Evolution of p53 pathway-related genes provides insights into anticancer mechanisms of natural longevity in cetaceans

Xing Liu
Nanjing Normal University
Zhenpeng Yu
Nanjing Normal University
Fei Yang
Nanjing Normal University
Xin Huang
Nanjing Normal University
Linxia Sun
Nanjing Normal University
Wenhua Ren
Nanjing Normal University
Guang Yang
Nanjing Normal University
Shixia Xu (xushixia@njnu.edu.cn)
Nanjing Normal University

Research Article

Keywords: Cetaceans, Longevity, p53 pathway, Molecular evolution, Anticancer, Peto's paradox

Posted Date: June 15th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1727364/v1

Abstract

Background: Large, long-lived organisms typically have a higher risk of cancer, but cetaceans, which include the largest and longest-living mammals, may have evolved a mechanism to counter this. However, little is known about the genetic basis underlying this mechanism.

Cetaceans, which include the largest and longest-lived mammals, may have evolved effective anticancer mechanisms in response to a high risk of cancer caused by massive somatic cell divisions.

Results: The p53 pathway is ideal for studying the mechanisms behind cancer resistance, as nearly all types of cancer have evolved ways to evade this suppressive mechanism. Here, comparative genetic analyses of the 73 genes involved in the p53 pathway were examined in cetaceans to explore potential anticancer mechanisms behind natural longevity. We found that long-lived species have three positively selected genes (*APAF1*, *CASP8*, *TP73*) and three duplicated genes (*IGFBP3*, *PERP*, *CASP3*) involved in the regulation of apoptosis, suggesting that apoptosis regulation may be an important strategy for cancer resistance in these long-lived cetaceans. Inhibition of angiogenesis may be another strategy for protecting cetaceans from cancer because the evolutionary rates of three angiogenesis-regulated genes (*SERPINE1*, *CD82* and *TSC2*) were significantly related to longevity traits (longevity quotient and maximum lifespan). Interestingly, some tumor suppressor genes were determined to be under positive selection (*APAF1* and *TP73*), have high copy numbers (*CASP3* and *CCNG1*), and be related to body size in the large-bodied and long-lived cetacean lineages. The above is all molecular evidence for Peto's paradox.

Conclusions: Our study identified several candidate genes for the p53 pathway with different evolutionary paths in the large, long-lived cetaceans and suggested that tumor suppression in cetaceans might be mediated through multiple synergistic effects rather than a single pathway/target.

Background

Lifespan is a characteristic of life history [1]. Extant mammals show a wide variety of lifespans and body masses (BM), with the shortest- and longest-living mammals differing by more than 100-fold and smallest and largest mammals differing by ~100-million-fold [2, 3]. The cetaceans contain extremes in both categories, as this group includes the largest and longest-lived lineages of mammals. The largest living mammal on earth is the blue whale (*Balaenoptera musculus*), which has an average reported adult weight of 136,000 kg and maximum lifespan (MLS) of 110 years [4]. The longest living mammal is the bowhead whale (*Balaena mysticetus*), weighing more than 100,000 kg and living up to 211 years [5]. Generally, large, long-lived organisms have an increased risk of developing cancer because big bodies with more cells undergo more cell divisions, increasing the probability that a) DNA damage will occur and b) a normal cell will become a cancerous one [6, 7]. However, large whales—which have 1,000 times more cells than humans—have a lower cancer risk and longer lifespan than might be expected based on this pattern [8]; this is consistent with Peto’s paradox, the observation that there is no correlation between cancer incidence and body size or longevity across species [9]. For example, only two benign tumors were

identified in 2,000 baleen whales shunted in South Africa [10] and few cancers were identified in wild toothed whales that were killed by hunters or died of natural causes [8]. Thus, it is reasonable to suggest that these large, long-lived cetaceans evolved an effective mechanism for suppressing cancer.

Recent studies have sequenced the genomes of several extremely large and long-lived species of cetaceans and identified the genetic basis by which they resist cancer [11-13]. Comparative genomic studies of the long-lived bowhead whales and humpback whales (*Megaptera novaeangliae*, MLS: 95 years) found that a number of genes related to cancer and aging are under positive selection (e.g. *SOCS2*, *APTX*, *PRDM1* & *PRDM2*, and *WDHD1*), have undergone gene duplication (e.g. *LAMTOR1*), and have lineage-specific amino acid changes (e.g. *PRDM13*, *ERCC1*) [11, 12]. Moreover, transcriptomic analyses of the gray whale (*Eschrichtius robustus*)—which can live up to 77 years—revealed that pathways related to DNA repair, autophagy, and ubiquitination appear to have higher activity that other mammals, which is crucial for preventing cell transformation and eliminating cancer cells [13]. A handful of candidate genes and pathways related to lifespan extension were identified in single-species studies, but the broader mechanisms underlying lifespan regulation in cetaceans remain poorly explored.

The p53 pathway has received considerable attention for its essential roles in tumor-suppression and extending lifespan; in addition, nearly all types of cancer have evolved ways to evade this suppressive mechanism. The p53 pathway mainly exerts its tumor-suppressing function through the well-known tumor suppressor p53. Numerous studies have shown that loss or mutation of p53 results in over half of all human cancers, and spontaneous tumorigenesis was shown to occur in p53 knockout mice [14]. As a transcription factor, p53 can prevent tumor formation and development through selective transcriptional regulation of many target genes to induce cell cycle arrest (e.g. *SFN*, *CDK1*, *CDK4*), promote cell apoptosis or senescence (e.g. *EI24*, *APAF1*, *FAS*), and accelerate DNA repair (e.g. *CHEK1*, *CHEK2*, *TP53*) [15]. For example, comparative genomic studies identified multiple copies of the tumor suppressor gene *TP53* in the long-lived elephants [16], and *in vitro* experiments further found that the expansion of the *TP53* enhanced the DNA damage response and augmented apoptosis, thus decreasing cancer incidence and increasing longevity [17]. In addition, p53 negatively regulates the IGF1-mTOR pathway by inducing the expression of related genes (e.g. *PTEN*, *TSC2*, *IGFBP3*) and inhibits cell growth and division with high error rates caused by stress [18]. Previous studies have shown that the IGF1 pathway is related to lifespan regulation in nematodes and mammals [19], and inhibition of the mTOR signaling pathway can extend the lifespans of many species—e.g., mice (*Mus musculus*) [20, 21]. Together, studying the p53 pathway is a promising way to uncover mechanism by which large and long-lived species inhibit cancer.

In the present study, we investigated the molecular evolution of 73 genes (downloaded from KEGG; the Kyoto Encyclopedia of Genes and Genomes, **Table S1**) involved in the p53 pathway in 17 cetacean species with high-quality genomes to explore the potential anticancer mechanisms underlying natural longevity. We first traced the gene copy number variation across cetacean genomes to test whether duplicated genes coincide with the evolution of large, long-lived species. For single copy orthologs, we determined whether positive selection is limited to long-lived cetaceans classified by MLS and longevity

quotient (LQ) correcting for body size. Finally, we investigated the relationship between gene evolutionary rates and lifespan variables (MLS, LQ, and BM) in cetaceans. We aimed to use the results from the above analyses to understand the molecular mechanism behind the suppression of tumors in long-lived cetaceans.

Results

According to the MLS and BM records of nonflying eutherian mammals from the AnAge online dataset, we obtained a new allometric equation: $\mathrm{Y}\left(\text{expected longevity}\right) = 3.7136\ \mathrm{BM}^{0.1842}$. The LQ of all cetacean species was calculated based on the allometric equation $\mathrm{LQ} = \mathrm{MLS} \ / \ (3.7136 * \mathrm{BM} \hat{} \ 0.1842)$. The mean LQ value from 65 cetaceans was 0.97 and the 0.5 SD was 0.17 (LQ = 0.97 ± 0.17). Six cetacean species had LQ > 1.14, so we classified them as long-lived species: bowhead whales, long-finned pilot whales (*Globicephala melas*), Pacific white-sided dolphins (*Lagenorhynchus obliquidens*), killer whales (*Orcinus orca*), bottlenose dolphins ((*Tursiops truncatus*), and Indo-Pacific bottlenose dolphins (*T. aduncus*). In contrast, we classified five species short-lived (LQ < 0.80): vaquita (*Phocoena sinus*), harbor porpoises (*P. phocoena*), beluga whales (*Delphinapterus leucas*), baiji (*Lipotes vexillifer*), and common minke whales (*B. acutorostrata*). The remaining species (0.8 < LQ < 1.14) were used as the control **(**Fig. 1). For the MLS, the mean MLS value ± 0.5 SD for all cetaceans was 47.90 ± 15.67. Five species were classified as the long-lived group with MLS > 63.57: blue whales, bowhead whales, humpback whales, killer whales, and sperm whales (*Physeter catodon*). Four species had MLS < 32.23 and were thus considered short-lived: vaquita, harbor porpoises, Yangtze finless porpoises (*Neophocaena asiaeorientalis*), and baiji. In addition, the ancestral state of both MLS and LQ was reconstructed to further classify long-lived lineages in the ancestral nodes (**Figure S1**). The long-lived cetacean species identified by the two standards (LQ > 1.14, MLS > 63.57) were used for subsequent analyses.

Gene duplications in cetacean genomes

To gain insight into the molecular mechanisms of cancer resistance in the p53 pathway, we identified genes that had undergone duplications in cetaceans. In our study, 23 genes were detected to have copy number gains in at least one cetacean lineage **(Figure 2**, **Table S2**). We leveraged the COSMIC v92 [22] and TSGene 2.0 [23] databases and found that, among genes that underwent duplications, more than 50% were tumor suppressor genes (i.e. *CASP3*, *CDK1*, *CDK2*, *CDK6*, *EI24*, *GADD45A*, *IGFBP3*, *PERP*, *RCHY1*, *SFN*, *SIAH1*, *THBS1*). Four genes with two copies were unique to the long-lived cetacean lineages (LQ > 1.14)—*BCL2L1* (long-finned pilot whales), *IGFBP3* (Indo-Pacific bottlenose dolphins), *PERP* (bottlenose dolphins), and *STEAP3* (Indo-Pacific bottlenose dolphins)—whereas only one copy was identified in other cetacean lineages. In addition, three copies of *CCNB2* and two copies of *MDM4* were detected only in the large, long-lived sperm whale, but only one copy of each was found in the other cetacean lineages. We also found that both *CASP3* and *CCNG1* had undergone duplication in all the large long-lived cetacean species (MLS > 63.57) except the common minke whale. In contrast, only one copy of *RCHY1* was identified in the five large, long-lived species, but two copies were found in the other cetacean

species. To further confirm whether the above eight genes with gene copy gains (*BCL2L1*, *IGFBP3*, *PERP*, *STEAP3*, *CCNB2*, *MDM4*, *CASP3*, *CCNG1*) and one gene with copy loss (*RCHY1*) are unique to long-lived cetaceans, we added 17 non-cetacean mammals and found higher copies of both *CASP3* and *PERP* in the well-known long-lived species (including large cetaceans, primates and the naked mole-rat), although there were some exceptions (e.g. cow). In addition, we did not find an effect of genome assembly length or scaffold N50 numbers on estimated gene copy number (**Figure S2**).

Positive selection of p53 pathway-related genes in cetaceans

A total of 46 “one-to-one” orthologous genes were identified among the 73 genes involved in the p53 pathway. To detect signatures of episodic selection in genes occurring along the long-lived cetaceans, we used four different methods: free-ratio and branch-site model from the PAML4.9 package and aBSREL and BUSTED from Datamonkey. The LRT revealed that the free-ratio model that assumes an independent ω on each branch fit the data significantly better than the one-ratio model for three genes (*APAF1*, *CASP8*, *AIFM2*;$P < 0.05$, **Figure 1**, **Table S3**). $\omega > 1$ was only identified in four long-lived branches of both *APAF1* and *CASP8*: the LCA of the humpback whale and terminal branch of the Indo-Pacific bottlenose dolphin of *APAF1*, and the LCA of delphinids and the branch leading to the Pacific white-sided dolphin of *CASP8*. However, *AIFM2* was found to be under positive selection in both long-lived long-finned pilot whales and short-lived vaquita. Similar results were obtained from the more stringent branch-site model—which can detect stronger positive selection acting on only a few sites within a landscape of overall purifying selection—from CODEML. Evidence of positive selection was observed in the two long-lived branches leading to the sperm whale for *TP73* and the long-finned pilot whale for *AIFM2* (**Figure 1**, **Table S3**). Two positively selected sites identified using the BEB approach in both genes (*TP73*: 506; *AIFM2*: 459) had undergone radical changes in at least one property. In addition, three genes (*TP53I3*, *SIVA1*, *GTSE1*) were identified in the short-lived groups and another two (*CCNE1* and *TP53*) along the non-long-lived branch leading to the common minke whale. We then ran another branch-site model in the Datamonkey program aBSREL, which appeared to be markedly more sensitive in detecting episodic selection than branch-site methods from PAML. The result showed evidence of episodic selection on two branches from two genes after correcting for multiple testing. *TP73* was identified to be under positive selection along the large long-lived humpback whale, whereas *GTSE1* was subject to positive selection along the non-long-lived lineage of the LCA of Phocoenidae and Monodontidae (**Table S4**). The BUSTED program, which may be particularly effective at testing for selection limited to foreground branches, further revealed that the *TP73* in the long-lived cetaceans undergoes positive selection ($p < 0.05$, **Table S5**). The above four methods of selection testing identified three positively selected genes (*APAF1*, *CASP8*, *TP73*) unique to the long-lived cetacean species.

Neutral theory predicts that the ω value is higher in species with small effective population sizes (*Ne*), like cetaceans. To test whether the selective signs we identified in the three genes (*APAF1*, *CASP8*, *TP73*) of long-lived cetaceans were due to a low *Ne*, we used branch-site and free ratio models implemented in PAML4.9 to evaluate the ω value in each lineage across 17 non-cetacean mammals. Two genes were also identified to be under positive selection in other long-lived mammals: the ancestral branch of

primatesfor *CASP8* and *TP73*, and *TP73* in the terminal branch of Brandt's bat (*Myotis brandtii*). The results suggest that positively selected genes were not identified in the long-lived cetaceans due to a low *Ne* (**Table S6**).

Gene-phenotype evolution

The phylogenetic generalized least squares (PGLS) regressions were performed between the evolutionary rate of each orthologous gene (represented by root-to-tip ω) and three lifespan-associated traits (MLS, BM, and LQ). Regression analyses revealed that the evolutionary rates of the two genes were significantly positively correlated with LQ: *CD82* (R^2 = 0.392, *P* = 0.004) and *SERPINE1* (R^2 = 0.343, *P* = 0.010, **Figure 3**). For MLS, a positive association between $\log_{10}$ (root-to-tip ω) and $\log_{10}$ (MLS) was identified at *TSC2* (R^2 = 0.364, *P* = 0.008). Moreover, the *TP73* evolution rate was positively related to BM (R^2 = 0.226, *P* = 0.031, **Figure 3**).

Discussion

Larger cetaceans tend to have a longer lifespan than small ones. A previous study suggested that large and long-lived cetaceans might have evolved effective anticancer abilities, since their incidences of cancer were reportedly low [8]. Until now, however, anticancer mechanisms have only been revealed for individual, extremely long-lived species. The p53 pathway plays a key role in tumor suppression because the activation of p53 can induce cell cycle arrest, regulate autophagy, accelerate DNA repair, regulate cell metabolism, and promote cell apoptosis. Here, we comprehensively analyzed the 73 genes involved in the p53pathway across 17 cetacean lineages with relatively high-quality genomes and found evidence that long-lived cetaceans may have a mechanism for preventing cancer (**Figure 4**). Four duplicated genes (i.e. *IGFBP3*, *STEAP3*, *CASP3*, *PERP*) and three positively selected genes (*APAF1*, *CASP8*, *TP73*) were detected in the long-lived cetaceans. Additionally, evolutionary rates of four genes (*CD82*, *SERPINE1*, *TSC2*, *TP73*) were significantly related to three lifespan traits (MLS, LQ, BM) in cetaceans. Genes involved in the p53 pathway of long-lived cetaceans underwent different evolutionary paths, indicating that cancer resistance might be mediated through multiple synergistic effects rather than a single pathway/target.

Tumor suppression by regulation of apoptosis and inhibition of angiogenesis

p53-dependent apoptosis is a key regulator of tumorigenesis because cell apoptosis is significantly reduced in tumors of mice without p53 [24]. The apoptosis rate of African elephant (cancer mortality of 4.81%) cells is reportedly twice that of human (cancer mortality of 11% to 25%) cells [25]. Accordingly, several studies have suggested—and our study corroborated—that regulating apoptosis is important for preventing cancer in long-lived mammals.

In the present study, 100% (3/3) of positively selected genes (*APAF1*, *CASP8*, *TP73*) and 75% (3/4) of genes with multiple numbers of copies (*IGFBP3*, *PERP*, *CASP3*) identified in the long-lived cetacean species were involved in regulating apoptosis, thus inhibiting tumor growth. *APAF1* is a key regulator in the intrinsic or mitochondrial pathway of apoptosis, as was evident from the deletion of this gene in mice

showing clear defects in apoptosis functioning [26, 27]. *CASP3* and *CASP8* are included in the Caspases (CASP) gene family, which is known to play a major role in the execution of apoptotic cascades. *CASP3* plays a key role in removing DNA-damaged cells and preventing the occurrence of tumors—impairing the expression of *CASP3* induces various tumors [28]—whereas overexpression of *CASP8* in breast cancer cells can lead to cell apoptosis [29]. In addition, *IGFBP3*, a duplicated gene that was identified solely in long-lived cetaceans, reportedly exerts antitumor effects through IGF-independent mechanisms, which involve the activation of caspase-dependent apoptosis and angiogenesis suppression by regulating multiple potent angiogenic factors [30, 31]. The *PERP* (p53 apoptosis effector related to PMP22), a critical effector involved in the p53-dependent apoptotic pathway, functions as a tumor suppressor in several human cancers [32]. Interestingly, positive selection and duplication events related to apoptosis identified in long-lived cetaceans were also observed in long-lived lineages of 17 non-cetacean mammals —e.g., *CASP8* and *TP73* were positively selected in the ancestral branch of primates; *TP73* was positively selected in the terminal branch of Brandt's bat; and *PERP* and *CASP3* underwent gene expansion in human, bonobo (*Pan paniscus*), and naked mole rat (*Heterocephalus glaber*). A similar result was reported in a recent study, which found human cancer gene duplications across long-lived mammals [33]. Thus, we suggest that these long-lived lineages might have evolved convergent mechanisms of apoptosis to ward off cancer.

Angiogenesis is a key process in the development of cancer because it allows for the delivery of oxygen, nutrients, and growth factors and the dissemination of tumors to distant organs [34]. There is growing evidence that p53 also helps inhibit angiogenesis by activating downstream genes in the p53 pathway [35]. The evolution rates of two angiogenesis-inhibiting genes (*SERPINE1* and *CD82*) were significantly correlated with LQ in our study. *SERPINE1*, which encodes endothelial plasminogen activator inhibitor-1 (PAI1), plays an important role in inhibiting VEGF (vascular endothelial growth factor)-induced angiogenesis in mice [36]. Another tumor suppressor gene—*TSC2* (tuberous sclerosis complex 2)—exhibits a significantly positive relationship with MLS in cetaceans. *TSC2* reportedly contributes to the regulation of angiogenesis, since the loss of this gene was accompanied by increased levels of HIF-1α (hypoxia-induced factor 1α) and VEGF, whereas the activation of the HIF-1α/VEGF pathway was responsible for hypoxia-induced angiogenesis [37]. In addition, *TSC2* was negatively regulated by mTOR signaling, whereas a genetic inhibition of TOR activity led to a two-fold extension in the lifespan in *Caenorhabditis elegans* [38]. Thus, the inhibition of angiogenesis may be another strategy by which cetaceans protect against cancer.

Evidence for Peto's paradox in large, long-lived cetaceans

Peto's paradox states that large-bodied, long-lived species do not have a greater lifetime risk of cancer than small, short-lived ones [9]. For instance, the African elephant was not found to have a higher cancer incidence than humans, despite the 100-fold difference in number of cells between them [25]. This phenomenon was also found in large-bodied and long-lived cetaceans [8]. The evolution of giant body size and low cancer incidence may be related to the selection of genes and pathways related to cancer in cetaceans.

In the present study, we searched genes related to cancer resistance in cetaceans and found three lines of evidence that support Peto's paradox at the molecular level. First, two positively selected genes (*APAF1* and *TP73*) involved in apoptosis and senescence were identified across the large, long-lived lineages: the sperm whale and humpback whale for *TP73* and the ancestral branch of three baleen whales (blue whale, common minke whale, and humpback whale) for *APAF1*. Second, a significantly positive relationship between the gene evolution and body size was identified for the tumor suppressor gene *TP73*, suggesting that cetaceans evolved to overcome the risk of cancer caused by the accumulation of mutations within the cell. *TP73* was linked to senescence as knockout mice revealed more pronounced aging with increased oxidative damage [39]. Finally, increases in tumor suppression gene copy numbers might be another mechanism of cancer resistance in large, long-lived cetaceans [40]. In our study, both tumor suppression genes that participated in cell apoptosis (*CASP3* and *CCNG1*) were duplicated in the large, long-lived cetaceans, whereas only one copy was found in small-bodied species. Big-bodied baleen whales generally have a long lifespan, with the exception of the common minke whale (MLS = 50 years). The last common ancestor of baleen whales was estimated to have lived over 110 years according to the ancestral state reconstruction in our study (**Figure S1**). Here, it might be understandable that the copy number gains of both genes in the common minke whale were the same as those of other baleen whales classified as long-lived species based on the MLS. Taken together, our results might account for the relatively lower incidence of cancer in large and long-lived species and provide new insights into Peto's paradox. Of course, laboratory experiments are needed to verify this.

Conclusions

In our study, we analyzed 73 genes of the p53 pathway in 17 cetaceans to reveal their potential anticancer mechanisms. Apoptosis-related genes were identified to be under positive selection (*APAF1* and *CASP8*) or have undergone duplications (*IGFBP3*, *PERP*, *CASP3*) in long-lived cetaceans, suggesting that the regulation of apoptosis is a key anti-cancer strategy in the extending the lifespan of cetaceans. Moreover, we suggest that cetaceans also inhibit angiogenesis to protectect against cancer, as the evolutionary rates of three angiogenesis-regulated genes (*SERPINE*, *TSC2*, *CD82*) were significantly related to longevity traits LQ and MLS. Tumor suppressor genes were determined to be under positive selection (*APAF1* and *TP73*) and to have more gene copies (*CASP3* and *CCNG2*) in the extremely large and long-lived cetaceans, this is all molecular evidence for Peto's paradox. In conclusion, our study identified several candidate genes for cancer resistance in cetaceans, which may provide new ideas for lifespan extension in humans.

Methods

Identification of long-lived species

In this study, we used two different standards to define long-lived species: MLS and LQ. LQ is the ratio of observed longevity to expected longevity following Austad and Fisher [41], which can indicate how long a species' lifespan compares to that of other species of a similar size. The expected longevity was

calculated for each species by fitting a linear regression [42]: , where Y is the expected MLS, M is BM, a is a constant representing the proportionality coefficient, and b is the scaling exponent. We first obtained MLS and BM records from the AnAge database [3] for all nonflying eutherian mammals (n = 823, **Table S7**) to calculate the a and b values. This regression equation also applies to cetaceans. The MLS and BM records of 65 cetacean species were collected from the AnAge & Amniote life history databases [43] and published books such as Jefferson et al. [44] (**Table S8**). Species with an LQ or MLS greater than 0.5 SD (standard deviations) from the mean of 65 cetaceans were classified as long-lived. This is because we calculated the sequential threshold from 0 to 1.0 SD from the cetacean mean and found that the same species were included in the long-lived species with any threshold between 0.4 and 0.8 **(Figure S3**). Thus, the species with LQ or MLS less than 0.5 SD from the mean were classified as the short-lived group. To assess whether ancestral nodes in the tree belong to the long-lived species, the ancestral state of MLS and LQ was reconstructed through the maximum-likelihood method implemented in the APE R-package [45]. We used the time-calibrated supertree from McGowen et al. [46] as the reference tree for the cetacean phylogeny.

Identification of multiple copy genes in cetacean genomes

We used 16 cetacean genomes (three mysticetes and 13 odontocetes) for this study with scaffold N50 > 1M downloaded from NCBI and the genome of longest-lived bowhead whale from its corresponding resource (http://www.bowhead-whale.org/, **Table S8**). The reference gene set used in this study came from the human p53 signaling pathway in KEGG (hsa04115). We utilized the method of Tollis et al. [33] to test if multiple copies involved in the p53 pathway occurred in cetaceans. We first used BLAT software [47] to look for p53 pathway orthologous genes in each cetacean genome (minscore = 55, minidentity = 60), using all the known human protein sequences as initial references. To collect p53 pathway gene paralogs, the putative homologs in cetaceans with a length at least 2/3 that of the coding sequence were taken as queries in blast searches against the human protein database using BLASTx [48], and cetacean copies were only kept when the top hit had over 65% sequence identity to the human sequence. Additionally, genes with multiple copies identified in cetaceans were scanned in the genomes of 17 non-cetacean species to detect whether the duplicated genes were unique to cetaceans (**Table S8**). To assess the influence of genome assembly quality on copy number detection, we analyzed the correlation between the normalized copy numbers (the total number of copies divided by the total number of genes) and genome assembly length and scaffold N50.

Sequence retrieval and alignment

We performed TBLASTN searches [49] against cetacean genomes using bottlenose dolphins genes as queries to obtain the coding sequence of high-confidence single-copy orthologs of the p53 pathway, setting the expected cut off value to 1e-5 (**Table S9**). The downloaded sequences that covered at least 75% of the entire coding sequences were retained for further analysis. Multiple alignments of orthologous sequences were first carried out with MACSE, a tool that prevents the disruption of frameshifts and stop codons [50]. Sequences were then aligned with PRANK, which outperforms other alignment tools and

produces the fewest false positives [51]. Finally, potentially unreliable regions of multiple alignments were removed using the Gblocks program under parameters “-t=c -b1 = 5 -b2 = 6 -b3 = 8 -b4 = 10 -b5 = h” [52].

Selection detection

To elucidate the molecular evolution of orthologous genes involved in the p53 pathway, we estimated the rates of nonsynonymous (d_N) and synonymous (d_S) ($\omega = d_N / d_S$) using the maximum likelihood (ML) method in the CODEML program implemented in PAML 4.9e [53]. Values of $\omega > 1$, $= 1$, and < 1 indicate positive selection, neutral selection, and purifying selection, respectively. A well-accepted phylogeny of cetaceans [46] was used as an input tree for analyses of each orthologous gene.

To evaluate whether positive selection was restricted to long-lived cetacean lineages, we used the free-ratio model and branch-site model implemented in the CODEML program. The free-ratio model (M1), which allow independent ω values for each branch, was compared with the null one-ratio model (M0) with the same ω for all branches in a tree. The branch-site model, which assumes that codons are under positive selection along a specific lineage with $\omega_2 > 1$, was compared with a null model Ma0 with a fixed $\omega_2 = 1$. The positively selected sites were identified using a Bayes Empirical Bayes (BEB) analysis [54] with posterior probabilities ≥ 0.80. A false discovery rate (FDR, cutoff = 0.05) correction for multiple tests was conducted in the branch-site model analysis [55]. All nested models were compared using the likelihood ratio test (LRT) to determine which models were statistically different from the null model, *P* values < 0.05 after FDR correction were considered significant. We further manually checked the putative positively selected sites identified by the branch-site model to minimize the impact of potential false positives. The putative positively selected sites were removed if the aligned positions of the positively selected sites were located nearly or within a poorly conserved section, such as close to insertions or deletions or surrounded by large gaps. The positively selected sites further employed TreeSAAP v3.2 to measure the selective influences on 31 structural and biochemical amino acid properties, the residues that had category Z-scores greater than six were regarded as radical amino acid changes [56].

Additional robustness of statistical significance of aBSREL (adaptive branch-site random effects likelihood) [57] and BUSTED (branch-site unrestricted statistical test for episodic diversification) [58] implemented in Datamonkey were further used to investigate whether episodic positive selection acted on the long-lived lineages. aBSREL identifies branches under positive selection without a priori knowledge about which lineages are of interest by sequential likelihood ratio tests. BUSTED is capable of detecting episodic positive selection that acts on a subset of branches in the phylogeny in at least one site within the alignment [58]. BUSTED, which splits branches into the foreground and background partitions, includes three ω classes ($\omega 1 \leq \omega 2 \leq \omega 3$, unconstrained model) and tests for positive selection against a constrained null model ($\omega = 1$, disallowing positive selection) on the foreground branches.

Association analysis between ω and longevity-associated traits

PGLS regression was used to further assess the relationships between gene evolution and lifespan associated traits (i.e., MLS, BM, and LQ). The maximum-likelihood method was used as a quantitative measure of phylogenetic correlation (λ). Values of λ range between 0 and 1, with λ close to 0 indicating traits that are phylogenetically independent and λ of 1 or close to 1 indicating that the genes show a strong phylogenetic signal. All analyses were performed in R 3.4.2 [59]. The gene evolutionary rate, ω value, for each gene was calculated using the free-ratio model implemented in the CODEML program of PAML 4.9e. The root-to-tip ω (average ω ratio from the ancestral cetacean to each terminal species tip), which includes gene evolutionary history, is more suitable for regressions against phenotypic data from extant species. The ultrametric tree of 17 cetacean species from McGowen et al. [46] was used as an input tree. Lifespan variables and root-to-tip ω were $\log_{10}$ transformed to improve normality for the regression analysis.

Abbreviations

BLAST: Basic local alignment search tool; BLAT: The BLAST-like alignment tool; FDR: False discovery rate; LCA: Last common ancestor; LRT: Likelihood ratio test; PAML: Phylogenetic analysis by maximum likelihood; PGLS: Phylogenetic generalized least squares.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent to publish

Not applicable.

Availability of data and materials

The data generated and analyzed during this study are included in this article and its additional files, including 9 tables and 7 figures.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the National Natural Science Foundation of China (NSFC, grant nos. 32070409, 31772448 to SX), the Key Project of the NSFC (grant nos. 32030011, 31630071 to GY), the NSFC (grant nos. 31872219 to WR), the Priority Academic Program Development of Jiangsu Higher Education Institutions to GY and SX, and the Qing Lan Project of Jiangsu Province to SX. These funding

bodies played no role in study design, data collection, analysis, interpretation of data, and writing the manuscript.

Authors' contributions

S.X. designed the study. X.L. was responsible for the data collection and analysis. S.X. and X.L. drafted the manuscript. Z.Y., S.X. revised the manuscript. F.Y., X.H. and L.S. participated in the data collection. W.R. and G.Y. helped edit the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank members of the Jiangsu Key Laboratory for Biodiversity and Biotechnology, Nanjing Normal University, for their contributions to this paper. We thank Mr. Tianzhen Wu and Mr. Xu Zhou for their helpful suggestions. We are particularly grateful to Dr. Ran Tian and Dr. Weijian Guo for their technical support.

References

1. Carnes BA. What is lifespan regulation and why does it exist? Biogerontology. 2011;12(4):367–374.
2. Fushan AA, Turanov AA, Lee SG, Kim EB, Lobanov AV, Yim SH, Buffenstein R, Lee SR, Chang KT, Rhee H. Gene expression defines natural changes in mammalian lifespan. Aging cell. 2015;14(3):352–365.
3. Tacutu R, Thornton D, Johnson E, Budovsky A, Barardo D, Craig T, Diana E, Lehmann G, Toren D, Wang J. Human ageing genomic resources: new and updated databases. Nucleic Acids Res. 2018;46(D1):D1083-D1090.
4. Haley D. Marine mammals of eastern North Pacific and Arctic waters. Washington: Pacific Search Press; 1978.
5. George JC, Bada J, Zeh J, Scott L, Brown SE, O'Hara T, Suydam R. Age and growth estimates of bowhead whales (Balaena mysticetus) via aspartic acid racemization. Can J Zool. 1999;77(4):571–580.
6. Tollis M, Boddy AM, Maley CC. Peto's Paradox: how has evolution solved the problem of cancer prevention? BMC Biol. 2017;15(1):1–5.
7. Tejada-Martinez D, de Magalhães JP, Opazo JC. Positive selection and gene duplications in tumour suppressor genes reveal clues about how cetaceans resist cancer. P Roy Soc B-Biol Sci. 2021;288(1945):20202592.
8. Nagy JD, Victor EM, Cropper JH. Why don't all whales have cancer? A novel hypothesis resolving Peto's paradox. Integr Comp Biol. 2007;47(2):317–328.
9. Peto R, Roe F, Lee P, Levy L, Clack J. Cancer and ageing in mice and men. Brit J Cancer. 1975;32(4):411–426.

10. Uys C, Best P. Pathology of lesions observed in whales flensed at Saldanha Bay, South Africa. J Comp Pathol. 1966;76(4):407-IN416.
11. Keane M, Semeiks J, Webb AE, Li YI, Quesada V, Craig T, Madsen LB, van Dam S, Brawand D, Marques PI. Insights into the evolution of longevity from the bowhead whale genome. Cell Rep. 2015;10(1):112–122.
12. Tollis M, Robbins J, Webb AE, Kuderna LF, Caulin AF, Garcia JD, Bèrubè M, Pourmand N, Marques-Bonet T, O'Connell MJ. Return to the sea, get huge, beat cancer: an analysis of cetacean genomes including an assembly for the humpback whale (Megaptera novaeangliae). Mol Biol Evol. 2019;36(8):1746–1763.
13. Toren D, Kulaga A, Jethva M, Rubin E, Snezhkina AV, Kudryavtseva AV, Nowicki D, Tacutu R, Moskalev AA, Fraifeld VE. Gray whale transcriptome reveals longevity adaptations associated with DNA repair and ubiquitination. Aging Cell. 2020;19(7):e13158.
14. Sabapathy K, Lane DP. Corrigendum to 'Understanding p53 functions through p53 antibodies'. J Mol Cell Biol. 2019;11(12):1105.
15. Cha H-J, Yim H. The accumulation of DNA repair defects is the molecular origin of carcinogenesis. Tumor Biol. 2013;34(6):3293–3302.
16. Caulin AF, Graham TA, Wang L-S, Maley CC. Solutions to Peto's paradox revealed by mathematical modelling and cross-species cancer gene analysis. Philos T R Soc B. 2015;370(1673):20140222.
17. Sulak M, Fong L, Mika K, Chigurupati S, Yon L, Mongan NP, Emes RD, Lynch VJ. TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. elife. 2016; 5:e11994.
18. Feng Z, Levine AJ. The regulation of energy metabolism and the IGF-1/mTOR pathways by the p53 protein. Trends Cell Biol. 2010;20(7):427–434.
19. Friedman DB, Johnson TE. A mutation in the age-1 gene in Caenorhabditis elegans lengthens life and reduces hermaphrodite fertility. Genetics. 1988; 118(1):75–86.
20. Kaeberlein M, Powers RW, Steffen KK, Westman EA, Hu D, Dang N, Kerr EO, Kirkland KT, Fields S, Kennedy BK. Regulation of yeast replicative life span by TOR and Sch9 in response to nutrients. Science. 2005;310(5751):1193–1196.
21. Miyazaki M, McCarthy JJ, Esser KA. Insulin like growth factor-1-induced phosphorylation and altered distribution of tuberous sclerosis complex (TSC) 1/TSC2 in C2C12 myotubes. FEBS J. 2010;277(9):2180–2191.
22. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E. COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res. 2019;47(D1):D941-D947.
23. Zhao M, Kim P, Mitra R, Zhao J, Zhao Z. TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. Nucleic Acids Res. 2016; 44(D1):D1023-D1031.
24. Symonds H, Krall L, Remington L, Saenz-Robles M, Lowe S, Jacks T, Van Dyke T. p53-dependent apoptosis suppresses tumor growth and progression in vivo. Cell. 1994;78(4):703–711.

25. Abegglen LM, Caulin AF, Chan A, Lee K, Robinson R, Campbell MS, Kiso WK, Schmitt DL, Waddell PJ, Bhaskara S. Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans. Jama. 2015;314(17):1850–1860.
26. Yoshida H, Kong Y-Y, Yoshida R, Elia AJ, Hakem A, Hakem R, Penninger JM, Mak TW. Apaf1 is required for mitochondrial pathways of apoptosis and brain development. Cell. 1998;94(6):739–750.
27. Bao Y, Wu X, Yuan D, Shi W, Shi J. High expression of Pirh2 is associated with poor prognosis in glioma. Cell Mol Neurobiol. 2017;37(8):1501–1509.
28. Jiang XH, Chun-Yu Wong B, Yuen ST, Jiang SH, Cho CH, Lai KC, Lin MC, Kung HF, Lam SK. Arsenic trioxide induces apoptosis in human gastric cancer cells through up-regulation of P53 and activation of caspase-3. Int J Cancer. 2001;91(2):173–179.
29. Fritsch M, Günther SD, Schwarzer R, Albert M-C, Schorn F, Werthenbach JP, Schiffmann LM, Stair N, Stocks H, Seeger JM. Caspase-8 is the molecular switch for apoptosis, necroptosis and pyroptosis. Nature. 2019;575(7784):683–687.
30. Kim EB, Fang X, Fushan AA, Huang Z, Lobanov AV, Han L, Marino SM, Sun X, Turanov AA, Yang P. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature. 2011;479(7372):223–227.
31. Williams AC, Collard TJ, Perks CM, Newcomb P, Moorghen M, Holly JM, Paraskeva C. Increased p53-dependent apoptosis by the insulin-like growth factor binding protein IGFBP-3 in human colonic adenoma-derived cells. Cancer Res. 2000;60(1):22–27.
32. Roberts O, Paraoan L. PERP-ing into diverse mechanisms of cancer pathogenesis: regulation and role of the p53/p63 effector PERP. BBA-Rev Cancer. 2020;188393.
33. Tollis M, Schneider-Utaka AK, Maley CC. The evolution of human cancer gene duplications across mammals. Mol Biol Evol. 2020;37(10):2875–2886.
34. Al-Ostoot FH, Salah S, Khamees HA, Khanum SA. Tumor angiogenesis: Current challenges and therapeutic opportunities. Cancer Treat Res Commun. 2021;28:100422.
35. Teodoro JG, Evans SK, Green MR. Inhibition of tumor angiogenesis by p53: a new role for the guardian of the genome. J Cell Mol Med. 2007;85(11):1175–1186.
36. Wu J, Strawn TL, Luo M, Wang L, Li R, Ren M, Xia J, Zhang Z, Ma W, Luo T. Plasminogen activator inhibitor-1 inhibits angiogenic signaling by uncoupling vascular endothelial growth factor receptor-2-αVβ3 integrin cross talk. Arterioscl Throm Vas. 2015;35(1):111–120.
37. Brugarolas JB, Vazquez F, Reddy A, Sellers WR, Kaelin Jr WG. TSC2 regulates VEGF through mTOR-dependent and-independent pathways. Cancer cell. 2003; 4(2):147–158.
38. Vellai T, Takacs-Vellai K, Zhang Y, Kovacs AL, Orosz L, Müller F. Influence of TOR kinase on lifespan in C. elegans. Nature. 2003;426(6967):620–620.
39. Rufini A, Niklison-Chirou MV, Inoue S, Tomasini R, Harris IS, Marino A, Federici M, Dinsdale D, Knight RA, Melino G. TAp73 depletion accelerates aging through metabolic dysregulation. Gene Dev. 2012;26(18):2009–2014.

40. Leroi AM, Koufopanou V, Burt A. Cancer selection. Nat Rev Cancer. 2003; 3(3):226–231.
41. Austad SN, Fischer KE. Mammalian aging, metabolism, and ecology: evidence from the bats and marsupials. J. Gerontol. 1991;46(2):B47-B53.
42. Schmidt-Nielsen K, Knut S-N. Scaling: why is animal size so important?. Cambridge: Cambridge university press; 1984.
43. Myhrvold NP, Baldridge E, Chan B, Sivam D, Freeman DL, Ernest SM. An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles: Ecological Archives E096-269. Ecology. 2015; 96(11):3109–3109.
44. Jefferson TA, Webber MA, Pitman RL. Marine mammals of the world: a comprehensive guide to their identification. Oxford: Elsevier; 2011.
45. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 2004;20(2):289–290.
46. McGowen MR, Tsagkogeorga G, Álvarez-Carretero S, Dos Reis M, Struebig M, Deaville R, Jepson PD, Jarman S, Polanowski A, Morin PA. Phylogenomic resolution of the cetacean tree of life using target sequence capture. Syst Biol. 2020;69(3):479–501.
47. Kent WJ. BLAT—the BLAST-like alignment tool. Genome Res. 2002;12(4):656–664.
48. Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. Domain enhanced lookup time accelerated BLAST. Biol Direct. 2012;(1):1–14.
49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–410.
50. Ranwez V, Douzery EJ, Cambon C, Chantret N, Delsuc F. MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Mol Biol Evol. 2018;35(10):2582–2584.
51. Löytynoja A. Phylogeny-aware alignment with PRANK. Methods Mol Biol. 2014;155–170.
52. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17(4):540–552.
53. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–1591.
54. Yang Z, Wong WS, Nielsen R. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol. 2005;22(4):1107–1118.
55. Anisimova M, Yang Z. Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. Mol Biol Evol. 2007;24(5):1219–1228.
56. Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA. TreeSAAP: selection on amino acid properties using phylogenetic trees. Bioinformatics. 2003;19(5):671–672.
57. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol. 2015;32(5):1342–1353.

58. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM. Gene-wide identification of episodic selection. Mol Biol Evol. 2015;32(5):1365–1371.
59. Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N, Pearse W. Caper: comparative analyses of phylogenetics and evolution in R. Methods Ecol Evol. 2012;3:145–151.

Figures

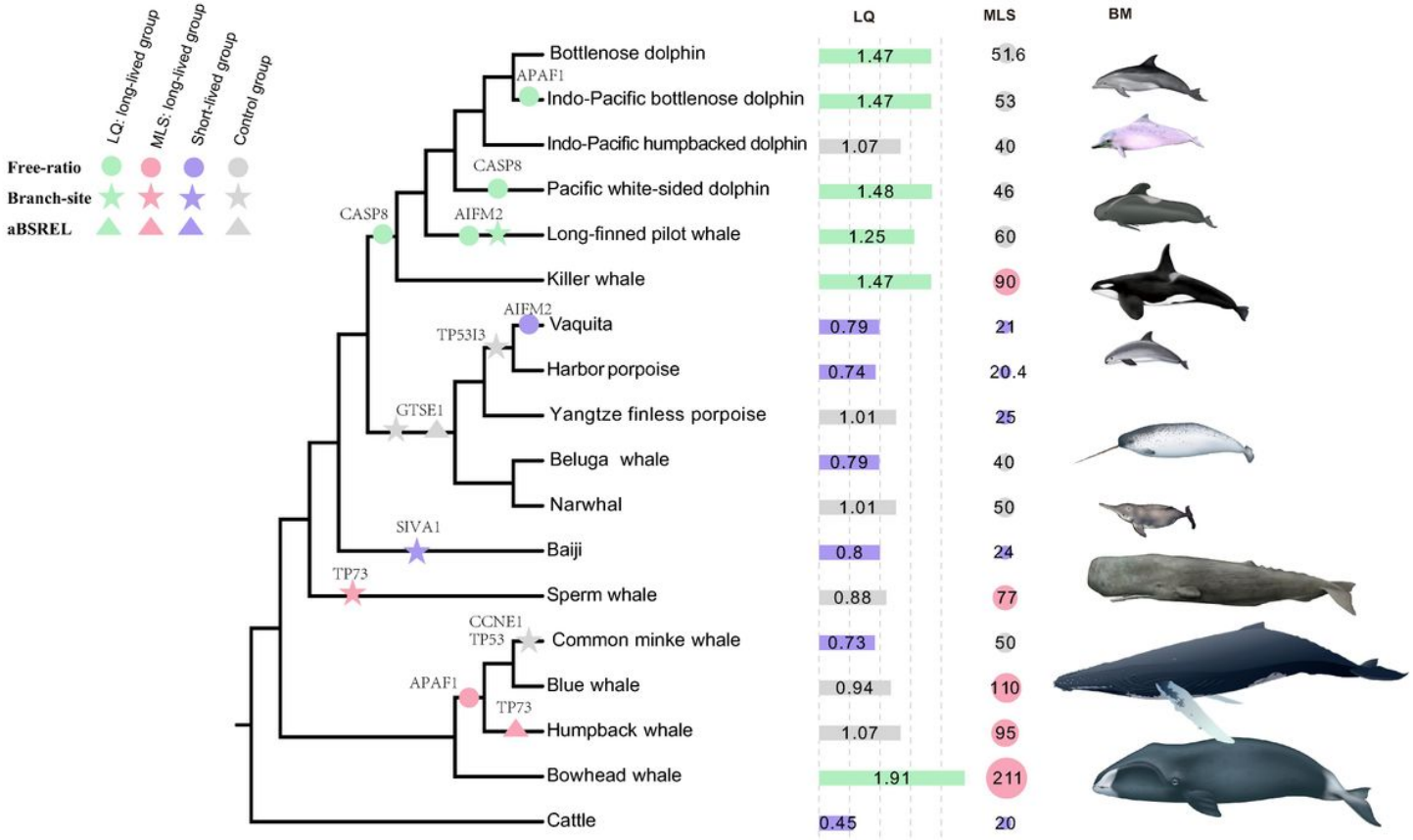

Figure 1

Evidence of positive selection across the phylogeny of cetaceans. The long-lived species identified based on the longevity quotient (LQ) and maximum lifespan (MLS) are marked in green and red, respectively. In contrast, short-lived species and the control group are in purple and grey, respectively. Significant positive selection identified by the free-ratio model, branch-site model, and aBSREL is indicated by a circle, pentagram, and triangle, respectively.

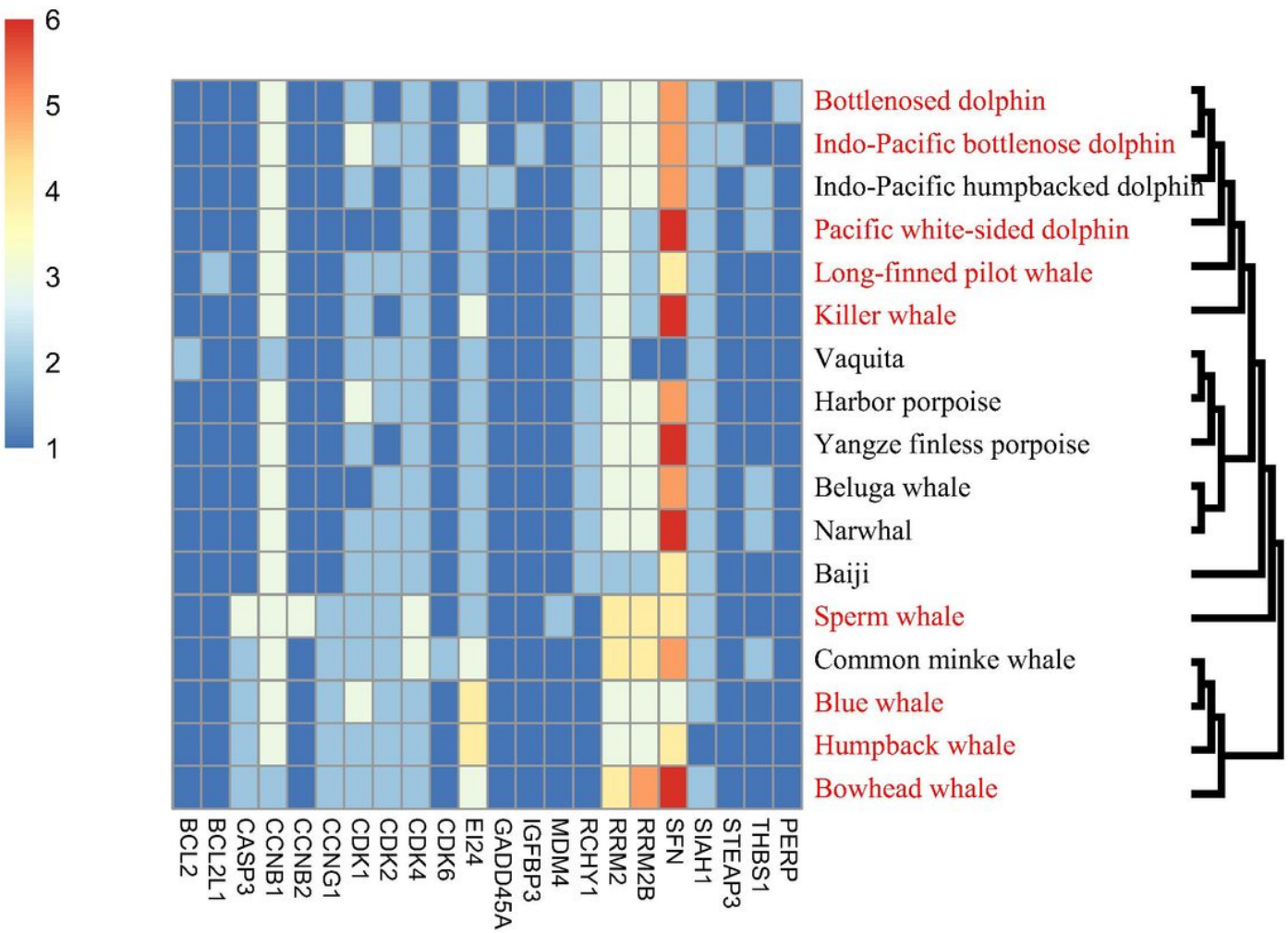

Figure 2

The copy number variation of p53 pathway-related genes in cetaceans. The heat map shows copy number variation in cetaceans. The colors correspond to the number of copies, with red indicating increasing copy number. Red bars represent long-lived species.

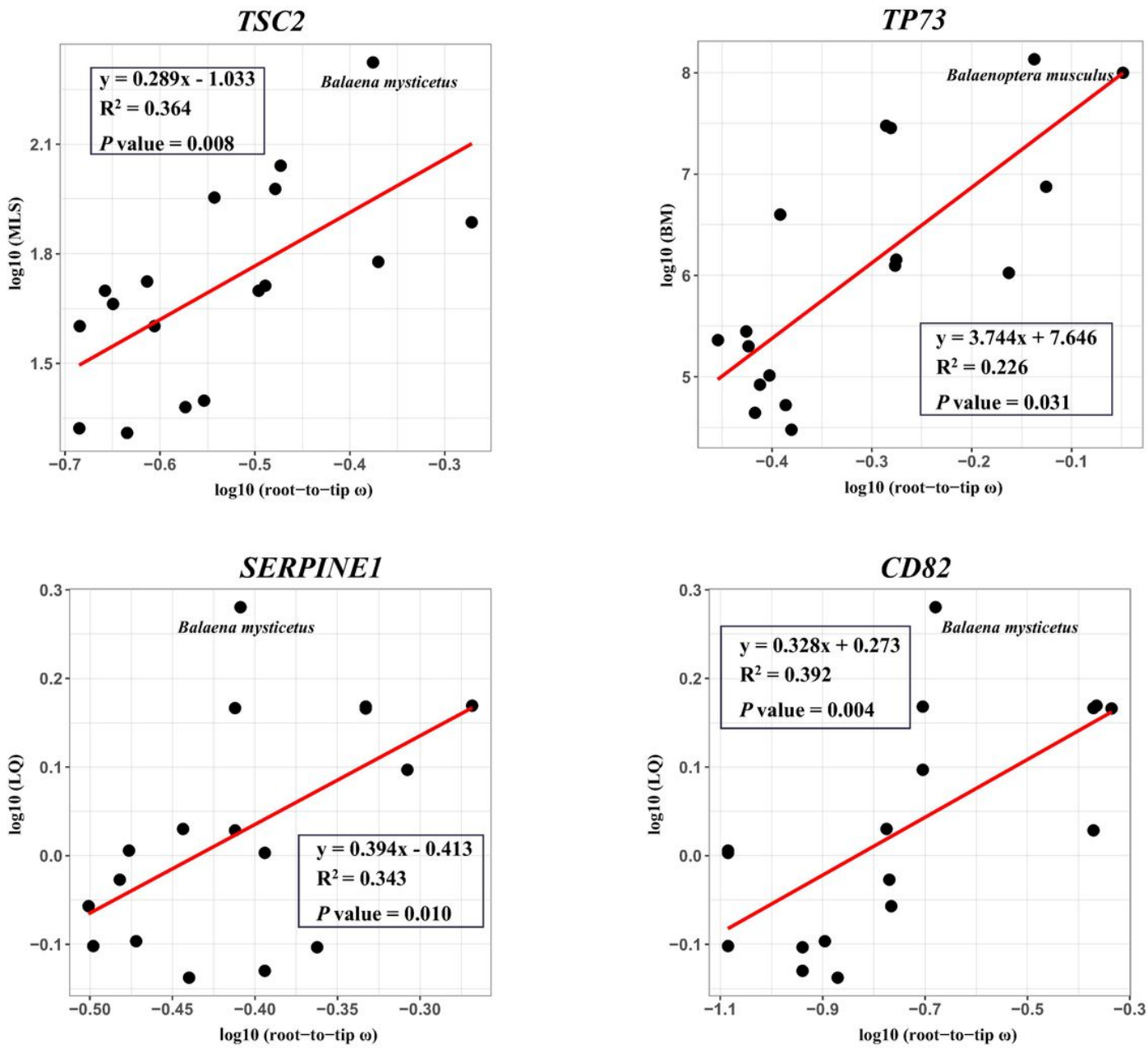

Figure 3

Regression analyses between root-to-tip (ω) and three longevity traits (MLS, BM, and LQ).

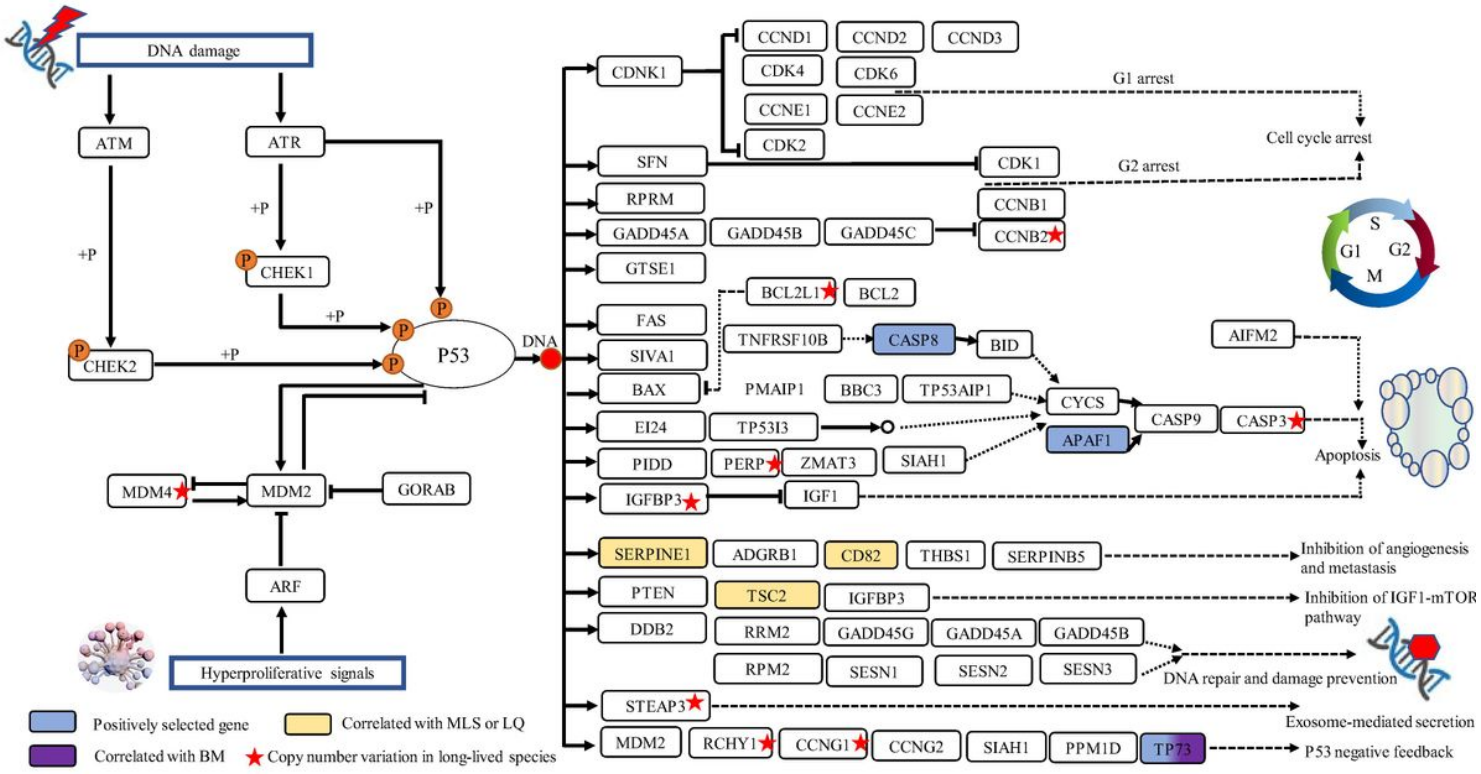

Figure 4

Diagram of p53 pathway-related genes evolution in the long-lived cetaceans. Positively selected genes (blue), genes related to three longevity phenotypes [MLS or LQ (gold) and BM (purple)] identified in the long-lived cetacean lineages are marked with different colors. The red star indicates gene duplications examined in the long-lived cetaceans.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfiles.docx
- Additionalfile1.xlsx
- Additionalfile2.xlsx
- Additionalfile3.docx
- Additionalfile4.xlsx
- Additionalfile5.xlsx
- Additionalfile6.pdf