

# New insights into the genetic mechanisms of body size evolution in Carnivora

## Xin Huang

Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Science, Nanjing Normal University, Nanjing 210023, China

## Di Sun

Jiangsu Key Laboratory of Biodiversity and Biotechnology, College of Life Science, Nanjing Normal University, Nanjing 210023, China

## Tianzhen Wu

Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210023, China

## Xing Liu

Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210023, China

## Shixia Xu

Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210023, China <https://orcid.org/0000-0002-0283-0986>

## Guang Yang (✉ [gyang@njnu.edu.cn](mailto:gyang@njnu.edu.cn))

Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing 210023, China <https://orcid.org/0000-0003-3202-3197>

---

## Research article

**Keywords:** Carnivora, Body size associated genes, Peto's Paradox, Rapid evolving genes, Fixed amino acid changes

**Posted Date:** November 24th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-112815/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## Background

The range of body sizes in Carnivora is unparalleled in any other mammalian order, with more than 130,000 times in body mass and 50 times in length. However, the molecular mechanisms underlying the huge difference in body size of Carnivora have not been explored so far.

## Results

Herein, we performed a comparative genomics analysis of 20 carnivores to explore the genetic basis of great body size variation in carnivores. Phylogenetic generalized least squares (PGLS) revealed that 337 genes were significantly related to both head body length and body mass, these genes were defined as body size associated genes (BSAGs). Fourteen positively related BSAGs were found to be associated with obesity and three of which were identified to be under rapid evolution in the extremely large body-sized carnivores, which suggested that these obesity-related BSAGs might have driven the body size expansion in carnivores. Interestingly, 100 BSAGs were examined to be associated with cancer control in carnivores, particularly 15 cancer-related genes were found to be under rapid evolution in extremely large carnivores. These results strongly suggested that large body-sized carnivores might have evolved effective mechanism to resist cancer, which could be regarded as molecular evidence to support for the Peto's paradox. For small carnivores, we identified 15 rapidly evolving genes and found six genes with fixed amino acid changes that were reported to reduce body size.

## Conclusion

This study brings new insights into the molecular mechanisms that drove the diversifying evolution of body size in carnivores, and provides new target genes for exploring the mysteries of body size evolution in mammals.

## Background

Carnivores, or the mammalian order Carnivora, feed primarily or exclusively on animal matter. They represent a highly diverse and successful group of mammals, which is on the top of the food chain. A total of 280 species were included in 11 families of Carnivora [1], which are widely distributed all over the world, covering most of the major land masses, rivers, and all of the oceans. Carnivores are well-known for their dietetic preferences, carnassial dentition, skull shape and body size [2].

Body size is closely related to habitat, life history, metabolism and risk of extinction, and etc. [3]. Previous studies revealed that the basal metabolic rate of carnivores decreased with the increase of body mass [4–6]. For feeding habits, the prey size and diversity increase with body size in predatory carnivores.

Interestingly, statistical analysis showed that among terrestrial carnivores, the herbivorous species are relatively large while the insectivorous species are relatively small [7]. A typical case is the Ursidae that may have increased in size due to their diet including a greater proportion of fruits and vegetation [8]. Importantly, the range of body sizes of carnivores is unparalleled in any other mammalian orders [9, 10]. The largest carnivore, male southern elephant seal (*Mirounga leonina*), is more than 4,000 kg in body mass and over 5.8 m in length [11, 12], whereas the smallest carnivore, the least weasel (*Mustela nivalis*), is only 29 g in body mass and 0.114 m in body length [13, 14]. Such a huge difference with more than 130,000 times in body mass and 50 times in length makes the carnivores a good target for investigating the mechanism of mammalian body size evolution.

The formation of the discrepant body size of species in carnivores is actually a manifestation of adapting to their respective niche and has their own ecological advantages. For small carnivores, they have access to a wide variety of food resources that not available to larger carnivores, owned improved reproductive efficiency and a greater ability to respond to environmental emergencies [15–17]. For example, small carnivores such as Mustelidae and Viverridae, were adapted for exploiting small rodent prey and invertebrates. Their smaller body makes them move swiftly enough to follow and pounce on prey and more inconspicuous in open vegetation during hunt [10]. In contrast, the increased body size brings a large of benefits including the ability to exploit vast food resources, increased competitiveness, increased defense against predation, and extended longevity [18, 19]. However, larger body size with more cells will have a higher risk of cancer in theory assuming that each cell has a consistent risk of mutations [20]. Notably, some extremely large carnivores such as the walrus (*Odobenus rosmarus*) and polar bear (*Ursus maritimus*) that can live longer than 40 years was not found have higher risk of cancer than smaller species such as Mustelidae species that have a lifespan of only several years [21]. This phenomenon is well-known as the Peto's paradox [20, 22]. And carnivores may be an excellent target for testing the Peto's paradox at the molecular level.

Until now, the molecular mechanisms regulating the body size of carnivores remain poorly explored. Previous studies mainly focused on the interspecific variation of body size, especially confined to the domestic dog (*Canis lupus familiaris*). A recent study showed that variants in *IGF1*, *COL11A2*, *ITGA10*, and *ADAMTS17* contributed to height that segregates within specific dog breeds [23]. The constantly updated high-quality genomes of carnivores provide new opportunities for insight into the mechanism of huge body size variation among carnivores. In the present study, 20 high-quality genomes of carnivores were used for comparative genomic analysis. First, phylogenetic generalized least squares (PGLS) methods were used to scan the body-size-associated genes (BSAGs). Second, we determined the rapidly evolving genes (REGs) in extremely large or small carnivores, respectively. And fixed amino acid changes were also identified in different body size groups. Finally, selective pressure variation on cancer-related BSAGs among different carnivores may partly verify the Peto's paradox. From above results, we expect to provide some novel insights into the molecular mechanism of body size evolution in carnivores and mammals.

# Results

## Genome-scanning of BSAGs and functional enrichment

A total of 6,667 one-to-one orthologous genes were identified in the 20 carnivores and one cow genomes (**Figure 1; Table S1**) by Orthofinder and our in-house Perl scripts. PGLS revealed that 1,132 and 668 genes were significantly associated with head body length and body mass, respectively. There were 337 genes were significantly related to both head body length and body mass after performing the two-step of the calibration procedure ( $P \text{ value.all/robust/max} < 0.05$ ), which were defined as body size associated genes (BSAGs; **Table S2**). According to the tendency or the correlation slope, 256 genes showed positive correlation and 81 genes exhibited negative correlation.

For 256 positively correlated BSAGs, functional enrichment analyses revealed that 61.7% (161/261) genes were enriched significantly ( $P < 0.05$ ) in 164 GO terms (**Table S3**). There were 28.6% (46/161) genes annotated to metabolic process, such as “NADP metabolic process”, “glycoprotein metabolic process” and “cellular amino acid metabolic process” (**Figure 2**). And 32.3% (52/161) were significantly enriched for GO categories that associated with growth and development, including “positive regulation of growth”, “cardiac chamber development”, “positive regulation of nervous system development” and “epidermis development” (**Figure 2**), etc. For instance, seven genes (*ADAM10*, *DBN1*, *NTRK3*, *PPIB*, *MAP2K5*, *WNT2* and *ZFPM2*) which played key role in maintaining normal development of organ or body were enriched in the GO term of the “positive regulation of growth”. However, due to the scattered functions of these BSAGs, there were only 21 (8.2%) genes significantly enriched in KEGG pathways (**Table S3**), like “cytokine-cytokine receptor interaction”, “fanconi anemia pathway” and so on.

For negative correlated BSAGs, 29.6% (24/81) genes were defined in GO categories such as “telomere maintenance”, “DNA repair”, “negative regulation of DNA metabolic process”, “negative regulation of cell activation” and “apoptotic signaling pathway” (**Figure 2; Table S3**). There were no significant enriched terms when mapped these negative correlated genes to KEGG database.

We then manually looked up the biological functions of these BSAGs by searching literature and multiple databases and identified 14 positively correlated BSAGs (*BRAP*, *CHCHD5*, *CPT1C*, *GPR1*, *LDLR*, *MAP2K5*, *PLEKHS1*, *SLC30A8*, *ST3GAL2*, *STX16*, *ZFH3*, *ZGRF1*, *ZNF395* and *ZPLD1*) were associated with “obesity” (**Table S4**), which is a manifestation of an enlarged-body-sized phenotype. For instance, a significant positive association between log (root-to-tip  $\omega$ ) and log (body mass) were tested at the *BRAP*, *STX16*, *ZGRF1* and *ZPLD1* gene (**Figure 3**) respectively, and these four genes were reported to cause obesity in human. Furthermore, we also found 100 BSAGs were associated with cancer control, including “DNA repair”, “cell cycle control, apoptosis, adhesion and autophagy”, “immune response” and so on (**Table S5**). A total of 21 BSAGs (*ADAM11*, *APC*, *BRCA2*, *CDH11*, *CERS2*, *DSC3*, *DTWD1*, *EPHB6*, *ERCC3*, *ERCC4*, *FANCC*, *HELQ*, *HRG*, *ING1*, *INTS6*, *POU6F2*, *STAG1*, *TEP1*, *TET1*, *TRMT2A* and *ZFH3*) were identified as tumor suppressors according to the literature or CGC database (**Table S6**). For example, *BRCA2* (breast cancer type 2 susceptibility protein) is a tumor suppressor gene that links with an increased risk for breast and ovarian cancers, and it plays a central role in DNA repair, transcription, and

cell cycle in response to DNA damage. In addition, 18 genes were related to immune response, development and maturation of immunocyte, such as *ITK* and *TNFRSF17* (**Table S5**).

### REGs in different body sizes groups

“Branch model” implemented in Codeml of PAML 4.9e program was used to identify rapidly evolving genes (REGs) respectively in the small and extremely large groups. These results showed that divergent selective pressure might have acted on carnivores with contrasting body size.

In the small-sized group, a total of 15 BSAGs were found to be under rapid evolution (**Table S7**). Among them, *CTLA4* (cytotoxic T-lymphocyte associated protein 4) is a member of the immunoglobulin superfamily and encodes a protein that transmits an inhibitory signal to T cells. In addition, *CTLA4* is related to lower body mass index in human, which was in keeping with small-body-sized phenotypes.

By contrast, 55 REGs were determined in extremely large-sized group (**Table S7; Figure 4**). Fifteen genes were significant enrichment in growth and development (such as nerve, muscle and cardiac development), or phenotypic changes in body size (*ATP8B1*, *DIS3*, *POMGNT1*, *SLITRK5*, *ST3GAL2*, *TENM3*, *ZGRF1* and *ZPLD1*). For instance, *ST3GAL2* (ST3 beta-galactoside alpha-2,3-sialyltransferase 2) encodes a type  $\alpha$  membrane protein that catalyzes the transfer of sialic acid from CMP-sialic acid to galactose-containing substrates, and this gene is associated with increased fat mass and lean body mass, which is consistent with large-body-sized phenotypes. Additionally, fifteen genes were associated with cancer control (**Table S5**), among which, three genes are related to immunity (*MAGT1*, *RFXANK* and *SKAP2*), *ADGRL3* and *TENM3* are relevant to cell adhesion, and two genes (*ADAM11* and *TEP1*) had been proved to be tumor suppressor genes. For example, *ADAM11* (ADAM metallopeptidase domain 11) has been identified as a candidate tumor suppressor gene in human breast cancer, it encodes a member of the ADAM (a disintegrin and metalloprotease) protein family which take part in a variety biological process involving cell-cell/matrix interactions, muscle and nerve development.

### Fixed amino acid changes in extremely small body size groups

Identifying fixed amino acid changes in a certain group may help to explain the molecular mechanism of the occurrence of specific phenotypes. In the present study, we identified six fixed amino acid changes of six genes in extremely small carnivores (*CDC7*, *ENG*, *LIG4*, *MMP2*, *POLE* and *TSPAN8*; **Figure 5**) but none was found in the extremely large carnivores. And these sites were located in their functional domains of respective proteins that were identified by Pfam. For instance, a unique change, S513L, was located in the protein kinase domain of *CDC7* gene and another mutation (S784P) locates in the DNA ligase IV domain of *LIG4* gene, both genes were associated with reduced-body-sized phenotype in human or mice.

## Discussion

The great difference in body size between the largest and smallest carnivores, with more than 130,000 times in body mass and 50 times in length, is significantly shocking in mammals. Reconstructions

incorporating fossil data supported small body size (<5kg) for the ancestors of the species in Caniformia [24]. Though a few studies have been conducted to reveal molecular mechanisms for body size variation within species (such as the domestic dog), the genetic mechanisms underlying the huge differences in body size among the living carnivores are not well understood. In the present study, we scanned the whole genomes of 20 carnivores covering contrasting body size and identified that 337 genes were significantly related with both head body length and body mass. Selective pressure analyses revealed different evolutionary patterns in the extremely large and small species. Furthermore, fixed amino acid changes have also been identified in extremely small-body-sized carnivores and suggest the potential functions of these loci to restrict body size growth. All these results provide a genetic basis for studying the body size evolution in carnivores.

### **Obesity-related genes contributing to increasing body size of carnivores**

In carnivores, some species of Pinnipedia and Fissipedia have evolved relatively large body size, containing some extremely large species such as walrus, northern sea lion, weddell seal and polar bear, even weighed exceed 350kg. Due to the semi-aquatic life habits, species in Pinnipedia had evolved a large body size to increase body surface area to reduce the rapid heat loss in water [25]. Similarly, the polar bear is the largest extant bear to adapt to the cold Arctic regions, which are averagely weighed 372 kg [26]. Interestingly, polar bears and seals have relatively thick subcutaneous fat, which accounted for more than 30% of their body weight and were much higher than that of other wild carnivores [27-29]. It has been suggested that pinnipeds and polar bears are covered with a thick layer of fat to protect them from cold environment. In general, obesity refers to certain degree of overweight and a thick fat layer, which is a state caused by excessive fat accumulation [30], and percent body fat  $\geq 25\%$  for men and  $\geq 30\%$  for women is an indication of obesity [31].

Fourteen positively correlated BSAGs were identified in our study and variations of these genes were reported to cause obesity (*BRAP*, *CHCHD5*, *CPT1C*, *GPR1*, *LDLR*, *MAP2K5*, *PLEKHS1*, *SLC30A8*, *ST3GAL2*, *STX16*, *ZFH3*, *ZGRF1*, *ZNF395* and *ZPLD1*). For instance, SNPs in *BRAP* (BRCA1 associated protein) were shown to associate with obesity and other metabolic abnormalities [32]. And *STX16* (syntaxin 16) encoded a protein that is a member of the syntaxin or t-SNARE family and deletion in this gene may cause obesity and macrosomia in human [33]. Our result revealed that *MAP2K5* (mitogen-activated protein kinase kinase 5) was enriched in the GO cluster of "positive regulation of growth (GO:0045927)" and genetic variations in this gene were reported to cause childhood obesity [34]. Furthermore, the evolutionary rates of 14 obesity-related BASGs in large carnivores were higher than that of small ones. Specially, three obesity-related BASGs (i.e. *ST3GAL2*, *ZGRF1* and *ZPLD1*) were determined rapid evolution in extremely large carnivores. For instance, the evolutionary rate of *ZPLD1* (zona pellucida like domain containing 1) gene is 0.68632 in the extremely large carnivores, 5.5 times of that identified in the control group. Deletions in *ZPLD1* has been proved to contribute to genetic susceptibility of common childhood obesity [35]. And it was reported that *ST3GAL2*-null mice have an increase of 50% in fat mass and 9% in lean body mass [36]. For *ZGRF1* (Zinc finger GRF-type containing 1), this gene encodes a protein that contains GRF zinc finger (zf-GRF) and transmembrane domains, and a recent genome-wide and exome

chip association study revealed its association with adiposity [37]. Thus, the 14 obesity-related BSAGs identified in this study may contribute to increased body size and accumulated body fat for the large carnivores.

### **Molecular evidence for the Peto's Paradox in carnivores?**

Animal gigantism is a recurring phenomenon that is seemingly influenced by available resources and natural selection [19]. Being larger brings an array of advantages for an organism, however, there also exists associated biological tradeoffs, including the increased risk of developing cancer due to with more cells [38]. Surprisingly, empirical cancer rates did not vary with body size, large and long-lived animals have a lower risk of suffering cancer than smaller, shorter-lived animals, and this phenomenon was called the Peto's Paradox [20, 22, 39]. Recent years, the Peto's Paradox has been studied in many large mammals, such as the bowhead whale and humpback whale [40, 41]. Their genomes provided the following two major evidences related to cancer suppression: 1) multiple duplications of tumor suppressor genes; 2) positive selection in genes related to cancer and aging. It was suggested that large species might have evolved multiple mechanisms to suppress cancer.

Carnivores are relatively long-lived mammals, and generally speaking, species in Pinnipedia have a longer lifespan than species in Fissipedia. Some extremely large species in carnivores, such as walrus and polar bears were reported to live 40 or more years in the wild [42]. Within particular taxa in carnivores, body size and lifespan also seem to be positively correlated [21]. For instance, the sea otter (about 27.4kg) may live as long as 27 years in the wild, as compared with the ferret (about 975.6g) which has lifespan much shorter as 11.1 years [21]. The relatively large polar bears (maximum lifespan: 43.8 years) and brown bears (maximum lifespan: 40 years) also live longer than the smaller giant pandas (maximum lifespan: 36.8 years). However, the molecular mechanism for maintaining the longevity of the large carnivores is not very clear, and there have been very few relevant studies on cancer in carnivores so far.

In our study, it was interestingly identified a total of 100 BSAGs in carnivores which were related to cancer control process, including tumor suppressor, DNA repair, as well as immunity. In addition, we also identified 15 cancer-related REGs in extremely large carnivores. Among them, 21 BSAGs were determined as tumor suppressor genes in previous studies. For example, the *APC* (APC regulator of WNT signaling pathway) gene encoded a multidomain protein that played a crucial role in tumor suppression by antagonizing the WNT signaling pathway. Variants in *APC* would induce various kinds of cancer such as colorectal cancer, pancreatic cancer and so on [43, 44]. The *DSC3* (Desmocollin 3) gene is required for cell adhesion and desmosome formation, and it has been proved that this gene had a tumor suppressive activity through inhibition of AKT pathway in colorectal cancer [45]. *ZFX3* (zinc finger homeobox 3) was essential to regulate myogenic and neuronal differentiation, and it was reported to function as a tumor suppressor in several cancers [46]. The evolutionary rates of these three genes are significantly positively correlated with both body size parameters, suggesting the higher evolutionary rate in large carnivores than small ones. Meanwhile, two tumor suppressor genes, *ADAM11* and *TEP1*, exhibited rapid evolution in the extremely large carnivores. The evolutionary rate of *ADAM11* is 0.48527 in the extremely large

carnivores, 15.4 times of that identified in the background group. And this gene has been identified as tumor suppressor gene in human breast cancer [47]. The *TEP1* (Telomerase associated protein 1) product was a component of the ribonucleoprotein complex responsible for telomerase activity. And alterations in *TEP1* has been confirmed to cause several types of human tumors including brain, breast, prostate and lung cancer [48]. These two genes all had relatively higher evolutionary rates in extremely large carnivores and suggested enhanced ability for suppressing cancer.

Additionally, 16 BSAGs were found to be related with “DNA repair” and it was well known that deficits in DNA repair capacity might lead to genetic instability and carcinogenesis [49]. Recent study revealed that loss of a single allele of the *CLSPN* (claspin) gene in mice was sufficient to drive earlier tumorigenesis [50]. And *HELQ* (Helicase, POLQ like) played a critical role for replication-coupled DNA repair, germ cell maintenance and tumor suppression in mammals [51]. Importantly, 18 immunity related genes were identified in BSAGs of carnivores, among which three genes exhibited elevated evolutionary rates (*MAGT1*, *RFXANK* and *SKAP2*). For instance, *ITK* (IL2 inducible T cell kinase) functions downstream of the T-cell receptor and was important for T-cell activation, development, differentiation, and production of many pro-inflammatory cytokines [52]. Moreover, *ITK* has been found to be prevalent in all main types of human neoplasia [53, 54]. And the evolutionary rate of *MAGT1* (magnesium transporter 1) was 0.33127 in extremely large carnivores, 6.1 times of that in background group. Loss of *MAGT1* would disrupt T cell signaling and lead to a novel human primary immunodeficiency [55], and furthermore, overexpression of *MAGT1* was associated with development and metastasis of colorectal cancer [56]. Here, we obtained 100 cancer-related genes that were significantly associated with body size evolution in carnivores including 15 cancer-related REGs that identified in extremely large group, which might protect them from cancer invasion, special for large and long-lived species. These results might provide novel molecular evidence for the Peto’s Paradox with regard to carnivores.

### **Fixed amino acid changes in extremely small bod-sized carnivores contributing to growth restriction**

There are some extremely small species in carnivores, such as meerkat and ferret with body mass less than 1 kg and body length less than 50 cm. The smaller size brings multiple superiority to species. For instance, smaller carnivores could take great advantage of food resources that are not available to some large animals to ensure survival when the environment changes dramatically and have more free energy and time to engage in activities that increase mating and reproductive success [15-17].

Compared with other carnivores in our dataset, six fixed amino acid changes from six genes were identified in extremely small-body-sized group: *CDC7* (S513L), *ENG* (V367I), *LIG4* (S784P), *MMP2* (S579T), *TSPAN8*(S178T), and *POLE* (E701D). These six genes have been shown to be related with the phenotype of reduced body size. *CDC7* (cell division cycle 7) was very conserved through mammalian evolution but a fixed amino acid change (S513L) was identified in extremely small carnivores. *CDC7* played essential roles in initiation of mitotic DNA replication, and previous study showed that *CDC7*<sup>-/-</sup> or low expression of *CDC7* protein would lead to reduced body size with decreased cell proliferation in mice [57]. Importantly, the fixed changes (S513L) were located in the protein kinase domain that functioned as

an on/off switch for many cellular processes including metabolism, cell cycle progression, transcription [58]. Another fixed change (V367I) was examined in the Zona\_pellucida domain of *ENG* (endoglin) that was reported *Eng*<sup>-/-</sup> mice three times smaller than wild type mice at E10.5 of the embryonic period [59]. The unique amino acid mutation (S784P) was determined in the key domain (DNA ligase IV domain) of the *LIG4* (DNA ligase 4) gene that was reported mutations in humans or mice would cause growth failure and microcephaly, and this might be the result of activation of the DNA damage response, leading to a large amount of apoptosis during development [60]. Fixed amino acid mutation was separately found in the *MMP2* (S579T) and *POLE* (E701D) gene. Previous studies have shown *MMP2* (matrix metalloproteinase 2) knockout in mice and mutations of *POLE* (DNA polymerase epsilon, catalytic subunit) in human could cause short stature [61, 62]. Finally, we found a fixed difference at site 178 of the Tetraspanin domain in *TPSAN8* between extremely small carnivores and others, this gene was reported that genetic ablation of in mice caused a reduction (-15.6%) in the body weight of male fed a normal chow diet [63]. Furthermore, the changes of S513L in *CDC7* and S784P in *LIG4* exhibit transitions of polarity might cause radical changes in the three-dimensional structure and function of proteins [64]. These six unique changes examined in the extremely small carnivores might have restricted body size growth. Of course, function experiments are need to further test if these changes cause growth retardation.

## Conclusions

Mammalian order Carnivora exhibited the huge variation of body size, with more than 130,000 times in body mass and 50 times in body length, but their molecular mechanism remains poorly explored. Here we scanned the 20 genomes of representative carnivores and found a total of 337 genes to be associated with the body size. Our analyses showed 14 obesity-related genes and three rapidly evolving genes might drive the body size expansion. Importantly, our results provided molecular evidence for the Peto's paradox that lack of correlation between body size and cancer risk, based on 100 BSAGs associated with cancer control and 15 cancer-related genes under rapid evolution in carnivores. By contrast, 15 rapidly evolving genes and unique amino acid change of six genes might have restricted the growth of small carnivores. This study brought some new insights into the molecular mechanisms that drove the diversifying evolution of body size in carnivores, and provided new target genes for exploring the mysteries of body size evolution in mammals.

## Methods

### Phenotypic data and orthologous genes acquisition

High-quality genomes of 20 carnivores and an outgroup cow (*Bos taurus*) were downloaded from NCBI database. Eight families in Carnivora (i.e. Felidae, Canidae, Mustelidae, Phocidae, Otariidae, Odobenidae, Herpestidae, Ursidae) were chosen to include diverse body sizes. Two kinds of phenotype data, head body length (cm) and body mass (g), were collected from the PanTHERIA database [65] that were used for subsequent correlation analysis. For the missing head body length data of the domestic cat (*Felis*

*catus*) in the database, we obtained the alternative data from the Animal Diversity Web resource (<http://animaldiversity.org/>) [66]. Head body length means the length from the snout of the nose to the root of the tail for an animal [67]. All the phenotype data came from adult individuals. High-confidence “one-to-one” orthologous gene clusters were identified using the Orthofinder [68] pipeline, which applied all-against-all BLSATP algorithm. For genes with various transcripts, the longest coding sequence was used in our study. And if transcripts shorter than 150bp, or its lengths were not multiples of three were eliminated by our in-house Perl scripts. The sequences were aligned using Prank [69] at the codon level, with poorly aligned regions with gaps and non-homologous fragments removed using Gblocks [70] with relatively strict parameters (“-t=c, -b5=h”). High-quality multiple sequence alignment (MSA) files were used for subsequent analysis.

### **PGLS scanning the body-size-associated genes (BSAGs)**

PGLS implemented in the “Caper” package in R [71] was used to test the potential association between evolutionary rates of each gene and each phenotypic data (i.e. head body length and body mass). The ultrametric tree of 20 carnivores for PGLS analysis was obtained from the TimeTree website [72]. The Brownian motion model was applied and the phylogenetic signal ( $\lambda$  parameter) was tested by maximum likelihood (ML) method. The lambda ( $\lambda$ ) value was used as a quantitative measure of phylogenetic signals [73]. When  $\lambda$  was estimated to be 1 or near to 1, it indicated that these genes showed a strong phylogenetic signal. To obtain more stringent correlation *P* values, we further employed an extra two-step calibration procedure as suggested by Ma et al. [74]. On the basis of ‘*P value.all*’ from the regression analysis for all carnivores, the following two *P* values were calculated: 1) ‘*P value.robust*’ from regression repeated after discarding the species with largest residual error; 2) ‘*P value.max*’ from the PGLS on the remaining species, to calculate the maximal *P* value after dropping each species one at a time. Genes that were significantly related to both head body length and body mass under the most stringent standard (*P value.all* < 0.05, *P value.robust* < 0.05, *P value.max* < 0.05) are defined as body-size-associated genes (BSAGs).

The evolutionary rate  $\omega$  means the ratio of non-synonymous ( $d_N$ ) and synonymous substitutions ( $d_S$ ), which was estimated by *free-ratios* model (model = 1) implemented in the CODEML program of PAML 4.9e [75]. The root-to-tip  $\omega$  of each species was calculated by averaging the  $\omega$  from the ancestral carnivore to each terminal branch according to the method suggested by Montgomery et al. [76]. And if the value of  $d_N$  or  $d_S$  in each single  $\omega$  is less than 0.0002, we marked it as an outlier “*n/a*” to prevent its adverse effect on the integral root-to-tip  $\omega$ . And all root-to-tip  $\omega$  were  $\log_{10}$ -transformed to improve normality for regression analysis [76].

### **Functional enrichment analysis**

Functional annotation clustering tool Metascape (<http://metascape.org/>) [77] was used to perform Gene Ontology (GO) and KEGG pathway enrichment for the BSAGs list. GO categories were discovered and grouped into annotation clusters, against a background of human genome. All GO terms with enrichment

score (ES) > 1.3 (corresponding to a  $p$ -value less than 0.05) were considered for significant enrichment for list of BSAGs examined.

We also used literature searches, GWAS catalog, Human Phenotype Ontology (HPO; [78]), DisGeNET [79], RefSeq [46], Cancer Gene Census (CGC; [54]), Online Mendelian Inheritance in Man (OMIM; [80]) databases to explore potential biological functions of these single BSAG in association with body size.

### **Rapidly evolving genes (REGs) test**

To detect whether divergent selective pressure acting on carnivores with contrasting body size among the significant BSAGs determined above. We divided the 20 species into three groups, small body-sized carnivores (body mass < 12kg and body length < 1m), including five species: meerkat, *Suricata suricatta*, ferret *Mustela putorius furo*, domestic cat *Felis catus*, Canada lynx *Lynx canadensis* and red fox *Vulpes vulpes*, and extremely large body-sized carnivores (body mass > 350kg), such as polar bear *Ursus maritimus*, walrus *Odobenus rosmarus*, northern sea lion *Eumetopias jubatus* and weddell seal *Leptonychotes weddellii*, and the remaining 11 species as the third group. Two-ratio model (model = 2) that allows different  $\omega$  values within the foreground and background branches was used to evaluate the selective pressures in the small and extremely large groups, respectively. The two groups were separately regarded as foreground branches and the remaining 11 species as the background branches. The null model, i.e. one-ratio model (model = 0) assumed all branches have the same  $\omega$ . The likelihood ratio test (LRT) was used to compare nested likelihood models. If the  $\omega$  of the foreground is higher than that of the background branches with  $p < 0.05$ , we defined these genes as REGs.

### **Identification of fixed amino acid changes in extremely small/large body size groups**

To explore the effects of changes in single amino acid sites contribute to body size development, we scanned all the orthologous genes set for the fixed amino acid changes in extremely large group (i.e. polar bear, walrus, northern sea lion and weddell seal) and extreme small species (body mass < 1kg, i.e. Meerkat and Ferret, Bekoff et al. 1984). FasParser [81] was used to pick out the fixed amino acid changes in these two groups compared with other carnivores. For extremely small groups, the stoat *Mustela erminea* (mMusErm1.Pri) was added to improve the reliability of identified amino acid sites. Amino acid sites that were the same in the three extremely small species and were consistently different in other carnivores were selected. Gaps-containing sites were also excluded, and positions were corrected using human (*Homo sapiens*) amino acid sequences. Pfam 1.6 [82] was used to determine whether the changes were located in the functional domains of the protein.

## **Abbreviations**

BLAST: basic local alignment search tool; *BRCA1*:breast cancer type 1 susceptibility protein; BSAGs:body size associated genes; CMP:cytidine 5'-monophosphate; GO:Gene Ontology; GRF:growth hormone-releasing factor; GWAS:Genome-Wide Association Studies; KEGG:Kyoto Encyclopedia of Genes and Genomes; LRT:likelihood ratio test; PAML:phylogenetic analysis by maximum likelihood;

PGLS:phylogenetic generalized least squares; *POLQ*:DNA polymerase theta; REGs:rapidly evolving genes; SNARE:soluble NSF attachment protein receptor; SNP:Single Nucleotide Polymorphism

## 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 7 tables and 5 figures.

### Competing interests

The authors declare that they have no competing interests.

### Funding

This study was supported by the Key Project of the the National Natural Science Foundation of China (grant nos. 32030011, 31630071 to GY), NSFC (grant nos. 32070409, 31772448 to SX), National Key Programme of Research and Development, Ministry of Science and Technology (Grant number 2016YFC0503200 to GY and SX), the Qing Lan Project of Jiangsu Province to SX, and the Priority Academic Program Development of Jiangsu Higher Education Institutions to GY and SX.

### Authors' contributions

SX and GY conceived and designed this project. XH and TW performed the data analysis and wrote this paper. SX improved the manuscript. DS and XL put forward valuable opinions and suggestions, assisted with manuscript editing and polished the paper. All authors agree to publish this paper.

### Acknowledgements

We thank members of the Jiangsu Key Laboratory for Biodiversity and Biotechnology, Nanjing Normal University, for their contributions to this paper. The authors thank Mr. Xinrong Xu, Dr. Yulin Gai, Dr. Yuan Mu, Dr. Simin Chai, and Dr. Zhenpeng Yu for some helpful discussion. Many thanks to Dr. Ran Tian and Dr. Daiqing Yin for literature assistant and important advice. Special thanks to Dr. Di Sun and Dr. Weijian Guo for technical supports.

# References

1. Nowak RM, Walker EP. Walker's Mammals of the World, vol. 1: JHU press; 1999.
2. Meloro C, Clauss M, Raia P. Ecomorphology of Carnivora challenges convergent evolution. *Organisms Diversity & Evolution*. 2015;15(4):711-720.
3. Clauset A, Erwin DH. The evolution and distribution of species body size. *Science*. 2008;321(5887):399-401.
4. Kleiber M. Body size and metabolic rate. *Physiological reviews*. 1947;27(4):511-541.
5. McNab BK. Food habits, energetics, and the population biology of mammals. *The American Naturalist*. 1980;116(1):106-124.
6. McNab BK. Basal rate of metabolism, body size, and food habits in the order Carnivora. In: *Carnivore behavior, ecology, and evolution*. Springer; 1989;335-354.
7. Gittleman JL. Carnivore body size: ecological and taxonomic correlates. *Oecologia*. 1985;67(4):540-554.
8. Eisenberg JF. *The mammalian radiations: an analysis of trends in evolution, adaptation, and behaviour*; 1981.
9. Savage RJ. *Evolution in carnivorous mammals*. 1977.
10. Bekoff M, Daniels TJ, Gittleman JL. Life history patterns and the comparative social ecology of carnivores. *Annual review of ecology and systematics*. 1984;15:191-232.
11. King JE. *Seals of the world*: Comstock Pub. Associates; 1991.
12. Hofmeyr G. *Mirounga leonina*. The IUCN Red List of Threatened Species; Species Survival Commission: Gland, Switzerland. 2015.
13. Heptner V, Sludskii A. *Mammals of the Soviet Union*. Vol. II, part 1b, Carnivores (Mustelidae and Procyonidae). Washington, DC: Smithsonian Institution Libraries and National Science Foundation. 2002.
14. Van Valkenburgh B, Wayne RK. Carnivores. *Current biology*. 2010;20(21):R915-R919.
15. Moors PJ. Sexual dimorphism in the body size of mustelids (Carnivora): the roles of food habits and breeding systems. *Oikos*. 1980:147-158.
16. King CM. The advantages and disadvantages of small size to weasels, *Mustela* species. In: *Carnivore behavior, ecology, and evolution*. Springer; 1989;302-334.
17. Blanckenhorn WU. The evolution of body size: what keeps organisms small? *The quarterly review of biology*. 2000;75(4):385-407.
18. Hone DW, Benton MJ. The evolution of large size: how does Cope's Rule work? *Trends in Ecology & Evolution*. 2005;20(1):4-6.
19. Vermeij GJ. Gigantism and its implications for the history of life. *PLoS One*. 2016;11(1):e0146092.
20. Peto R, Roe F, Lee P, Levy L, Clack J. Cancer and ageing in mice and men. *British journal of cancer*. 1975;32(4):411-426.

21. Estes JA. Adaptations for aquatic living by carnivores. In: *Carnivore behavior, ecology, and evolution*. Springer; 1989;242-282.
22. Peto R. Epidemiology, multistage models, and short-term mutagenicity tests. *International Journal of Epidemiology*. 2016;45(3):621-637.
23. Bannasch DL, Baes CF, Leeb T. Genetic Variants Affecting Skeletal Morphology in Domestic Dogs. *Trends in genetics*. 2020.
24. Finarelli JA, Flynn JJ. Ancestral state reconstruction of body size in the Caniformia (Carnivora, Mammalia): the effects of incorporating data from the fossil record. *Systematic Biology*. 2006;55(2):301-313.
25. Clauset A. How large should whales be? *PloS one*. 2013;8(1):e53967.
26. Pond CM, Ramsay MA. Allometry of the distribution of adipose tissue in Carnivora. *Canadian Journal of Zoology*. 1992;70(2):342-347.
27. Pond CM. Morphological aspects and the ecological and mechanical consequences of fat deposition in wild vertebrates. *Annual Review of Ecology and Systematics*. 1978;9:519-570.
28. Mustonen A-M, Puukka M, Saarela S, Paakkonen T, Aho J, Nieminen P. Adaptations to fasting in a terrestrial mustelid, the sable (*Martes zibellina*). *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*. 2006;144(4):444-450.
29. Shero MR, Costa DP, Burns JM. Scaling matters: incorporating body composition into Weddell seal seasonal oxygen store comparisons reveals maintenance of aerobic capacities. *Journal of Comparative Physiology B*. 2015;185(7):811-824.
30. Apovian CM. Obesity: definition, comorbidities, causes, and burden. *Am J Manag Care*. 2016;22(7 Suppl):s176-185.
31. Shah NR, Braverman ER. Measuring adiposity in patients: the utility of body mass index (BMI), percent body fat, and leptin. *PloS one*. 2012;7(4):e33308.
32. Imaizumi T, Ando M, Nakatochi M, Yasuda Y, Honda H, Kuwatsuka Y, Kato S, Kondo T, Iwata M, Nakashima T. Effect of dietary energy and polymorphisms in BRAP and GHRL on obesity and metabolic traits. *Obesity Research & Clinical Practice*. 2018;12(1):39-48.
33. de Lange IM, Verrijn Stuart AA, van der Looij RB, Ploos van Amstel HK, van Haelst MM. Macrosomia, obesity, and macrocephaly as first clinical presentation of PHP1b caused by STX16 deletion. *American Journal of Medical Genetics Part A*. 2016;170(9):2431-2435.
34. Lv D, Zhang D-D, Wang H, Zhang Y, Liang L, Fu J-F, Xiong F, Liu G-L, Gong C-X, Luo F-H. Genetic variations in SEC16B, MC4R, MAP2K5 and KCTD15 were associated with childhood obesity and interacted with dietary behaviors in Chinese school-age population. *Gene*. 2015;560(2):149-155.
35. Molerés A, Martínez J, Martí A. Genetics of obesity. *Current Obesity Reports*. 2013;2(1):23-31.
36. Lopez PH, Aja S, Aoki K, Seldin MM, Lei X, Ronnett GV, Wong GW, Schnaar RL. Mice lacking sialyltransferase ST3Gal-II develop late-onset obesity and insulin resistance. *Glycobiology*. 2017;27(2):129-139.

37. Gao C, Wang N, Guo X, Ziegler JT, Taylor KD, Xiang AH, Hai Y, Kridel SJ, Nadler JL, Kandeel F. A comprehensive analysis of common and rare variants to identify adiposity loci in hispanic Americans: the IRAS Family Study (IRASFS). *PLoS One*. 2015;10(11):e0134649.
38. Nunney L. Size matters: height, cell number and a person's risk of cancer. *Proceedings of the Royal Society B*. 2018;285(1889):20181743.
39. 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.
40. 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 reports*. 2015;10(1):112-122.
41. 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*). *Molecular biology and evolution*. 2019;36(8):1746-1763.
42. De Magalhaes J, Costa J. A database of vertebrate longevity records and their relation to other life-history traits. *Journal of evolutionary biology*. 2009;22(8):1770-1774.
43. Horii A, Nakatsuru S, Miyoshi Y, Ichii S, Nagase H, Ando H, Yanagisawa A, Tsuchiya E, Kato Y, Nakamura Y. Frequent somatic mutations of the APC gene in human pancreatic cancer. *Cancer research*. 1992;52(23):6696-6698.
44. Sparks AB, Morin PJ, Vogelstein B, Kinzler KW. Mutational analysis of the APC/ $\beta$ -catenin/Tcf pathway in colorectal cancer. *Cancer research*. 1998;58(6):1130-1134.
45. Cui T, Yang L, Ma Y, Petersen I, Chen Y. Desmocollin 3 has a tumor suppressive activity through inhibition of AKT pathway in colorectal cancer. *Experimental Cell Research*. 2019;378(2):124-130.
46. O'Leary NA, Wright MW, Brister JR, Ciufu S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids research*. 2016;44(D1):D733-D745.
47. Yamamoto S, Higuchi Y, Yoshiyama K, Shimizu E, Kataoka M, Hijiya N, Matsuura K. ADAM family proteins in the immune system. *Immunology today*. 1999;20(6):278-284.
48. Kohno T, Takahashi M, Manda R, Yokota J. Inactivation of the PTEN/MMAC1/TEP1 gene in human lung cancers. *Genes, Chromosomes and Cancer*. 1998;22(2):152-156.
49. Goode EL, Ulrich CM, Potter JD. Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiology and Prevention Biomarkers*. 2002;11(12):1513-1530.
50. Hunter JE, Butterworth JA, Sellier H, Luli S, Floudas A, Moore AJ, Thomas HD, Campbell KJ, Kenneth NS, Chiremba RT. Regulation of checkpoint kinase signalling and tumorigenesis by the NF- $\kappa$ B regulated gene, CLSPN. *bioRxiv*. 2018:358291.
51. Adelman CA, Lolo RL, Birkbak NJ, Murina O, Matsuzaki K, Horejsi Z, Parmar K, Borel V, Skehel JM, Stamp G. HELQ promotes RAD51 paralogue-dependent repair to avert germ cell loss and

- tumorigenesis. *Nature*. 2013;502(7471):381-384.
52. Andreotti AH, Schwartzberg PL, Joseph RE, Berg LJ. T-cell signaling regulated by the Tec family kinase, Itk. *Cold Spring Harbor perspectives in biology*. 2010;2(7):a002287.
53. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. *Nature Reviews Cancer*. 2007;7(4):233-245.
54. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. A census of human cancer genes. *Nature reviews cancer*. 2004;4(3):177-183.
55. Li F-Y, Lenardo MJ, Chaigne-Delalande B. Loss of MAGT1 abrogates a Mg<sup>2+</sup> flux required for T cell signaling and leads to a novel human primary immunodeficiency. *Magnesium research: official organ of the International Society for the Development of Research on Magnesium*. 2011;24(3):S109.
56. Zheng K, Yang Q, Xie L, Qiu Z, Huang Y, Lin Y, Tu L, Cui C. Overexpression of MAGT1 is associated with aggressiveness and poor prognosis of colorectal cancer. *Oncology letters*. 2019;18(4):3857-3862.
57. Kim JM, Masai H. Genetic dissection of mammalian Cdc7 kinase: cell cycle and developmental roles. *Cell Cycle*. 2004;3(3):298-302.
58. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science*. 2002;298(5600):1912-1934.
59. Li DY, Sorensen LK, Brooke BS, Urness LD, Davis EC, Taylor DG, Boak BB, Wendel DP. Defective angiogenesis in mice lacking endoglin. *Science*. 1999;284(5419):1534-1537.
60. Klingseisen A, Jackson AP. Mechanisms and pathways of growth failure in primordial dwarfism. *Genes & development*. 2011;25(19).
61. Bauters D, Scroyen I, Van Hul M, Lijnen HR. Gelatinase A (MMP-2) promotes murine adipogenesis. *Biochimica et Biophysica Acta (BBA)-General Subjects*. 2015;1850(7):1449-1456.
62. Pachlopnik Schmid J, Lemoine R, Nehme N, Cormier-Daire V, Revy P, Debeurme F, Debré M, Nitschke P, Bole-Feysot C, Legeai-Mallet L. Polymerase  $\epsilon$ 1 mutation in a human syndrome with facial dysmorphism, immunodeficiency, livedo, and short stature ("FILS syndrome") Polymerase  $\epsilon$ 1 mutation in humans. *The Journal of experimental medicine*. 2012;209(13):2323-2330.
63. Champy M, Le Voci L, Selloum M, Peterson L, Cumiskey A, Blom D. Reduced body weight in male Tspan8-deficient mice. *International journal of obesity*. 2011;35(4):605-617.
64. Zhang J. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *Journal of molecular evolution*. 2000;50(1):56-68.
65. Jones KE, Bielby J, Cardillo M, Fritz SA, O'Dell J, Orme CDL, Safi K, Sechrest W, Boakes EH, Carbone C. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals: *Ecological Archives E090-184*. *Ecology*. 2009;90(9):2648-2648.
66. Myers P, Espinosa R, Parr C, Jones T, Hammond G, Dewey T. The animal diversity web. 2020. <https://animaldiversity.org/>. Accessed 30 Oct 2020.

67. Kolb H. Variation in the size of foxes in Scotland. *Biological Journal of the Linnean Society*. 1978;10(3):291-304.
68. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome biology*. 2015;16(1):157.
69. Löytynoja A. Phylogeny-aware alignment with PRANK. In: *Multiple sequence alignment methods*. Springer; 2014;155-170.
70. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution*. 2000;17(4):540-552.
71. Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N, Pearse W. Comparative analyses of phylogenetics and evolution in R. R package version 05. 2013;2.
72. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. *Molecular biology and evolution*. 2017;34(7):1812-1819.
73. Pagel M. Inferring the historical patterns of biological evolution. *Nature*. 1999;401(6756):877-884.
74. Ma S, Yim SH, Lee S-G, Kim EB, Lee S-R, Chang K-T, Buffenstein R, Lewis KN, Park TJ, Miller RA. Organization of the mammalian metabolome according to organ function, lineage specialization, and longevity. *Cell metabolism*. 2015;22(2):332-343.
75. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*. 2007;24(8):1586-1591.
76. Montgomery SH, Capellini I, Venditti C, Barton RA, Mundy NI. Adaptive evolution of four microcephaly genes and the evolution of brain size in anthropoid primates. *Molecular biology and evolution*. 2011;28(1):625-638.
77. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature communications*. 2019;10(1):1-10.
78. Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gourdine J-P, Gargano M, Harris NL, Matentzoglou N, McMurry JA. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic acids research*. 2019;47(D1):D1018-D1027.
79. Piñero J, Ramírez-Anguaita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, Furlong LI. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic acids research*. 2020;48(D1):D845-D855.
80. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic acids research*. 2005;33(suppl\_1):D514-D517.
81. Sun Y-B. FasParser: a package for manipulating sequence data. *Zoological research*. 2017;38(2):110.
82. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A. The Pfam protein families database: towards a more sustainable future. *Nucleic*

## Additional Files

**Table S1.** Genome version and phenotypic data for 20 carnivores used in our study. **Table S2.** Significant results for the gene-phenotype association tests for both head body length and body mass ( $p$  value.all/robust/max < 0.05). **Table S3.** GO (Biological Process) and KEGG enrichment results for positively and negatively correlated BSAGs. **Table S4.** Biological roles of fourteen positively correlated BSAGs in Obesity. **Table S5.** Biological Function of 100 cancer-control related BSAGs. **Table S6.** Roles in cancer for Twenty-one tumor suppressor genes that are significantly associated with the evolution of body size in carnivores. **Table S7.** Rapidly evolving genes in small and extremely large carnivores

## Figures

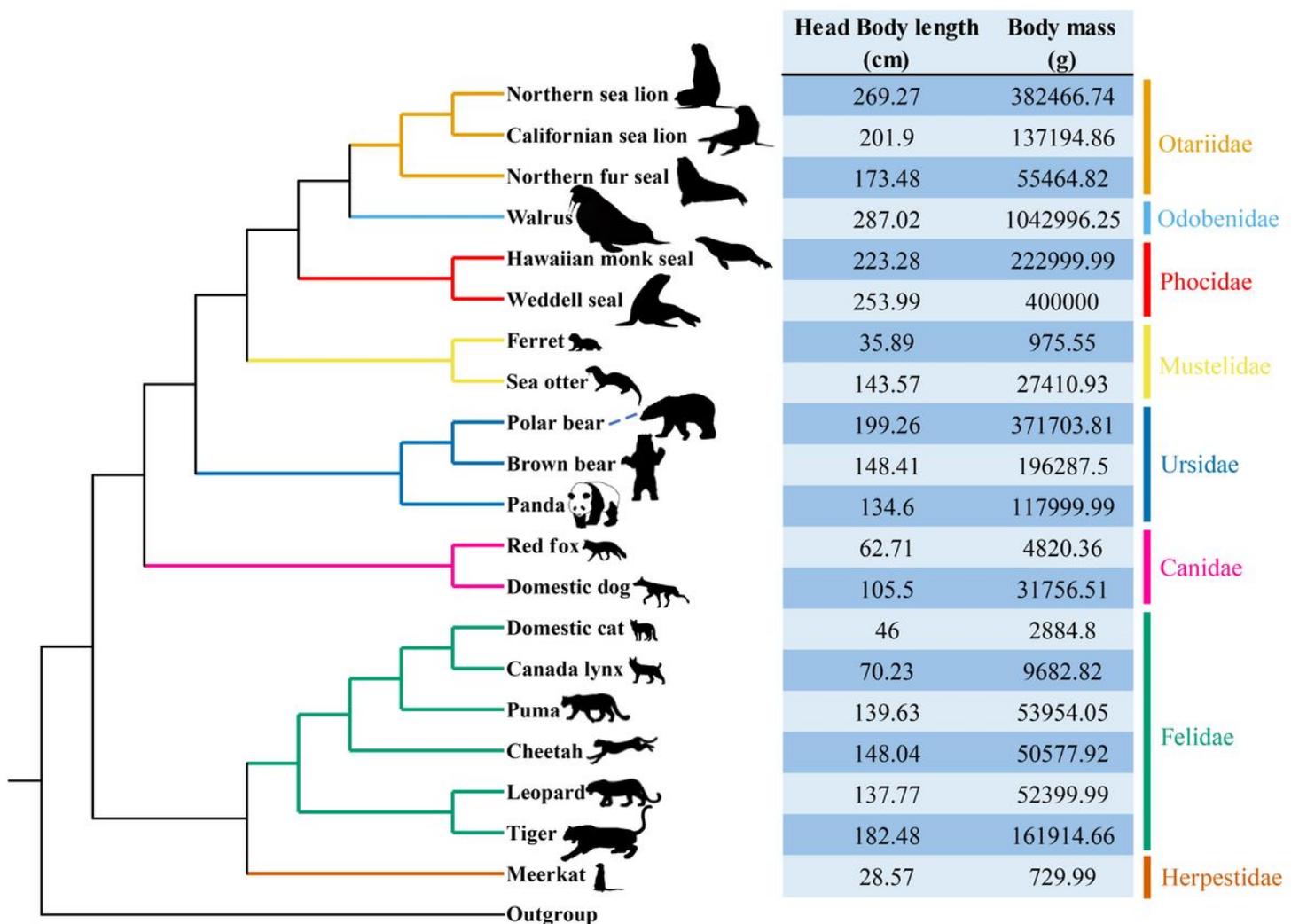


Figure 1

Phylogenetic tree and phenotype data of 20 carnivores used in this study. Phylogenetic tree and phenotypic data including head body length (cm) and body mass (g) of 20 carnivores were shown in this figure. All phenotypic data were collected from PanTHERIA database and all animal illustrations were taken from phylopic.org.

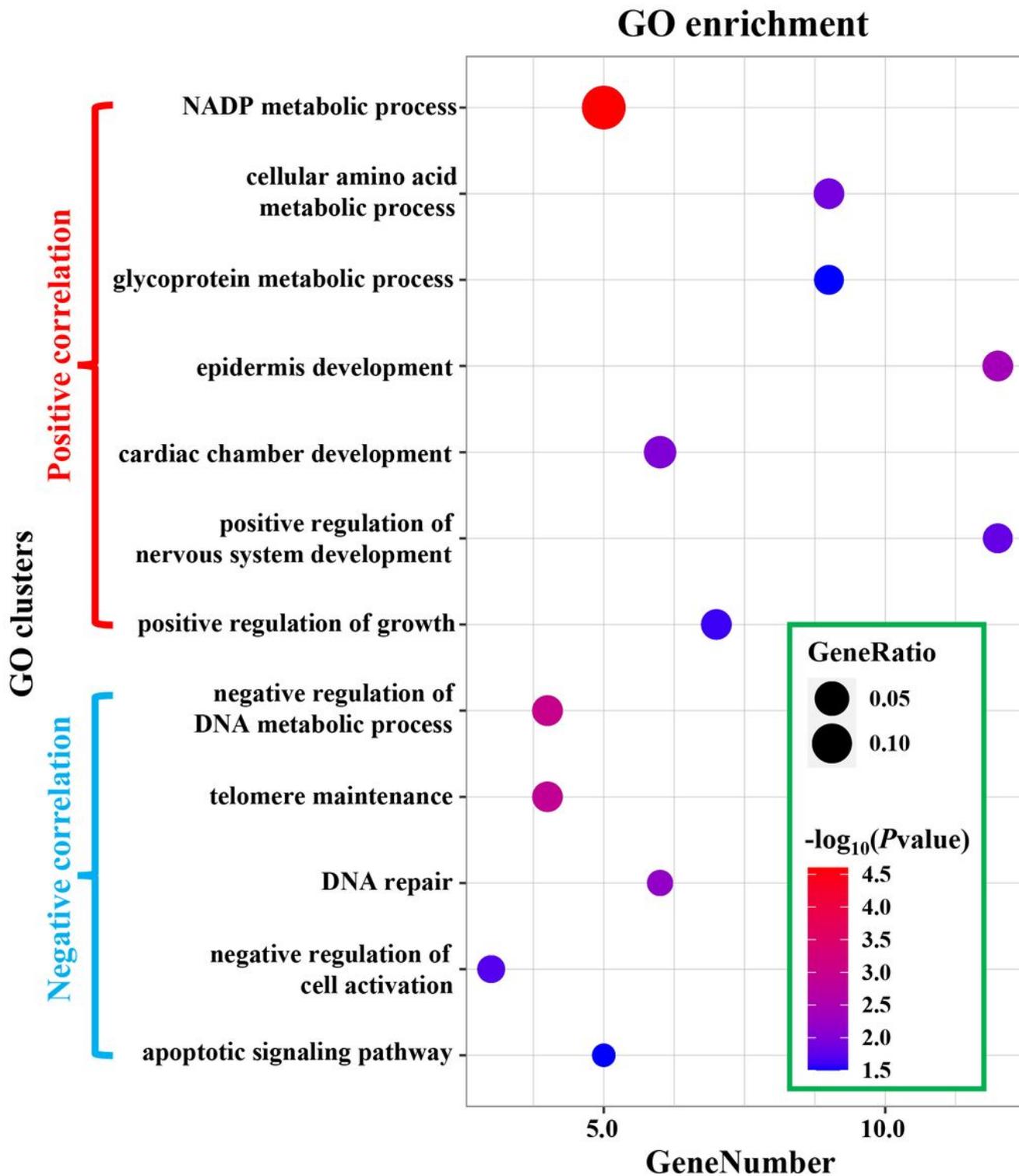
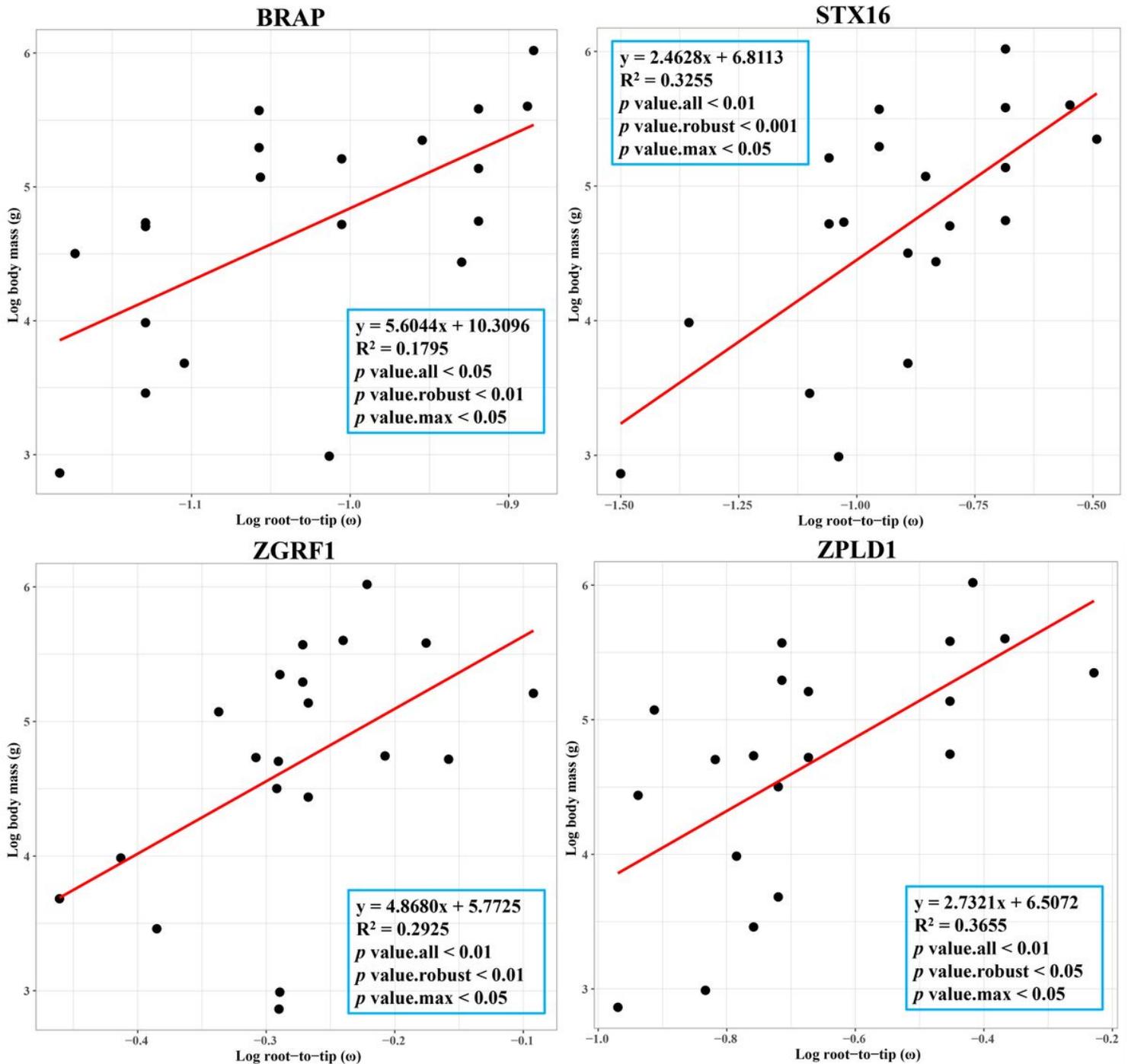


Figure 2

Functional analysis of BSAGs based on the GO enrichment. GO clusters with a p-value < 0.05 significantly enriched in positively or negatively correlated BSAGs. The circle size represents the “GeneRatio”, which means the ratio of enriched genes to all genes in this cluster. The color of the circles indicates the p-value of GO clusters. The left y-axis represents the GO clusters, and the x-axis shows the gene number.



**Figure 3**

Regression analyses between root-to-tip  $\omega$  and body mass. Regression analyses between root-to-tip ( $\omega$ ) and body mass (g) by PGLS in R for four obesity-related BSAGs: BRAP, STX16, ZGRF1 and ZPLD1.

## Rapidly evolving genes (REGs)

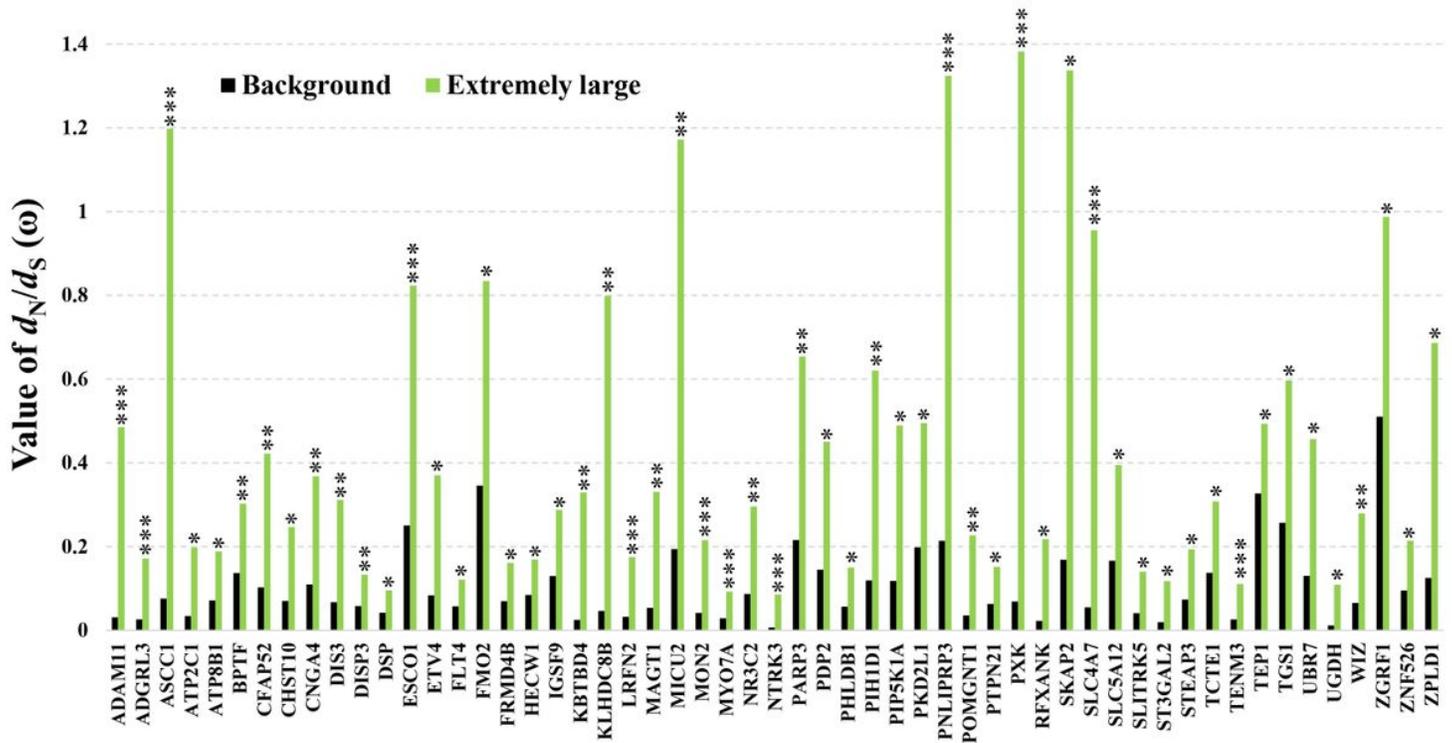
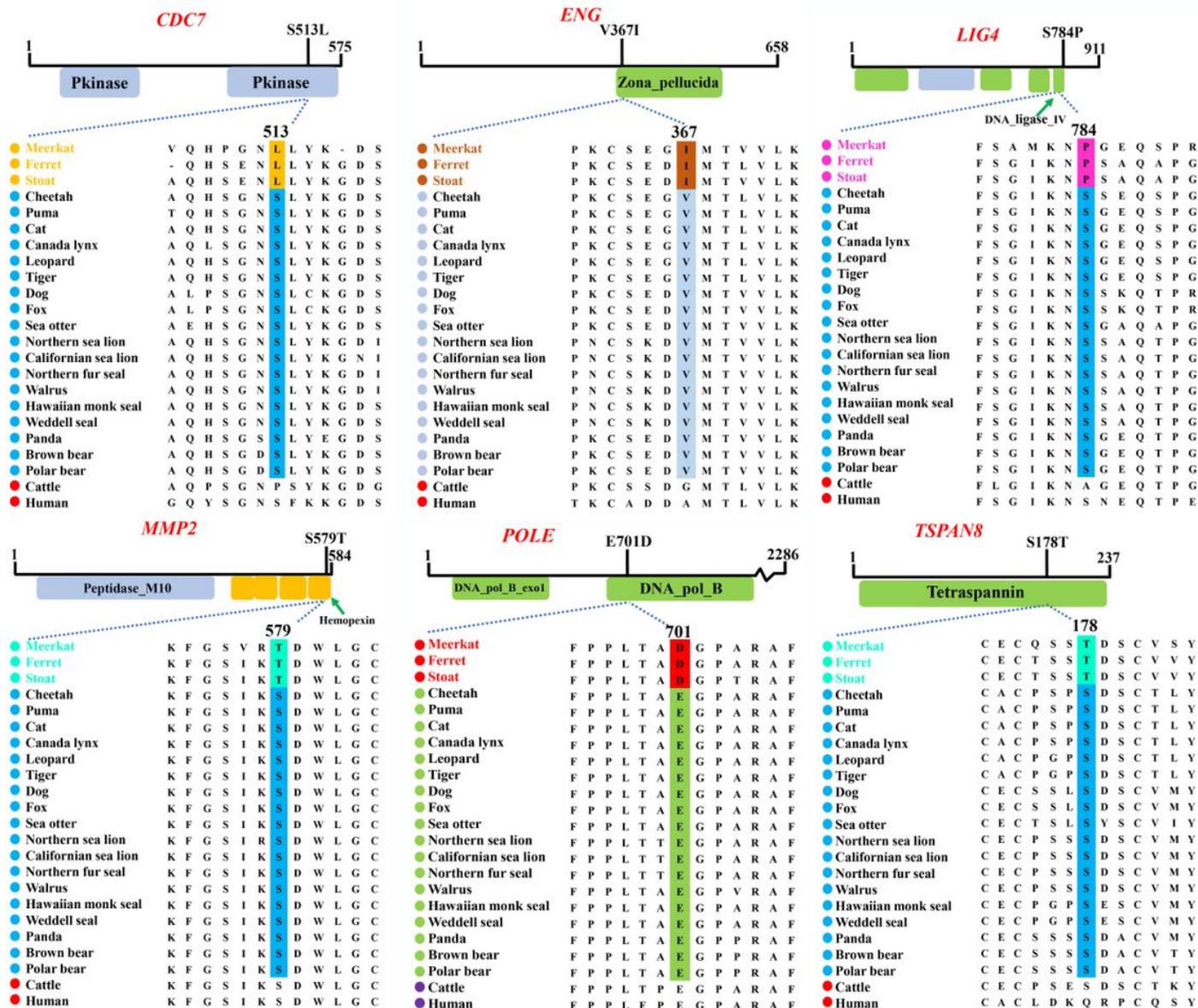


Figure 4

Rapidly evolving genes (REGs) in extremely large body size groups of carnivores. The background group are labeled as "grey", meanwhile, the extremely large carnivores are labeled by "green". The left y-axis in histogram represents the value of  $d_N/d_S$  ( $\omega$ ). Significance is represented by several asterisks, where "\*" indicates a p-value less than 0.05, "\*\*" indicates a p-value less than 0.01 and "\*\*\*" represents a p-value less than 0.001.



**Figure 5**

Fixed amino acid changes in six reduced body size-related genes of extremely small carnivores. The fixed amino acid changes of extremely small carnivores observed in six genes that related to the phenotype of reduced body size. These six sites are located in the functional domain of the CDC7, ENG, LIG4, MMP2, POLE, TSPAN8 gene, respectively.

## Supplementary Files

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

- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)

- [TableS4.docx](#)
- [TableS5.docx](#)
- [TableS6.docx](#)
- [TableS7.xlsx](#)