

Comparative analysis of molecular basis underlying mammalian lifespan

Dongming Xu^{1,2,3}, Shaogang Qu^{4,1,2,3*}

¹Central Laboratory, Shunde Hospital, Southern Medical University (The First People's Hospital of Shunde Foshan), Foshan, Guangdong, 528300, China

²Guangdong-Hong Kong-Macao Greater Bay Area Center for Brain Science and Brain-Inspired Intelligence, Guangzhou, Guangdong 510515, China.

³Key Laboratory of Mental Health of the Ministry of Education, Southern Medical University, Guangzhou, Guangdong 510515, China.

⁴Department of Neurology, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong 510515, China.

*To whom correspondence should be addressed: Shaogang Qu, Department of Neurology, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong 510515, China.

E-mail: sgq9528@163.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 span 74 mammals. We identified 370 negatively correlated genes and 300 positively correlated genes. Evolutionary analyses suggested that most positively correlated genes were experienced relaxation of evolutionary constraints 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 regulate lifespan and aging, respectively, but also were overrepresented in several categories with adverse or closely functions simultaneously, such as oxygen utilization and response to hypoxia, purine nucleotide biosynthesis and purinergic nucleotide receptor signaling pathway. Finally, our comprehensive network analyses revealed several functional modules with different patterns of evolutionary constraints and highlighted the core genes that might play a central role in regulating mammalian lifespan. Our genomic-wide analyses revealed the common molecular mechanism underlying lifespan determination, which may provide new routes for delaying human aging and extending human health lifespan.

Introduction

Aging and aging-associated diseases present a major threat to human health and are affected by inherently complex process(Jin, et al., 2015). However, the genetic mechanisms of human aging determination are still poorly understood(Sahm, et al., 2018). Most of our knowledge about the genetic mechanisms that govern aging was obtained by studying genetic manipulations of short-lived laboratory animal models(Folgueras, et al., 2018;Tian, et al., 2017). It is unclear if insights from such studies can be transferred to long-lived mammals like humans (Austad, 2009;Kowalczyk, et al., 2020;Ma, et al., 2017).

Fortunately, the rate of aging varies dramatically across wild mammals(Tacutu, et al., 2018). Maximum lifespan (MLS), which can reflect the inherent longevity and “rate of aging” in organisms(JE, 25 August 2008), is positively correlated with their body size(de Magalhaes, et al., 2007). However, many mammals are known to deviate their expected MLS estimated from body

mass(Nowak, et al., 1999). 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 concerning the predicated lifespan based on their body mass, respectively(Tacutu, et al., 2018). On the contrary, the rat, mouse, and shrew can only live about half of their expected MLS even in the well-cared laboratory(Schmidt, et al., 2016). 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 fertility(Buffenstein, 2008), and the longest-lived genus (*Myotis*) maintain the length of their telomeres with aging and does not develop cancer(Foley, et al., 2018), while cancer-related mortality could be up to 90% in short-lived mice(Lipman, et al., 2004;Tian, et al., 2017). Therefore, the repeated changes of MLS in mammals could provide a good opportunity for exploring the molecular mechanisms underlying the regulation of lifespan and aging.

The evolution of protein-coding sequences plays a central role in the regulation of lifespan. For example, DNA repair and p53 signaling pathway-associated genes were often showed positive selection and convergent evolution in long-lived mammals(Gorbunova, et al., 2014;Ma, et al., 2017;Muntane, et al., 2018;Sahm, et al., 2018). However, most of the previous studies only focused on a few long-lived lineages(Kim, et al., 2011;Muntane, et al., 2018;Sahm, et al., 2018;Seim, et al., 2013), thus it was also difficult to differentiate whether the uncovering molecular mechanisms are lineage-specific or common across mammals, or just coincidental neutral changes(Kowalczyk, et al., 2020). Moreover, most genomic-wide comparative studies emphasized the contributions of positive selection or convergent evolution in the regulation of lifespan(Muntane, et al., 2018;Sahm, et al., 2018), little concerned about the roles of relaxation of evolutionary constraints. Although

relaxed selection has been documented as a dominant driving force in the evolution of many traits(Lahti, et al., 2009), and depletion of numerous genes can significantly extend MLS and delay aging in model animals(Barbieri, et al., 2003;van Heemst, 2010).

Recently, a new method that tests for association between relative evolutionary rates of genetic elements and the evolution of traits across a phylogeny (*Reconverge*)(Kowalczyk, et al., 2019;Partha, et al., 2019) have been developed. This method can search for convergent shifts in evolutionary rates of individual protein-coding genes that respond to 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 intensified selection (positive selection)(Kowalczyk, et al., 2019). A previous study that used *RERconverge* methods to uncover the molecular basis underlying lifespan found that negatively correlated genes (NCGs) between protein evolutionary rate and lifespan were enriched in DNA repair, immunity, cell cycle, and cell death-related pathways(Kowalczyk, et al., 2020). However, the authors only used 35 mammals in their main analyses and they did not explore the positively correlated genes (PCGs), the rapid evolution genes in long-lived mammals, and their evolutionary driving forces(Kowalczyk, et al., 2020).

Many rapid evolution genes in long-lived mammals have been verified to affect lifespan or aging, such as genes involved in regulation of energy metabolism (Barbieri, et al., 2003;Seim, et al., 2013;van Heemst, 2010). Rapid evolution genes are always driven by positive selection or relaxed selection and distinguishing them is important for understanding the molecular mechanisms. A

recently developed framework, RELAX, can determine the relaxed selection and intensified selection (positive selection) in protein-coding gene according to the distribution of the nonsynonymous to synonymous substitution rates (ω or dN/dS) in the codon sites(Wertheim, et al., 2015). Moreover, lifespan is affected by complex genetic factors(Jin, et al., 2015;Tian, et al., 2017) and the ubiquity of pleiotropy, the rapid and slow evolutionary genes might interact with each other to regulate lifespan in mammals.

Here, we integrated the lifespan phenotypes from the HAGR (Human Ageing Genomic Resources) database(Tacutu, et al., 2018) and the orthologous genes from the OrthoMaM database(Scornavacca, et al., 2019) to obtain the largest genomic data of protein-coding genes ever used to reveal the common molecular basis underlying the regulation of mammalian lifespan. We genomic-wide identified the significantly correlated genes between protein evolutionary rate and lifespan across mammals. Further, we determined their evolutionary driving forces and explored the extend to which relaxed selection contributed to extend lifespan in mammals, especially for the positively correlated genes. Finally, we combined analyses of the PCGs and NCGs, and with their evolutionary driving forces to provide more insights into the molecular mechanisms underlying lifespan regulation in mammals. Our study will deepen our knowledge of the molecular mechanisms underlying the 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 alignments of one-to-one orthologous genes were downloaded from the OrthoMaM database (V10b)(Scornavacca, et al., 2019), which includes 14509 genes from 116 species. The observed maximum lifespan and adult weight were extracted from the Human Aging Genomic Resource (HAGR) database(Tacutu, et al., 2018), with 999 mammals having both traits. Expected maximum lifespan was calculated according to their adult weight using the formula: *expect lifespan (years)*= $3.34 * (\text{adult weight (g)}^{0.193})$ (de Magalhaes, et al., 2007), 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 identification of one-to-one orthologous and multiple sequence alignment are still full challenges. Removing the fake one-to-one orthologous sequences and poorly aligned regions will promote the reliability of the following comparative and evolutionary analyses(Capella-Gutierrez, et al., 2009;Natsidis, et al., 2021). To exclude the fake orthologous sequences in the one-to-one orthologous genes, we first 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 protein references using the *blastp* method, respectively. If the best hit of a

sequence is different from that of the human sequence in the same one-to-one orthologous gene when searched for the human protein reference or is different from that of the mouse sequence when searched for the mouse protein reference, we removed the 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*(Capella-Gutierrez, et al., 2009) to filter the original alignments with parameter “-resoverlap 0.70 -seqoverlap 50-automated1 -colnumbering”. Finally, only orthologous alignments 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 the 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)(Scornavacca, et al., 2019) (**Figure 1a**). The *nw_prune* module in the *newick-utils* (v1.6) package(Junier, et al., 2010) 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) packages(Yang, 2007) was used to calculate the branch lengths with the Empirical + F model. Then, we used R package *RERconverge* (v0.1.0)(Kowalczyk, et al., 2019) 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 calculates 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 the 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, indicates 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 values as positively correlated genes (PCGs), and the significantly correlated genes with negative Rho values as negatively correlated genes (NCGs) (**Supplementary Table 3-4**).

Determination of evolutionary forces

The coding sequence alignments of 74 mammals in this study were downloaded and extracted from the OrthoMaM database (v10b)(Scornavacca, et al., 2019). We removed the non-orthologous sequences and poorly aligned regions that were 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 the top 10 LQ values in the 74 mammals as long-lived mammals in this study. The rest species were treated as background groups. The *Hyphy* package developed by RELAX framework(Wertheim, et al., 2015) 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 the alternative hypothesis is better than the null hypothesis. The null hypothesis indicates no different evolutionary selection between foreground group and background group. A 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 model(Zhang, et al., 2005) to detect positive selection among all significantly correlated genes for long-lived species. The species with the top 10 LQ values were label as foreground. A 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 package(Yu, et al., 2012) to search for the significantly correlated genes' overrepresented GO Biological Process (GO BP) terms. The PCGs and the NCGs were used as the foreground genes, respectively. All the one-to-one orthologous genes were considered as the 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**. To further explore the uniquely enriched categories by PCGs and NCGs, we also searched the relatively overrepresented GO BP terms in the PCGs when compared with the NCGs, and reverse (**Supplementary Table 11-12**). Fisher's single-tailed test with a *P value* at 0.05 was used as the threshold for statistical significance. In addition, we used REVIGO(Supek, et al., 2011) to remove the redundantly significant GO BP terms with a 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)(Wu, et al., 2010). 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 using the *ReactomeFIViz* module in software *Cytoscape* (v3.8.0)(Shannon, et al., 2003). Then, we manually removed the predicted interactions and added correlation and evolutionary selection information to construct a comprehensive 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 observed maximum lifespan and adult weight of mammals were extracted from the HAGR database (Tacutu, et al., 2018). 999 mammals include both traits. The expected maximum lifespans were calculated according to their adult weights based on a previous study (de Magalhaes, et al., 2007) (**seen methods**). The Longevity quotients (LQ) is considered as the ratio of the observed maximum lifespan to the expected maximum lifespan (Muntane, et al., 2018). The original protein alignments of one-to-one orthologous genes were downloaded from the OrthoMaM database (v10b) (Scornavacca, et al., 2019), including 14509 genes span 116 mammals. Finally, 74 mammals included genomic sequences and lifespan phenotypes were used in the following analyses (**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 dependent on the quality of the orthologous identification and the alignment accuracy(Capella-Gutierrez, et al., 2009). We removed 7838 artificial one-to-one orthologous sequences from 1169 genes through the *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 annotation(Capella-Gutierrez, et al., 2009). We dropped the orthologous genes with less than 50 species or alignment lengths shorter than 50 amino acids. Finally, 13381 one-to-one orthologous genes 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 the RERconverge method, we identified 370 negatively correlated genes (NCGs) (**Supplementary Table 3**) and 300 positively correlated genes (PCGs) (**Supplementary Table 4**). These genes are not enriched in longevity genes collected by the HAGR database(Tacutu, et al., 2018) (**Supplementary Table 5**). However, this may be caused by most of the longevity-associated genes in the HAGR database were obtained from the human population or the short-lived laboratory animal models(Tacutu, et al., 2018), and thus was not a good representative of the common molecular basis underlying lifespan evolution. For example, 5 of the six most significantly correlated genes might affect lifespan or aging (**Figure 2**), none of them was collected by the HAGR database. Such as a minor SNP in *ADAMTS6* (ADAM metalloproteinase with thrombospondin type 1 motif, 6), the most positively correlated gene (**Figure 2a**), has a protective effect on aging in a GWAS study(Shi, et al., 2012). A locus (rs12199884) in *PKHD1* (polycystic kidney and hepatic

disease 1), the most negatively correlated gene (**Figure 2d**), is negatively associated with longevity in males from the Han Chinese population GWAS study(Zeng, et al., 2018). Deletion of *YbeY* (YbeY Metalloendoribonuclease), the second most negatively correlated gene (**Figure 2e**), will shorten the bacteria longevity(Yin, et al., 2019). *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 mice(Tian, et al., 2017). *SLC11A1* (Solute Carrier Family 11 Member 1), the third most negatively correlated gene (**Figure 2f**), is different expression and methylation in multiple human aging tissues(Reynolds, et al., 2014;Shokhirev, et al., 2021). These results suggested that our significantly correlated genes are highly associated with regulating lifespan or anti-aging across mammals and supported that the *RERconverge* correlative analysis is an effective way to reveal the genetic mechanisms underlying trait evolution.

Determination of the 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 the PCGs. We 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**). The branch-site model also found the ratio of positively selected genes in PCGs was significantly less than that of the 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 three times higher than that of

non-significantly correlated genes ($P=3.03E-11$) and was over six times more than that of the NCGs ($P=1.84E-52$) (**Figure 3b**). These analyses suggested that relaxation of evolutionary constraints is an important force that driving the extending of lifespan in mammals.

Enrichment analyses

Biological Process of Gene Ontology (GO BP) enrichment analyses showed that the PCGs and the NCGs were uniquely enriched in many canonical pathways regulating 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 study(Kowalczyk, et al., 2020). 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 for the 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**). Not surprisingly, these categories enriched by the PCGs were driven by relaxed selection in long-lived mammals, little was driven by positive selection (**Supplementary Table 10**).

Moreover, we further found several categories with adverse or closely relative functions that were

enriched by the PCGs and the NCGs simultaneously. If categories with adverse functions were enriched by the NCGs and the PCGs simultaneously, it was about to double confirm that these functional categories might affect longevity or aging. If categories with closely relative functions were enriched by the NCGs and the PCGs simultaneously, it might reflect the pleiotropy of these genes or categories. Our results showed that categories associated with oxygen utilization and response to hypoxia were uniquely enriched by PCGs and NCGs simultaneously. 7 PCGs, including 4 relaxed selected genes (*ACTN3/SUCLG2/ACO2/OGDHL*), were overrepresented in the category “aerobic respiration” ($P=2.80E-04$) (**Supplementary Table 10**), which was significantly more than that of the NCGs (0 genes, $P=0.0062$) (**Supplementary Table 12 and Figure 5a**). In adverse, 4 NCGs (*CYBB/PSMB11/STOX1/BCL2*) were involved in the 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 longevity. For example, hypoxia can extend the lifespan in fruit fly(Harrison, et al., 2008) and nematode(Mehta, et al., 2009). Also, elderly people living in hypoxia in the Tibetan Plateau tend to have a longer life than similarly aged people in other Chinese regions(Li, et al., 2017). In addition, 5 NCGs (*CYBB/RECQL5/BLM/ADCY6/SPIDR*) were enriched in the category “cellular response to alcohol” ($P=0.0096$) (**Supplementary Table 9 and Figure 5a**), while no PCGs 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 death(Cunningham, et al., 2003). These results suggested that enhancing hypoxia resistance and weakening the efficiency of 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 the PCGs and the NCGs simultaneously. For example, 4 NCGs (*P2RY2/P2RY10/GPR171/P2RY14*) were enriched in the category “G-protein coupled purinergic nucleotide receptor signaling pathway” ($P=2.38E-04$) (**Supplementary Table 9**), while no PCGs was involved in this pathway ($P=0.0929$) (**Figure 5b**). Interestingly, all these four genes belong to the P2Y receptors family. Although we know little about the relationship between P2Y receptors and lifespan, P2Y receptors affect aging-related diseases, such as Alzheimer’s disease(Ajit, et al., 2014;Erb, et al., 2015). Surprisingly, we found that the PCGs were enriched in many other purine biosynthetic-associated categories. Such as 7 PCGs were involved in the category “purine ribonucleoside diphosphate metabolic process” ($P=0.0076$) (**Supplementary Table 10**), which was significantly more than that of the NCGs (1 gene, $P=0.0167$) (**Supplementary Table 12 and Figure 5b**). In addition, the PCGs were also enriched 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 the addition of dietary adenine could shorten the lifespan by 48%(Lang, et al., 2019) and blocks the longevity effects of dietary restriction(Stenesen, et al., 2013). These results suggested that genes involved in purine biosynthetic and purine signaling transduction might suffer different evolutionary constraints in determining the lifespan or anti-aging in mammals.

Network analyses

To further explore the extent to which NCGs and PCGs interacted with each other to affect lifespan, we constructed a comprehensive network using all significantly correlated genes with additional

information of their evolutionary driving forces (**Figure 6**). From the network, we can identify at least five modules with different patterns of evolutionary constraints (**Figure 6 Module1-5**). For example, for Module2 and Module 3, which are involved in DNA repair and ribosome biogenesis, respectively. Almost all genes in these modules were experienced more evolutionary constraints in long-lived mammals. Enhancing the capability of DNA repair(Tian, et al., 2019) and slowing down the ribosome turnover(MacInnes, 2016) are two fundamental mechanisms for extending lifespan across eukaryotic speciesa(MacInnes, 2016).WD Repeat-Containing Protein 46 (*WDR46*) might be the hub gene in Module 2. All genes in this module activated or catalyzed *WDR46* (**Figure 6 Module 2**). A previous study showed that depletion of *wdr-46* reduced the lifespan in nematode(Leung, et al., 2012). 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 *PRIMI*. Thus, these three genes may be the critical mediators between the capability of DNA repair and lifespan in mammals (**Figure 6 Module 3**).

The functions of Module 1 are 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 cell proliferation and differentiation(Plotnikov, et al., 2011). Here, we found that most 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 recent study(Yu,

et al., 2021). Moreover, several GO categories that were regulated by insulin-associated signaling pathways were also overrepresented by the relaxed selected PCGs, such as “positive regulation of glucose import” and “positive regulation of growth” (**Figure 4b and Supplementary Table 10**). However, we unexpectedly found five genes in the most upstream of the pathways that were consistently experienced increasing evolutionary constraints in long-lived mammals, especially three of them (*MUC20/EPHB2/SYNJI*) were experienced positive selection because the positively selected gene were relatively rare either in our network or in our significantly correlated genes (**Figure 3a and Supplementary Figure 3**). These genes might also affect lifespan or aging. For example, mucin 20 (*MUC20*), a negative regulator of MAPK cascade and RTKs, was one of the most prominent calorie restriction (CR)-associated genes in mice (Kok, et al., 2018). The pattern of evolutionary constraint in Module 5 was a little bit like Module 1. Module 5 appeared to be a protein complex and mainly involved in the cell cycle and cell division. Most of the genes in this module were very conservative in long-lived mammals. However, Nucleoporin 107 (*NUP107*), the only gene in this module that directly regulated genes outside the module, was a positively correlated gene and suffered relaxed selection. *NUP107* may combine other proteins, such as *NUP88* and *NUP155*, to activate *CPSF3* and *NCBP1* (**Figure 6 Module 5**) and thus control cell cycle and cycle division.

The functions of Module 4 are mainly involved in cell adhesion, cell-matrix interaction, extracellular matrix organization. Studies in both *Caenorhabditis elegans* and *Drosophila melanogaster* have revealed that loss-of-function mutations and knockdown of genes encoding components of the integrin-signaling complex can extend lifespan and improve anti-aging

performance(De Luca, 2019). Here, we found that most of the regulator genes upstream of Module 4 were experienced rapid evolution and driven by relaxed selection in long-lived mammals. However, the integrin subunit beta 7 (*ITGB7*), the core gene in this module, was experienced increased evolutionary constraint in long-lived mammals (**Figure 6 Module 4**). *ITGB7* was an integrin receptor that functions in signaling from the extracellular matrix to the cell(Erle, et al., 1991). Although we know little information about the biological implication of *ITGB7*, it is worth exploring the effects of *ITGB7* in the regulation of lifespan in the mammal.

Besides the five modules above, other genes might also be involved in regulating lifespan or aging in our network. For example, cyclin D1 (*CCND1*), a positively correlated gene, was targeted by the largest number of members in our comprehensive network (**Figure 6**). Overexpression of *CCND1* will lead to cell senescence(Atadja, et al., 1995). Another most targeted gene was BCL2 apoptosis regulator (*BCL2*), a negatively correlated gene that was regulated by four genes. *BCL2* is a negative regulator of autophagy, disruption of the beclin 1-Bcl2 autophagy regulatory complex promotes longevity in mice(Fernandez, et al., 2018), and down-regulation of Bcl-2 expression controls murine dendritic cell longevity(Nopora, et al., 2002).

DISCUSSION

Pearson correlation in the *RERconverge* package was used to determine whether a gene significantly correlates between protein evolutionary rate and lifespan across the phylogeny. The reliability and sensitivity of the correlative analysis largely depend on the sample size. Here, we collected 74 mammals, which means 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 the fewest species (50 mammals) also include 97 pairs of data. However, the previous study used 34 mammals as the maximum in their main study. Although the authors also analyzed 61 mammals in their additional analyses, they did not limit their analyses on one-to-one orthologous genes, while gene duplications will affect its evolutionary constraints and evolutionary rates (Crow, et al., 2006). In addition, large-scale identification of one-to-one orthologous and multiple sequence alignment are still full challenges so far. Here, we removed the fake one-to-one orthologous genes and dropped the poorly aligned region, which will further promote the accuracy of calculating the branch lengths. Thus, our analyses might be more reliable and sensitive in identifying significantly correlated genes than the previous study.

Our evolutionary selection analyses supported that relaxed selection plays an essential role in the regulation of lifespan in mammals. We found 62.7% of the PCGs were suffered relaxation of evolutionary constraints in long-lived mammals, which was much more than only 1% of the PCGs were suffered positive selection. However, we could not arbitrarily declare that relaxed selection plays more important roles than that of positive selection in the determination of lifespan in mammals because the period of positive selection usually was very short, and the signals of positive selection can quickly be overwhelmed over time (Yang, 2007).

Our GO enrichment analyses found the NCGs and the PCGs were uniquely enriched in many canonical pathways known to regulate lifespan or aging (**Figure 4**). These categories were not only significantly overrepresented when comparing to the background genes, but also were significantly

overrepresented when comparing to their adverse correlated genes. This further supported different functional categories-associated genes were consistently suffered different evolutionary constraints in long-lived mammals, and both the NCGs and the PCGs play an essential role in the regulating lifespan. For example, both our study and many previous studies found the NCGs were enriched in DNA repair pathways. However, our study further found that the number of NCGs involved in DNA repair pathways was also much more than that of the PCGs. Almost no genes involved in DNA repair pathways were experienced rapid evolution in long-lived mammals (**Supplementary Table 11**). As well, the number of PCGs involved in many categories, such as positive regulation of glucose import, ionotropic glutamate receptor signaling pathway, positive regulation of ubiquitin-dependent protein catabolic process, was higher than expected and significantly more than that of the NCGs (**Supplementary Table 12**).

Our comprehensive network analyses of the NCGs and PCGs showed the five modules have three different patterns of evolutionary constraints. The first pattern was almost all genes in the modules were suffered more evolutionary constraints in long-lived mammals, such as Module 1 and Module 4, which mainly involved maintaining genome stability. The second pattern was the genes upstream of signaling pathways were suffered more evolutionary constraints in long-lived mammals while the genes downstream of the pathways were suffered relaxation of evolutionary constraints, such as Module 1 and Module 5, which mainly involved in cell division and cell proliferation. For the signaling pathways that control cell division and proliferation, the upstream genes may primarily function as a signal to initiate cell division or proliferation. In contrast, the downstream genes may function as directly inducing cell division or proliferation. Long-lived

species usually undergo more cell division and cell proliferation in their lifetime, and more cell division will increase the risk of developing cancer. However, many long-lived mammals do not have a higher risk of developing cancer. This is called Peto's Paradox (Tollis, et al., 2017). Here, we found that genes responded for maintaining genomic stability and regulating the signals that initial cell division or proliferation process were suffered more evolutionary constraints in long-lived mammals. In contrast, genes directly induced cell division or proliferation were suffered relaxed selection. This may partly explain Peto's Paradox.

The third pattern was contrary to the second pattern, the upstream genes were suffered relaxed selection in the long-lived mammals, while the downstream genes were suffered more evolutionary constraints, such as Module 4, which involved extracellular matrix organization and cell-matrix interaction. Interesting, several overrepresented GO categories had similar evolutionary constraints. Such as the process responsible for purine nucleotide metabolic and biosynthetic was enriched by the PCGs which were driven by relaxed selection (**Supplementary Table 10 and Figure 5b**), while the process responsible for purinergic nucleotide receptors signaling transduction was enriched by the NCGs (**Supplementary Table 9 and Figure 5b**). This may be caused by gene pleiotropy. Both maintaining their functional conservation and decreasing their activation of these pathways are essential to extend lifespan in mammals. For example, heterozygous purine nucleotide biosynthesis mutation can extend lifespan in model animals (De Luca, 2019; Stenesen, et al., 2013), while complete depletion of them will cause lethality in the early development (Holland, et al., 2011).

In addition, comparing to ordinary protein-protein interaction network, our comprehensive network

analyses includes the information of directions and signs, which will make the causality of the network prominent and highlight the core genes in the modules(Yim, et al., 2018). For example, it will be difficult to distinguish the hub genes in Module 2 and Module 4 in the ordinary protein-protein interaction network. However, with the additional information of directions and signs, we can clearly identify *WDR46* and *ITGB7* as the core gene in Module 2 and Module 4, respectively.

CONCLUSIONS

Our study collected the largest genomic data ever to systematically uncover the common molecular mechanisms underlying the evolution of mammalian lifespan. We identified 370 NCGs and 300 PCGs which showed a convergent shift between protein evolutionary rate and lifespan. Further, our evolutionary analyses suggested that the relaxation of evolutionary constraint is a major force to extend lifespan or enhance anti-aging in mammals. Enrichment analyses not only confirmed many canonical pathways in the regulation of longevity or aging but also found many other pathways might also play an essential roles, such as oxygen utilization and response to hypoxia, purine nucleotide biosynthesis, and purinergic nucleotide receptor signaling pathway. Finally, our comprehensive network analyses uncover several modules with different patterns of evolutionary constraints and highlighted core genes that might play central roles in commonly regulate lifespan and aging in mammals. Our study revealed many vital pathways and genes underlying lifespan evolution across mammals. Deepen studies of the molecular mechanisms of these genes and pathways might contribute to extending 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.

References

- Ajit D, Woods LT, Camden JM *et al.* 2014. Loss of p2y(2) nucleotide receptors enhances early pathology in the tgrnd8 mouse model of alzheimer's disease. *Mol Neurobiol*, **49**(2):1031-1042.
- Atadja P, Wong H, Veillette C *et al.* 1995. Overexpression of cyclin d1 blocks proliferation of normal diploid fibroblasts. *Exp Cell Res*, **217**(2):205-216.
- Austad SN. 2009. Comparative biology of aging. *J Gerontol A Biol Sci Med Sci*, **64**(2):199-201.
- Barbieri M, Bonafe M, Franceschi C *et al.* 2003. Insulin/igf-i-signaling pathway: An evolutionarily conserved mechanism of longevity from yeast to humans. *Am J Physiol Endocrinol Metab*, **285**(5):E1064-1071.
- Buffenstein R. 2008. Negligible senescence in the longest living rodent, the naked mole-rat: Insights from a successfully aging species. *J Comp Physiol B*, **178**(4):439-445.
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. Trimal: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**(15):1972-1973.
- Crow KD, Wagner GP, Investigators ST-NY. 2006. What is the role of genome duplication in the evolution of complexity and diversity? *Mol Biol Evol*, **23**(5):887-892.
- Cunningham CC, Van Horn CG. 2003. Energy availability and alcohol-related liver pathology. *Alcohol Res Health*, **27**(4):291-299.
- De Luca M. 2019. The role of the cell-matrix interface in aging and its interaction with the renin-angiotensin system in the aged vasculature. *Mech Ageing Dev*, **177**:66-73.
- de Magalhaes JP, Costa J, Church GM. 2007. An analysis of the relationship between metabolism, developmental schedules, and longevity using phylogenetic independent contrasts. *J Gerontol A Biol Sci Med Sci*, **62**(2):149-160.
- Erb L, Cao C, Ajit D *et al.* 2015. P2y receptors in alzheimer's disease. *Biol Cell*, **107**(1):1-21.
- Erle DJ, Rüegg C, Sheppard D *et al.* 1991. Complete amino acid sequence of an integrin beta subunit (beta 7) identified in leukocytes. *J Biol Chem*, **266**(17):11009-11016.
- Fernandez AF, Sebt S, Wei Y *et al.* 2018. Disruption of the beclin 1-bcl2 autophagy regulatory complex promotes longevity in mice. *Nature*, **558**(7708):136-140.

Foley NM, Hughes GM, Huang Z *et al.* 2018. Growing old, yet staying young: The role of telomeres in bats' exceptional longevity. *Sci Adv*, **4**(2):eaao0926.

Folgueras AR, Freitas-Rodriguez S, Velasco G *et al.* 2018. Mouse models to disentangle the hallmarks of human aging. *Circ Res*, **123**(7):905-924.

Gorbunova V, Seluanov A, Zhang Z *et al.* 2014. Comparative genetics of longevity and cancer: Insights from long-lived rodents. *Nat Rev Genet*, **15**(8):531-540.

Harrison JF, Rascon B. 2008. Moderate hypoxia (10 kpa) extends lifespan with adult but not juvenile exposure in the fruitfly, *drosophila melanogaster*. In.: Wiley Online Library.

Holland C, Lipsett DB, Clark DV. 2011. A link between impaired purine nucleotide synthesis and apoptosis in *drosophila melanogaster*. *Genetics*, **188**(2):359-367.

JE B. 25 August 2008. Living longer, in good health to the end. In. The New York Times; p. D7.

Jin K, Simpkins JW, Ji X *et al.* 2015. The critical need to promote research of aging and aging-related diseases to improve health and longevity of the elderly population. *Aging Dis*, **6**(1):1-5.

Junier T, Zdobnov EM. 2010. The newick utilities: High-throughput phylogenetic tree processing in the unix shell. *Bioinformatics*, **26**(13):1669-1670.

Kim EB, Fang XD, Fushan AA *et al.* 2011. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature*, **479**(7372):223-227.

Kok DEG, Rusli F, van der Lugt B *et al.* 2018. Lifelong calorie restriction affects indicators of colonic health in aging c57bl/6j mice. *J Nutr Biochem*, **56**:152-164.

Kowalczyk A, Meyer WK, Partha R *et al.* 2019. Rerconverge: An r package for associating evolutionary rates with convergent traits. *Bioinformatics*.

Kowalczyk A, Partha R, Clark NL *et al.* 2020. Pan-mammalian analysis of molecular constraints underlying extended lifespan. *Elife*, **9**.

Lahti DC, Johnson NA, Ajie BC *et al.* 2009. Relaxed selection in the wild. *Trends Ecol Evol*, **24**(9):487-496.

Lang S, Hilsabeck TA, Wilson KA *et al.* 2019. A conserved role of the insulin-like signaling pathway in diet-dependent uric acid pathologies in *drosophila melanogaster*. *PLoS Genet*, **15**(8):e1008318.

Leung CK, Empinado H, Choe KP. 2012. Depletion of a nucleolar protein activates xenobiotic detoxification genes in *caenorhabditis elegans* via *nrf-1* and *p53/cep-1*. *Free Radic Biol Med*, **52**(5):937-950.

Li Y, Wang MS, Otecko NO *et al.* 2017. Hypoxia potentially promotes tibetan longevity. *Cell Res*, **27**(2):302-305.

Lipman R, Galecki A, Burke DT *et al.* 2004. Genetic loci that influence cause of death in a heterogeneous mouse stock. *J Gerontol A Biol Sci Med Sci*, **59**(10):B977-B983.

Ma S, Gladyshev VN. 2017. Molecular signatures of longevity: Insights from cross-species comparative studies. *Semin Cell Dev Biol*, **70**:190-203.

MacInnes AW. 2016. The role of the ribosome in the regulation of longevity and lifespan extension. *Wiley Interdiscip Rev RNA*, **7**(2):198-212.

Mehta R, Steinkraus KA, Sutphin GL *et al.* 2009. Proteasomal regulation of the hypoxic response modulates aging in *c. Elegans*. *Science*, **324**(5931):1196-1198.

Muntane G, Farre X, Antonio Rodriguez J *et al.* 2018. Biological processes modulating longevity across primates: A phylogenetic genome-phenome analysis. *Mol Biol Evol*.

Natsidis P, Kapli P, Schiffer PH *et al.* 2021. Systematic errors in orthology inference and their effects on evolutionary analyses. *iScience*, **24**(2):102110.

Nopora A, Brocker T. 2002. Bcl-2 controls dendritic cell longevity in vivo. *J Immunol*, **169**(6):3006-3014.

Nowak RM, Walker EP. 1999. Walker's mammals of the world, vol. 1: JHU press.

Partha R, Kowalczyk A, Clark NL *et al.* 2019. Robust method for detecting convergent shifts in evolutionary rates. *Mol Biol Evol*, **36**(8):1817-1830.

Plotnikov A, Zehorai E, Procaccia S *et al.* 2011. The mapk cascades: Signaling components, nuclear roles and mechanisms of nuclear translocation. *Biochim Biophys Acta*, **1813**(9):1619-1633.

Reynolds LM, Taylor JR, Ding J *et al.* 2014. Age-related variations in the methylome associated with gene expression in human monocytes and t cells. *Nat Commun*, **5**:5366.

Sahm A, Bens M, Szafranski K *et al.* 2018. Long-lived rodents reveal signatures of positive selection in genes associated with lifespan. *PLoS Genet*, **14**(3):e1007272.

Schmidt K, Steiner K, Petrov B *et al.* 2016. Short-lived mammals (shrew, mouse) have a less robust metal-responsive transcription factor than humans and bats. *BioMetals*, **29**(3):423-432.

Scornavacca C, Belkhir K, Lopez J *et al.* 2019. Orthomam v10: Scaling-up orthologous coding sequence and exon alignments with more than one hundred mammalian genomes. *Mol Biol Evol*, **36**(4):861-862.

Seim I, Fang XD, Xiong ZQ *et al.* 2013. Genome analysis reveals insights into physiology and longevity of the brandt's bat myotis brandtii. *Nat Commun*, **4**.

Shannon P, Markiel A, Ozier O *et al.* 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res*, **13**(11):2498-2504.

Shi H, Belbin O, Medway C *et al.* 2012. Genetic variants influencing human aging from late-onset alzheimer's disease (load) genome-wide association studies (gwas). *Neurobiol Aging*, **33**(8):1849 e1845-1818.

Shokhirev MN, Johnson AA. 2021. Modeling the human aging transcriptome across tissues, health status, and sex. *Aging Cell*, **20**(1):e13280.

Stenesen D, Suh JM, Seo J *et al.* 2013. Adenosine nucleotide biosynthesis and ampk regulate adult life span and mediate the longevity benefit of caloric restriction in flies. *Cell Metab*, **17**(1):101-112.

Supek F, Bosnjak M, Skunca N *et al.* 2011. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One*, **6**(7):e21800.

Tacutu R, Thornton D, Johnson E *et al.* 2018. Human ageing genomic resources: New and updated databases. *Nucleic Acids Res*, **46**(D1):D1083-D1090.

Tian X, Firsanov D, Zhang Z *et al.* 2019. Sirt6 is responsible for more efficient DNA double-strand break repair in long-lived species. *Cell*, **177**(3):622-638 e622.

Tian X, Seluanov A, Gorbunova V. 2017. Molecular mechanisms determining lifespan in short- and long-lived species. *Trends Endocrinol Metab*, **28**(10):722-734.

Tollis M, Boddy AM, Maley CC. 2017. Peto's paradox: How has evolution solved the problem of cancer prevention? *BMC Biol*, **15**(1):60.

van Heemst D. 2010. Insulin, igf-1 and longevity. *Aging Dis*, **1**(2):147-157.

Wertheim JO, Murrell B, Smith MD *et al.* 2015. Relax: Detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol*, **32**(3):820-832.

Wu G, Feng X, Stein L. 2010. A human functional protein interaction network and its application to cancer data analysis. *Genome Biol*, **11**(5):R53.

Yang ZH. 2007. Paml 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol*, **24**(8):1586-1591.

Yim S, Yu H, Jang D *et al.* 2018. Annotating activation/inhibition relationships to protein-protein interactions using gene ontology relations. *BMC Syst Biol*, **12**(Suppl 1):9.

- Yin L, Ma H, Nakayasu ES *et al.* 2019. Bacterial longevity requires protein synthesis and a stringent response. *mBio*, **10**(5).
- Yu GC, Wang LG, Han YY *et al.* 2012. Clusterprofiler: An r package for comparing biological themes among gene clusters. *Omics-a Journal of Integrative Biology*, **16**(5):284-287.
- Yu Z, Seim I, Yin M *et al.* 2021. Comparative analyses of aging-related genes in long-lived mammals provide insights into natural longevity. *The Innovation*, **2**(2).
- Zeng Y, Nie C, Min J *et al.* 2018. Sex differences in genetic associations with longevity. *JAMA Netw Open*, **1**(4):e181670.
- Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*, **22**(12):2472-2479.