant driving forces in the evolution of many traits21 and depletion of numerous genes can significantly extend MLS and delay aging in model animals22, 23.

Recently, a new method that tests for association between relative evolutionary rates of genetic elements and the evolution of traits across a phylogeny (*Reconverge*)24, 25 has been developed. This method can search for convergent shifts in evolutionary rates of individual protein-coding genes that response for convergent phenotypes. The evolutionary rates can well reflect their evolutionary constraints. An increased selective constraint can lead to a slower evolutionary rate of protein-coding genes, whereas faster evolutionary rates can result from a relaxation of constraint or from intensified selection (positive selection)24. A previous study used *RERconverge* methods to uncover the molecular basis underlying lifespan found that negatively correlated genes (NCGs) between protein evolutionary rate and lifespan were enrichen in DNA repair, immunity, cell cycle and cell death-related pathways6. However, the authors only used 35 mammals in their mainly analyses and they did not explore the positively correlated genes (PCGs) and their evolutionary driving forces6.

In this study, PCGs indicate genes have been experienced rapid evolution in long-lived mammals. Many rapid evolution genes were suggested to affect lifespan in mammals, such as energy metabolism associated genes19, 22, 23. Rapid evolution gene is always driven by positive selection or relaxed selection and distinguishing them can promote the understandings and utilizations of the detailed molecular mechanisms. RELAX is a popular method to detect the relaxed selection in protein-coding genes26, which makes exploring the extent to which PCGs and relaxed selection contributing to the evolution of lifespan feasible. Moreover, lifespan is affected by complex genetic factors1, 2, combining analyses of PCGs and NCGs and their evolutionary driving forces might provide more insights into the molecular mechanisms underlying regulation of lifespan in mammals.

Here, we integrated the lifespan phenotypes from HAGR (Human Ageing Genomic Resources) database12 and the orthologous genes from OrthoMaM database27 to obtain the largest genomic data ever used to explore the universal molecular basis underlying regulation of mammalian lifespan. We genomic-wide identified significantly correlated genes between protein evolutionary rate and lifespan across mammals and further explored their evolutionary driving forces, especially for positively correlated genes. In addition, we constructed a signed functional interaction network to explore the important modules and core genes that might play central role in universal regulation of mammalian lifespan. Our study will deepen our knowledges of the molecular mechanisms underlying regulation of mammalian lifespan and provide targets for further experiment verification and drug development.

**MATERIALS AND METHODS**

**Integration of lifespan phenotypes and genomic data**

The protein alignment of one-to-one orthologous genes were downloaded from OrthoMaM database (V10b)27, which includes 14509 genes from 116 species. The data of observed maximum lifespan and adult weight were extracted from the Human Aging Genomic Resource (HAGR) database12, with 999 mammals having both traits. Expected maximum lifespan was calculated according to their adult weight using the formula: *expect lifespan (years)=* *3.34\*(adult weight (g)^0.193)*10 and the Longevity quotients (LQ) value is considered as the ratio of observed maximum lifespan to expected maximum lifespan. Eventually, 74 species with LQ values and genomics data were used in this study (**Figure 1 and Supplementary Table 1**). The HAGR database not only records the lifespan phenotypes, but also collects genes that affect longevity and aging. We downloaded the longevity-associated genes from 4 HAGR database, including LongevityMap (human genetic variants associated with longevity), AnAge (Curated genes affect aging and life history in animals), CellAge (genes affect Cell Senescence) and GenDR (Different expression genes after Dietary Restriction).

**Filtration of orthologous genes and alignment regions**

Large-scale one-to-one orthologous identification and multiple sequence alignment were still full challenges for . To exclude the fake orthologous sequences in the one-to-one orthologous genes, we downloaded the human and mouse protein references from the UCSC database. Then, we searched each protein sequence in each one-to-one orthologous gene to the human and mouse references using blastp method, respectively. If the best hit of protein sequence in one-to-one orthologous gene is different with that of human when searched for the human protein reference or is different with that of mouse protein when searched for the mouse protein reference, we removed the protein sequence from the one-to-one orthologous genes (**Supplementary Table 2**).

Next, to remove the poorly aligned regions in one-to-one orthologous genes, we used *trimAL*28 to filter the original alignments with parameter “-resoverlap 0.70 -seqoverlap 50-automated1 -colnumbering”. Finally, only orthologous alignment with ≥ 50 species and with length ≥ 50 amino acids were retained. The distribution of specie number and alignment length in one-to-one orthologous before and after filtering can be seen in **Supplementary Figure 1**. The distribution of the number of the species and the protein alignment length of each orthologous before and after filtering process were shown in **Supplementary Figure 2a and Supplementary Figure 2b**.

**Identification of significantly correlated genes**

The phylogenetic relationship of 74 species was obtained from the OrthoMaM database (v10b)27 (**Figure 1a**). The *nw\_prune* module in *newick-utils* (v1.6) package29 was used to generate the gene tree for each one-to-one orthologous gene based on the species tree of 74 mammals and the missing species. Next, the *codeml* module from the *PAML* (V4.7) packages30 was used to calculate the branch lengths with Empirical + F model. Then, we used R package *RERconverge* (v0.1.0)24 to compute the association statistic between LQ value changes and relative evolutionary rates (RER) for each orthologous gene. Briefly, RERconverge read all gene trees with their branch length values and calculate the average rate for each branch in the tree (**seen Phylogenetic tree section**). Then, RERconverge calculate gene-specific rates of evolution, termed relative evolutionary rates (RER) through average rate normalization. This correction can remove the non-specific factors affecting divergence on the branch such as time since speciation and mutation rate. Also, *RERconverge* read all LQ values of 74 mammals as continuous phenotypes and inferred the change of LQ value in each branch (**Supplementary Figure 2**) using maximum likelihood method. Finally, *RERconverge* computed the Pearson correlation between LQ value changes and relative evolutionary rates (RER) with default parameters, excluding the minimum number of species was set as 50. The Pearson correlative coefficient between relative evolutionary rates of genes and the evolution of traits across a phylogeny, Rho, indicate whether the relative evolutionary rate of a particular gene is positively or negatively correlated with the LQ values across the phylogeny. A rough *P* value at 0.05 was used as the threshold for determining significance. we called the significantly correlated genes with positive Rho value as positively correlated genes (PCGs), and the significantly correlated genes with negative Rho value as negatively correlated genes (NCGs) (**Supplementary Table 3-4**).

**Determination of evolutionary forces**

The coding sequence alignments of 74 mammals using in this study were download and extracted from OrthoMaM database (v10b)27. We removed the non-orthologous sequences and poorly alignment regions that dropped by the protein filtering process. Then, we considered the long-lived mammals as foreground group and considered other species as background group. However, it was difficult to determine which species were long-lived mammals, so we considered species with top 10 LQ values in the 74 mammals as long-lived mammals in this study. The rest species were treated as background group. *Hyphy* package developed by RELAX framework26 was used to determine the evolutionary selective forces in the foreground group according to the distribution of the nonsynonymous to synonymous substitution rates (ω or dN/dS) in the codon sites. Briefly, intensified selection (positive selection) will push all ω categories away from neutral evolution (leading to parameter k > 1), whereas relaxed selection will push all ω categories toward neutral (k < 1). Finally, the maximum likelihood method was used to evaluate whether alternative hypothesis is better than the null hypothesis. The null hypothesis indicates no different evolutionary selection between foreground group and background group. Two-sided Chi-square was used to statistically significant test. A rough *P* value at 0.05 was used as the threshold for statistical significance (**Supplementary Table 6-7**).

In addition, we also used branch-site model31 to detect positive selection among all significantly correlated genes for long-lived species. The species with top 10 LQ values were label as foreground. Two-sided Chi-square was used to statistically significant test. A rough *P* value at 0.05 was used as the threshold for statistical significance (**Supplementary Table 8**).

**Enrichment analyses**

Firstly, we used ClusterProfiler package32 to search for the overrepresented GO Biological Process (GO BP) terms from the significantly correlated genes. The PCGs and NCGs were used as foreground genes, respectively. All the one-to-one orthologous genes were considered as background genes. We removed the overrepresented GO BP terms having more than 300 background genes because these GO BP terms are usually at lower GO BP levels and can give us little information. The overrepresented GO BP terms can be seen in **Supplementary Table 9-10**. We also searched the relatively overrepresented GO BP terms in PCGs when compared with NCGs, and reverse. Fisher’s single-tailed test with *P value* at 0.05 was used as the threshold for statistical significance (**Supplementary Table 11-12**). In addition, we used REVIGO33 to remove the redundantly significant GO BP terms with medium cutoff standard (allowed similarity=0.7) (**Figure 3**).

**Network analyses**

Firstly, we extracted the direct interactions among all significantly correlated genes based on the Reactome Functional Interactions database (2020)34. The direct interaction includes catalyze, inhibit, activate and expression regulate. Next, we extracted genes from the direct interactions and constructed the signed functional interaction network of these genes using *ReactomeFIViz* module in software *Cytoscape* (v3.8.0)35. Then, we manually removed the predicted interactions and added the information of correlative and evolutionary forces analyses to the network (**Figure 6**).

**Phylogenetic tree**

Full tree with average branch lengths（74 mammals）:

(((((((((((*Acinonyx* *jubatus*: 0.00962, *Felis* *catus*: 0.00881): 0.00258, *Panthera* *pardus*: 0.00757): 0.05450, (((*Ailuropoda* *melanoleuca*: 0.02155, *Ursus* *maritimus*: 0.01937): 0.02431, *Mustela* *putorius*: 0.05682): 0.01400, *Canis* *familiaris*: 0.05820): 0.01480): 0.03970, (*Equus* *asinus*: 0.01069, *Equus* *caballus*: 0.01036): 0.08607): 0.00506, (((((*Bos* *taurus*: 0.01593, *Bubalus* *bubalis*: 0.01885): 0.01128, (*Capra* *hircus*: 0.00931, *Ovis* *aries*: 0.01601): 0.01906): 0.07207, ((*Delphinapterus* *leucas*: 0.01318, (*Orcinus* *orca*: 0.00593, *Tursiops* *truncatus*: 0.01308): 0.00928): 0.00697, *Lipotes* *vexillifer*: 0.02230): 0.05676): 0.01510, *Sus* *scrofa*: 0.09043): 0.00793, ((*Camelus* *bactrianus*: 0.00595, *Camelus* *dromedarius*: 0.00749): 0.01449, *Vicugna* *pacos*: 0.02803): 0.07332): 0.02829): 0.00441, ((*Eptesicus* *fuscus*: 0.03530, (*Myotis* *brandtii*: 0.00951, *Myotis* *lucifugus*: 0.01344): 0.03228): 0.10311, ((*Pteropus* *alecto*: 0.00744, *Pteropus* *vampyrus*: 0.01511): 0.02461, *Rousettus* *aegyptiacus*: 0.04263): 0.06533): 0.02508): 0.00780, ((*Condylura* *cristata*: 0.13130, *Sorex* *araneus*: 0.19886): 0.01742, *Erinaceus* *europaeus*: 0.20942): 0.02365): 0.01808, ((((((*Callithrix* *jacchus*: 0.03419, (*Cebus* *capucinus*: 0.01994, *Saimiri* *boliviensis*: 0.02460): 0.00400): 0.03614, (((*Colobus* *angolensis*: 0.01296, *Rhinopithecus* *roxellana*: 0.01132): 0.00462, (((*Macaca* *fascicularis*: 0.00433, *Macaca* *mulatta*: 0.00613): 0.00164, *Macaca* *nemestrina*: 0.00636): 0.00291, *Mandrillus* *leucophaeus*: 0.00562): 0.00640): 0.01721, ((*Gorilla* *gorilla*: 0.00854, (*Homo* *sapiens*: 0.00739, (*Pan* *paniscus*: 0.00359, *Pan* *troglodytes*: 0.00578): 0.00413): 0.00201): 0.00821, *Nomascus* *leucogenys*: 0.02502): 0.00898): 0.01535): 0.04728, *Carlito* *syrichta*: 0.09858): 0.00922, (*Microcebus* *murinus*: 0.07135, *Otolemur* *garnettii*: 0.09512): 0.03209): 0.01214, *Tupaia* *belangeri*: 0.14817): 0.00665, ((((*Castor* *canadensis*: 0.10185, *Dipodomys* *ordii*: 0.14725): 0.02250, (*Jaculus* *jaculus*: 0.14680, ((*Meriones* *unguiculatus*: 0.09506, (*Mus* *musculus*: 0.05877, *Rattus* *norvegicus*: 0.07042): 0.04346): 0.01597, (*Mesocricetus* *auratus*: 0.06874, (*Microtus* *ochrogaster*: 0.07757, *Peromyscus* *maniculatus*: 0.06254): 0.00724): 0.02165): 0.09537): 0.03727): 0.01749, ((((*Cavia* *aperea*: 0.01958, *Cavia* *porcellus*: 0.00697): 0.09341, (*Chinchilla* *lanigera*: 0.05949, *Octodon* *degus*: 0.09791): 0.01558): 0.02183, *Heterocephalus* *glaber*: 0.09525): 0.08295, *Ictidomys* *tridecemlineatus*: 0.09997): 0.01068): 0.01837, (*Ochotona* *princeps*: 0.15556, *Oryctolagus* *cuniculus*: 0.07011): 0.06955): 0.01203): 0.01683): 0.01862, (*Choloepus* *hoffmanni*: 0.11838, *Dasypus* *novemcinctus*: 0.10245): 0.05085): 0.01123, (*Echinops* *telfairi*: 0.23641, (*Loxodonta* *africana*: 0.07234, *Procavia* *capensis*: 0.14928): 0.02905): 0.03295): 0.20999, (*Monodelphis* *domestica*: 0.08938, (*Phascolarctos* *cinereus*: 0.06598, *Sarcophilus* *harrisii*: 0.08278): 0.02853): 0.17655, *Ornithorhynchus* *anatinus*: 0.33830);

**RESULTS**

**Mammals with genomic data and lifespan phenotype**

The values of observed maximum lifespan and adult weight of mammals were extracted from HAGR database12. There are 999 mammals include both traits. The expected maximum lifespans were calculated according to their adult weights based on previous studies10 (**seen methods**). The Longevity quotients (LQ) is considered as the ratio of the observed maximum lifespan to the expected maximum lifespan17. The original protein alignment of one-to-one orthologous were downloaded from the OrthoMaM database (v10b)27, including 14509 genes span 116 mammals, Finally, 74 mammals include both genomic sequences and lifespan phenotypes were used in following study (**Supplementary Table 1 and Figure 1a**). The adult weight of our species ranges from 4,800,000g (elephant) to 7g (Brandt's bat) and the maximum lifespan ranges from 122.5 years (human) to 2.5 years (Star-nosed mole), which can cover the major diversity of mammals (**Figure 1b**).

**Filtration of one-to-one orthologous sequences**

The accuracy of calculating the evolutionary rate is largely depended on the quality of the orthologous identification and the alignment accuracy28. We totally removed 7838 artificial one-to-one orthologous sequences from 1169 genes through *blastp* method (**seen methods, Supplementary Table 2**). Next, we used *TrimAL* to remove the poorly aligned fragments which may be caused by inaccurate alignment or incorrect genomic annotation28. We dropped the orthologous genes with less than 50 species or with alignment length shorter than 50 amino acids. Finally, 13381 one-to-one orthologous genes were remained. The average number of species was 65.9 and the average length of protein alignment was 583.0 after the filtering process (**Supplementary Figure 2a and Supplementary Figure 2b**).

**Identification of significantly correlated genes**

Using RERconverge package, we identified 370 negatively correlated genes (NCGs) (**Supplementary Table 3**) and 300 positively correlated genes (PCGs) (**Supplementary Table 4**). Although these genes are not enriched in longevity genes collected by HAGR database12 (**Supplementary Table 5**). This may be caused by most of HAGR longevity-associated genes were obtained from human population and short-lived laboratory animal models and thus was not a good representative of the common molecular basis underlying lifespan evolution. For example, 5 of the 6 most significantly correlated genes might affect lifespan or aging (**Figure 2**), none of them was collected by HAGR database. Such as, a minor SNP in *ADAMTS6* (ADAM metallopeptidase with thrombospondin type 1 motif, 6), the most positively correlated gene (**Figure 2a**), has protective effective on aging in a GWAS study36. Another locus (rs12199884) in *PKHD*1 (polycystic kidney and hepatic disease 1), the most negatively correlated gene (**Figure 2d**), is negatively associated with longevity in males from Han Chinese population GWAS study37. Deletion of *YbeY* (YbeY Metalloendoribonuclease), the second most negatively correlated gene (**Figure 2e**), will shorten the bacteria longevity38. *COL4A2* (Collagen type IV alpha2), the third most positively correlated gene (**Figure 2c**), is positively selected in long-live naked mole rat and down-expression in ant-aging mutant mice2. *SLC11A1* (Solute Carrier Family 11 Member 1), the third most negatively correlated gene (**Figure 2f**), is different expression and methylation in multiple human aging tissues39, 40. These results suggested our significantly correlated genes are highly reliable and further supported RERconverge correlative analysis is a good method to uncover genes underlying traits evolution.

**Determine of evolutionary driving forces**

We used RELAX method to uncover the evolutionary driving forces underlying the evolution of significantly correlated genes in long-lived mammals, especially for PCGs. We totally identified 2716 relaxed selected genes (**Supplementary Table 6**) and 514 intensified selected genes in long-lived mammals (**Supplementary Table 7**). Our results showed that only 3 (1.0%) PCGs showed intensified selection in long-lived mammals, which was significantly less than that of non-significantly correlated genes (4.1%) and that of NCGs (7.0%) (**Figure 3a**). Branch-site model also found the ratio of positively selected genes in PCGs was significantly less than that of NCGs (**Supplementary Figure 3 and Supplementary Table 8**). In adverse, we found 188 (62.7%) PCGs showed relaxed selection in long-lived mammals, which was almost 3 times higher than that of non-significantly correlated genes (*P*=3.03E-11) and was over 6 times more than that of NCGs (*P*=1.84E-52) (**Figure 3b**). These results suggested that most of the PCGs might be driven by relaxed selection in long-live mammals, which indicates that knock-out or down-regulation of these genes might contribute to extend lifespan or enhance anti-aging in mammals.

**Enrichment analyses**

Biological Process of Gene Ontology (GO BP) enrichment analyses showed that both the PCGs and the NCGs were uniquely enriched in many canonical pathways that regulate lifespan and aging, respectively. The most overrepresented categories by NCGs were involved in cell division, cell cycle, DNA repair, and p53 signaling pathway (**Supplementary Table 9 and Figure 4a**), which was very similar to the previous study6. However, we further found that many of these overrepresented categories did not include any PCGs, such as “regulation of DNA repair”, “regulation of cell division” and “regulation of DNA recombination” (**Supplementary Table 11 and Figure 4a**). The most overrepresented categories by PCGs were involved in energy metabolism, insulin/IGF-1 signaling pathway, glutamate receptor signaling pathway, ubiquitin-dependent proteolysis, calcium ion transmembrane transport and so on (**Supplementary Table 10 and Figure 4b**). Similarity, many of these pathways also did not include any NCGs, such as “positive regulation of glucose import”, “positive regulation of ubiquitin-dependent protein catabolic process” and “ionotropic glutamate receptor signaling pathway” (**Supplementary Table 12 and Figure 4b**). Combing the results of evolutionary driving forces analyses, we found that many PCGs in their overrepresented categories were driven by relaxed selection in long-lived mammals, little were driven by positive selection (**Supplementary Table 10**).

Moreover, we further found several groups of categories with adverse functions or closely relative functions that were enriched by PCGs and NCGs simultaneously. If a group of categories with adverse functions was enriched by NCGs and PCGs respectively, it was about to double verify that these functional categories might affect longevity or aging. If a group of categories with closely relative functions that were enriched by NCGs and PCGs respectively. It might reflect the pleiotropy of these genes or categories. Our results showed categories associated with oxygen utilization and response to hypoxia might be uniquely enriched by PCGs and NCGs, respectively. 7 PCGs, including 4 relaxed selected genes (ACTN3/SUCLG2/ACO2/OGDHL), were overrepresented in category “aerobic respiration” (P=2.80E-04) (**Supplementary Table 10**), which was significantly more than that of NCGs (0 gene, *P*=0.0062) (**Supplementary Table 12 and Figure 5a**). In adverse, 4 NCGs (*CYBB/PSMB11/STOX1/BCL2*) were associated with category “cellular response to hypoxia”, while no PCGs involved in this pathway (*P*=0.0929) (**Figure 5a**). Several studies have suggested that hypoxia may affect the longevity. For example, hypoxia can extend the lifespan in fruit fly41 and nematode42. Also, elderly people living in hypoxia in the Tibetan Plateau tend to have a longer life than similarly aged people in other Chinese regions43. In addition, 5 NCGs (*CYBB/RECQL5/BLM/ADCY6/SPIDR*) were enriched in category “cellular response to alcohol” (*P*=0.0096) (**Supplementary Table 9 and Figure 5a**), while no positively correlated gene involved in this pathway (*P*=0.0511) (**Figure 5a**). Alcohol metabolism can cause oxygen deficiencies or hypoxia, which in turn impedes ATP production and thus contributes to cell death44. These results suggested enhancing hypoxia resistance and weakening aerobic metabolism might be a common molecular mechanism in the regulation of lifespan across mammals.

We also found a group of categories with closely relative functions that were enriched by PCGs and NCGs, simultaneously. For example, 4 NCGs (*P2RY2/P2RY10/GPR171/P2RY14*) were associated with “G-protein coupled purinergic nucleotide receptor signaling pathway” (*P*=2.38E-04) (**Supplementary Table 9**), while no positively correlated gene was associated with this pathway (*P*=0.0929) (**Figure 5b**). Interesting, all these 4 genes belong to P2Y receptors family. Although, we know little about the relationship between P2Y receptors and lifespan, P2Y receptors affect aging-related disease, such as Alzheimer disease45, 46. Surprisingly, we found that PCGs were enriched in many other purine-related categories. 7 PCGs involved in “purine ribonucleoside diphosphate metabolic process” (*P*=0.0076) (**Supplementary Table 10**), which was significantly more than that of NCGs (1 gene, *P*=0.0167) (**Supplementary Table 12 and Figure 5b**). In addition, PCGs were also overrepresented in other purine biosynthetic categories, such as “regulation of purine nucleotide biosynthetic process” and “purine-containing compound biosynthetic process” (**Supplementary Table 12 and Figure 5b**). Previous studies showed that addition of dietary adenine can shorten the lifespan by 48%47 and blocks the longevity effects of dietary restriction48. These results suggested that genes involved in purine biosynthetic and purine signaling transduction might be suffered different evolutionary constraints in the regulation of lifespan or aging.

**Network analyses**

To further explore the modules and core genes underlying the regulation of lifespan in mammals, we constructed a signed functional interaction network using all significantly correlated genes and integrated the information of their evolutionary driving forces (**Figure 6**). From the network, we can obviously identify at least five modules with different patterns of evolutionary constraints (**Figure 6 Module1-5**) For example, for Modue2 and Module 3, which involved in ribosome biogenesis and DNA repair, respectively. Almost all genes in these modules were experienced more evolutionary constraints in long-lived mammals. Enhancing the capability DNA repair49 and slowing down the ribosome turnover50 are two fundamental mechanism for extending lifespan across eukaryotic speciesa50. WD Repeat-Containing Protein 46 (*WDR46*) might be the key gene in Module 2. All genes in this module activated or catalyzed *WDR46* (**Figure 6 Module 2**). Previous study showed that depletion of wdr-46 reduced the lifespan in nematode51. So, WDR46 might also play an essential role in the regulation of lifespan in mammals. In Module3, we found that the genes in this module mainly activated *FANCG*, *FANCC* and *PRIM1*. Thus, these 3 genes may play central role in mediation of the capability of DNA repair and regulation of lifespan in mammals (**Figure 6 Module 3**).

The function of Module 1 mainly involved in insulin-associated signaling pathways, such as positive regulation of MAPK cascade and transmembrane receptor protein tyrosine kinase signaling pathway (RTKs). These pathways are central signaling pathways that regulate a wide variety of stimulated cellular processes, including proliferation, differentiation, apoptosis and stress response52. Here, we found that most of genes directly activate these pathways were PCGs and many of them were experienced relaxed selection in long-lived mammals, such as *INSR*, *SH3GL2*, *ANGPT1* and *PRKCZ* (**Figure 6 Module 1**). The insulin receptor (INSR) was the core gene in this module, which was also suggested to be experienced rapid evolution in long-lived mammals by a rent study53. What is more, several GO categories that regulated by insulin-associated signaling pathways were also overrepresented by PCGs, such as “positive regulation of glucose import” and “positive regulation of growth” (**Figure 4b**). However, we also found 5 genes in the upstream of the pathways were consistently experienced increasing evolutionary constraints, especially three of them (*MUC20/EPHB2/SYNJ1*) were experienced positive selection in long-lived mammal. Positively selected gene were relatively rare either in our network or in our significantly correlated genes (**Figure 5b and Supplementary Figure 3**). These genes might also regulate lifespan or aging in mammals. For example, mucin 20 (*MUC20*), which might be a negative regulator of MAPK cascade and RTKs, was one the most prominent calorie restriction (CR)-associated genes in mice54. The pattern of evolutionary constraint in Module 5 was similar to Module 1. Module 5 like a protein complex and involved in cell cycle and cell division. Most genes in this module were NCGs. However, *NUP107*, the only gene in this module that direct regulate other genes was a positively correlated gene. *NUP107* may combine other protein, such as *NUP88* and *NUP155* to activate CPSF3 and NCBP1 (**Figure 6 Module 5**).

The function of Module 4 mainly involved in cell adhesion, cell matrix interaction, extracellular matrix organization and so on. This module might be regulated by MAPK cascade (**Figure 6 Module 4**) and 4 of 12 members in this module (*CAV1/PTK2B/ATCN3/COL15A1*) might affect longevity or aging according to HAGR database (**Supplementary Table 5**). Most genes in Module 4 were positively correlated gene and many of them were experienced relaxed selection in long-lived mammals. However, the integrin subunit beta 7 (*ITGB7*), which was activated by other 4 members in this module, was experienced increased evolutionary constraint in long-lived mammals (**Figure 6 Module 4**). *ITGB7* was an integrin receptor that function in signaling from the extracellular matrix to the cell. Although we know little information about the biological implication of *ITGB7*, it is worth exploring the effects of *ITGB7* in the regulation of lifespan or anti-aging in mammals in future. In addition, our GO enrichment analyses above also suggested some categories with similar pattern of evolutionary constraint in Module 4. For example, PCGs were enriched in purine nucleotide biosynthetic (**Supplementary Table 10 and Figure 5b**), while NCGs were overrepresented in P2Y receptors, a family of purinergic G protein-coupled receptors (**Supplementary Table 9 and Figure 5b**).

Beside five modules above, other genes might also involve in regulation of lifespan or aging in our network. For example, cyclin D1 (*CCND1*), a positively correlated gene, was targeted by largest number of members in our signed functional interaction network (**Figure 6**). Overexpression of *CCND1* will lead to cell senescence55. Another most targeted gene was BCL2 apoptosis regulator (*BCL2*), a negatively correlated gene, which was expression regulated by 4 genes. *BCL2* is a negative regulator of autophagy, disruption of the beclin 1-Bcl2 autophagy regulatory complex promotes longevity in mice56 and down-regulation of Bcl-2 expression controls murine dendritic cell longevity57.

5 of 15 genes in this module (*IGFBP3\IGF1\INSR\RET\PIK3R5*) were associated with regulation of lifespan or aging according to HAGR database (**Supplementary Table 5**).

**DISCUSSION**

Pearson correlation in *RERconverge* package was used to determine whether a gene was significant correlation between protein evolutionary rate and lifespan across the phylogeny. The reliability and sensitivity of correlative analysis largely depends on the sample size. Here, we collected 74 mammals, which indicated our correlative analyses were based on up to 155 pairs of data (74 terminal branches and 71 internal branches (unroot tree), **Supplementary Figure 2**). Even the orthologous genes with fewest number of species (50 mammals), it also includes 97 pairs of data. However, the previous study used 34 mammals as maximum in their mainly study. Although the authors also analyzed 61 mammals in their additional analyses, but they did not limit their analyses on one-to-one orthologous genes and gene duplications will affect the evolutionary constraints and evolutionary rates58. Thus, our analyses might be more reliability and sensitivity in identifying significantly correlated genes than the previous study. In addition, we used simple principle to remove the fake one-to-one orthologous genes and used *TrimAL* to drop the poorly aligned region, which will further promote the accuracy of calculating branch length and thus promote the reliability of correlative analyses.

Combining the PCGs with NCGs not only can promote the reliability of our enrichment analyses, but also might help us find more interesting categories. For example, both our study and the previous study found NCGs were enriched in DNA repair pathways. However, our study also found that almost no PCGs involved DNA repair pathways (**Supplementary Table 11**). This further suggested that DNA repair associated genes might be overall experienced more evolutionary constraints in long-lived mammals. As well, many categories that were enriched by positive correlated genes have no any negatively correlated gene involved in. such as, positive regulation of glucose import, positive regulation of insulin-like growth factor receptor signaling pathway, ionotropic glutamate receptor signaling pathway, positive regulation of ubiquitin-dependent protein catabolic process and so on (**Supplementary Table 12**). We further found the PCGs and NCGs enriched in some adverse categories, simultaneously, such as PCGs were enriched in aerobic respiration, while NCGs were enriched in cellular response to hypoxia and cellular response to alcohol. This further suggestedgenes associated with oxygen utilization might play important role in universal regulation of lifespan in mammals.

Reverse categories were enriched by PCGs and NCGs simultaneously will further supported genes involved in these categories might play important roles in regulation of lifespan or anti-aging.

Closely relative categories were enriched by PCGs and NCGs suggested that different part of the pathway might be suffered different evolutionary constraints. This may be caused by the pleiotropy of genes. several

Evolutionary driving forces supported that relaxed selection play very import roles in universal regulation of lifespan in mammals. We found 62.7% of PCGs were suffered relaxation of evolutionary constraints in long-lived mammals, much more than only 1% of PCGs were suffered positive selection. However, we could not exclude the important roles of positively selected genes and we could not state that relaxed selection is more important than positive selection underlying the evolution of lifespan in mammals because the signals of positive selection cannot maintain for long time30.

Comparing to protein-protein interaction network, our signed functional network analyses includes the additional information of directions and signs, which can clearly show the causality of the network and contribute us to identify the hub genes in the modules59. For example, we will be difficult to distinguish the hub genes in Module 2 and Module 4 in the protein-protein interaction network because the proteins in these modules interacted with each other very closely. However, after adding the information of directions and signs, we can clearly determine that WDR46 and ITGB7 were the hub gene in Module 2 and Module 7, respectively. Moreover, combing the evolutionary constraints results, we found three patterns of evolutionary in different modules. The first pattern was almost all genes in the modules were suffered more evolutionary constraints in long-lived mammals, such as Module 2 and Module 3. These modules were associated with maintain genome stability. Long-lived mammals is well known can maintain more genomic stability in old ages.

The second pattern was the regulator genes in the upstream of signaling pathway were suffered more evolutionary constraints, but the target genes in the downstream of signaling pathways were suffered relaxation of evolutionary constraints, such as Module 1 and Module 5. These modules involved in cell division and cell proliferation. long-lived mammals usually need more cell division and cell proliferation, and more cell division will increase the risk of developing cancer. However, many long-lived mammals have lower risk of developing cancer, which is similar to Peto’s Paradox. Here, we found that genes associated with genomic stability and regulate cell division were suffered more evolutionary constraints in long-lived mammals, but the genes directed induced cell division were suffered relaxed selection in long-lived mammals, which might explain why long-lived mammals can resist cancer.

The third pattern was reverse to the second pattern, the upstream genes were suffered relaxed selection, while the downstream genes were suffered more evolutionary constraints, such as Module 4. Module 4 involved in cell adhesion, cell matrix interaction, extracellular matrix organization and so on.

Our GO enrichment analyses showed a similar pattern, for example, genes involved in purine nucleotide biosynthesis were suffered relaxed selection, but genes associated with purinergic nucleotide receptor signaling pathway were suffered more evolutionary constraints in long-lived mammals.

**CONCLUSIONS**

Our study collected the largest genomic data ever to systematically uncover the universal molecular mechanisms underlying the evolution of mammalian lifespan. We identified 370 NCGs and 300 PCGs which showed convergent shift between protein evolutionary rate and lifespan. Enrichment analyses not only uniquely confirmed many canonical pathways already known in universal regulation of longevity and aging in mammals, but also found many other pathways might also play important roles, such as synaptic transmission, ubiquitin-dependent proteolysis, oxygen utilization, purine metabolisms and purinergic nucleotide receptor signaling pathway and so on. Further our evolutionary analyses suggested that most of PCGs were driven by relaxed selection in long-live mammals, which supported relaxation of evolutionary constraint as a common mechanism to extend lifespan or enhance anti-aging. Finally, our signed functional interaction network analyses of significantly correlated genes with their evolutionary driving forces highlighted several modules and core genes that might play central roles in universally regulate lifespan and aging in mammals. Our study uncovered many important pathways and genes underlying lifespan evolution across mammals. Deepen studies of the molecular mechanisms of these genes and pathways might contribute us to delay human aging and extend human health lifespan.

**ACKNOWLEDGMENTS**

This work was supported by the China Postdoctoral Science Foundation (No. 2019M652951).

**AUTHOR CONTRIBUTIONS**

D.-M.X. designed the project. S.-G.Q. supervised the project. D.-M.X. performed the genomic analyses and wrote the paper.

**Competing interests**: The authors declare no competing interests.

Reference

1 Jin K *et al.* *Aging Dis* **6**:1-5 (2015).

2 Tian X *et al.* *Trends Endocrinol Metab* **28**:722-734 (2017).

3 Sahm A *et al.* *PLoS Genet* **14**:e1007272 (2018).

4 Folgueras AR *et al.* *Circ Res* **123**:905-924 (2018).

5 Austad SN. *J Gerontol A Biol Sci Med Sci* **64**:199-201 (2009).

6 Kowalczyk A *et al.* *Elife* **9**(2020).

7 Ma S *et al.* *Semin Cell Dev Biol* **70**:190-203 (2017).

8 Austad SN. *Exp Gerontol* **36**:599-605 (2001).

9 JE B. Living Longer, in Good Health to the End. The New York Times 25 August 2008:p. D7.

10 de Magalhaes JP *et al.* *Journals of Gerontology Series a-Biological Sciences and Medical Sciences* **62**:149-160 (2007).

11 Nowak RM *et al.* Walker's Mammals of the World: JHU press (1999).

12 Tacutu R *et al.* *Nucleic Acids Res* **46**:D1083-D1090 (2018).

13 Schmidt K *et al.* *BioMetals* **29**:423-432 (2016).

14 Buffenstein R. *J Comp Physiol B* **178**:439-445 (2008).

15 Foley NM *et al.* *Sci Adv* **4**:eaao0926 (2018).

16 Lipman R *et al.* *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **59**:B977-B983 (2004).

17 Muntane G *et al.* *Mol Biol Evol* (2018).

18 Gorbunova V *et al.* *Nat Rev Genet* **15**:531-540 (2014).

19 Seim I *et al.* *Nat Commun* **4**(2013).

20 Kim EB *et al.* *Nature* **479**:223-227 (2011).

21 Lahti DC *et al.* *Trends Ecol Evol* **24**:487-496 (2009).

22 Barbieri M *et al.* *Am J Physiol Endocrinol Metab* **285**:E1064-1071 (2003).

23 van Heemst D. *Aging Dis* **1**:147-157 (2010).

24 Kowalczyk A *et al.* *Bioinformatics* (2019).

25 Partha R *et al.* *Mol Biol Evol* **36**:1817-1830 (2019).

26 Wertheim JO *et al.* *Mol Biol Evol* **32**:820-832 (2015).

27 Scornavacca C *et al.* *Mol Biol Evol* **36**:861-862 (2019).

28 Capella-Gutierrez S *et al.* *Bioinformatics* **25**:1972-1973 (2009).

29 Junier T *et al.* *Bioinformatics* **26**:1669-1670 (2010).

30 Yang ZH. *Mol Biol Evol* **24**:1586-1591 (2007).

31 Zhang J *et al.* *Mol Biol Evol* **22**:2472-2479 (2005).

32 Yu GC *et al.* *Omics-a Journal of Integrative Biology* **16**:284-287 (2012).

33 Supek F *et al.* *PLoS One* **6**:e21800 (2011).

34 Wu G *et al.* *Genome Biol* **11**:R53 (2010).

35 Shannon P *et al.* *Genome Res* **13**:2498-2504 (2003).

36 Shi H *et al.* *Neurobiol Aging* **33**:1849 e1845-1818 (2012).

37 Zeng Y *et al.* *JAMA Netw Open* **1**:e181670 (2018).

38 Yin L *et al.* *mBio* **10**(2019).

39 Reynolds LM *et al.* *Nat Commun* **5**:5366 (2014).

40 Shokhirev MN *et al.* *Aging Cell* **20**:e13280 (2021).

41 Harrison JF *et al.* Moderate hypoxia (10 kPa) extends lifespan with adult but not juvenile exposure in the fruitfly, Drosophila melanogaster. Wiley Online Library 2008.

42 Mehta R *et al.* *Science* **324**:1196-1198 (2009).

43 Li Y *et al.* *Cell Res* **27**:302-305 (2017).

44 Cunningham CC *et al.* *Alcohol Res Health* **27**:291-299 (2003).

45 Ajit D *et al.* *Mol Neurobiol* **49**:1031-1042 (2014).

46 Erb L *et al.* *Biol Cell* **107**:1-21 (2015).

47 Lang S *et al.* *PLoS Genet* **15**:e1008318 (2019).

48 Stenesen D *et al.* *Cell Metab* **17**:101-112 (2013).

49 Tian X *et al.* *Cell* **177**:622-638 e622 (2019).

50 MacInnes AW. *Wiley Interdiscip Rev RNA* **7**:198-212 (2016).

51 Leung CK *et al.* *Free Radic Biol Med* **52**:937-950 (2012).

52 Plotnikov A *et al.* *Biochim Biophys Acta* **1813**:1619-1633 (2011).

53 Yu Z *et al.* *The Innovation* **2**(2021).

54 Kok DEG *et al.* *J Nutr Biochem* **56**:152-164 (2018).

55 Atadja P *et al.* *Exp Cell Res* **217**:205-216 (1995).

56 Fernandez AF *et al.* *Nature* **558**:136-140 (2018).

57 Nopora A *et al.* *J Immunol* **169**:3006-3014 (2002).

58 Crow KD *et al.* *Mol Biol Evol* **23**:887-892 (2006).

59 Yim S *et al.* *BMC Syst Biol* **12**:9 (2018).