Gene Duplication and Loss of AANAT in Mammals Driven by Rhythmic Adaptations

Daiqing Yin, RuRu Zhou, Mengxin Yin, Yue Chen, Shixia Xu,* and Guang Yang*

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

*Corresponding authors: E-mails: xushixia78@163.com; gyang@njnu.edu.cn.

Abstract

Arylalkylamine N-acetyltransferase (AANAT) plays a crucial role in synchronizing internal biological functions to circadian and circannual changes. Generally speaking, only one copy of AANAT gene has been found in mammals, however, three independent duplications of this gene were detected in several cetartiodactyl lineages (i.e., Suidae, Hippopotamidae, and Pecora), which originated in the middle Eocene, a geological period characterized with the increased climate seasonality. Lineage-specific expansions of AANAT and the associated functional enhancement in these lineages strongly suggest an improvement in regulating photoperiodic response to adapt to seasonal climate changes. In contrast, independent inactivating mutations or deletions of the AANAT locus were identified in the four pineal-deficient clades (cetaceans, sirenians, xenarthrans, and pangolins). Loss of AANAT function in cetaceans and sirenians could disrupt the sleep-promoting effects of pineal melatonin, which might contribute to increasing wakefulness, adapting these clades to underwater sleep. The absence of AANAT and pineal glands in xenarthrans and pangolins may be associated with their body temperature maintenance. The present work demonstrates a far more complex and intriguing evolutionary pattern and functional diversity of mammalian AANAT genes than previously thought and provides further evidence for understanding AANAT evolution as driven by rhythmic adaptations in mammals.

Key words: AANAT, gene duplication, gene loss, rhythmic adaptations, cetartiodactyls.

Introduction

The circadian and circannual rhythms of organisms reflect significant adaptations for coping with the daily and annual fluctuations of external environments (Yerushalmi and Green 2009; Gwinner 2012), and daylength (photoperiod) is the most powerful environmental cue (Bradshaw and Holzapfel 2007). The circadian system is responsible for measuring and transducing of photoperiod alterations, whereas the circannual mechanism fulfills the important function of the timely prediction of the changes in the food supply, weather conditions, and predator activity, etc. By helping anticipate environmental changes, endogenous rhythms allow for synchronizing major biological functions, including sleep, migration, and reproduction to environmental cycles, thus enabling organisms to maximize survival and use of resources (Hut and Beersma 2011).

The pineal organ and its hormone melatonin play a pivotal role in the transduction of photoperiod information to synchronize daily and seasonal rhythms (Falcón et al. 2009). As a hormone of darkness, melatonin is characterized by having circulating levels that are much higher at night than in the day (Utiger 1992; Hut and Beersma 2011). Moreover, melatonin acts as a critical input signal of timing system, involving in the circadian organization of physiological status as well as in photoperiodic measurement (Kumar 2017). Furthermore, the duration and amplitude of nocturnal melatonin levels vary with seasonal photoperiodic variations. The changing duration of melatonin provides a dominant time cue for seasonal behavioral cycles (Bartness and Goldman 1989; Lincoln 2006). Additionally, the rhythmicity of pineal melatonin is a hallmark of internal time in mammals, but some taxa such as cetaceans, sirenians, pangolins, and xenarthrans (i.e., anteaters, sloths, and armadillos) seem to lack pineal glands (Ralph 1975; Kappers 1983; Panin et al. 2012).

The regulation of melatonin rhythm primarily depends on the expression level and enzyme activity of arylalkylamine N-acetyltransferase (AANAT), the rate-limiting enzyme in melatonin biosynthesis (Klein 2007; Pevet and Challet 2011). The dramatic increase in the expression and activity of AANAT at nighttime leads to a significant elevation in melatonin levels, whereas weak activity during the daytime severely limits the amount of substrates available for conversion to melatonin (Klein et al. 1996; Ganguly et al. 2002). Animals exposed to light at night experience a rapid decline in AANAT activity, followed by a large decline in melatonin levels (Klein and Weller 1970). Therefore, AANAT has become the master regulator for adjusting the endogenous melatonin levels in response to external photoperiodic changes and thus plays a crucial part in regulating rhythmic processes (Klein 2007; Kumar 2017).
It was found that a knockout of the AANAT gene in zebrafish caused a lack of melatonin and the abolishment of circadian sleep regulation (Gandhi et al. 2015).

Preliminary investigations have shown that kinetic differences and the different temperature–activity relationships of multiple AANATs in teleosts were likely to represent evolutionary responses to different environmental pressures (Zilberman-Peled et al. 2004, 2011). The evolution of avian AANAT regions was potentially driven by the selection for nocturnality (Fidler et al. 2004). As for mammals, it has been well established that only one copy of an AANAT gene has been conserved within mammalian genomes (Saha et al. 2019). However, recent analyses identified double AANAT-like sequences in three artiodactyl species, including the hippopotamus (Hippopotamus amphibius), cattle (Bos taurus), and goat (Capra hircus) (Kim et al. 2016; Lopes-Marques et al. 2019). By contrast, manatees and several cetaceans have been reported to have lost the AANAT gene (Huelsmann et al. 2019; Lopes-Marques et al. 2019). Thus, the evolution of the AANAT locus across mammals remains an alluring area of inquiry. Similarly, our understanding of the association between AANAT evolution and rhythmic adaptations remains limited.

In the present study, a comprehensive survey of the AANAT locus in 256 mammalian lineages was conducted to uncover the evolutionary characteristics of the AANAT gene across mammalian phylogeny. Interestingly, we observed the duplications of AANAT in certain artiodactyl clades and inactivations of AANAT in pineal-deficient clades, which was in contrast to only one functional AANAT gene in other mammals. Gene duplication and loss are widely considered as important genetic sources for functional innovations and evolutionary adaptations (Zhang et al. 2002; Sharma, Hecker, et al. 2018). Therefore, we further explored what evolutionary forces have shaped the diversification of mammalian AANAT genes. Phylogenetic trees and the ancestral state of all AANAT sequences were first reconstructed to elucidate gene duplication events within Cetartiodactyla, and we then evaluated whether selection pressures relaxed in lineages with inactivated AANAT genes. Finally, AANAT activity assays for represented species were performed to explore the potential functional modifications of AANAT genes in mammals with different evolutionary patterns.

Results

Evolutionary Characteristics of AANAT Genes in Mammals

To obtain a broader picture of AANAT evolution in mammals, we identified AANAT sequences from 174 mammalian species with high-quality genome assemblies, including 12 artiodactyl species. Most mammals were found to have a single copy of the AANAT gene. Phylogenetic trees of AANAT sequences showed a major congruence with commonly acknowledged phylogeny, indicating that AANAT genes were orthologous (supplementary fig. S1, Supplementary Material online). Interestingly, besides AANAT gene duplications identified in three artiodactyls (i.e., hippopotamus, cattle, and goat) which were reported previously, two AANAT sequences were also detected in other five artiodactyl species (fig. 1), including pig (Sus scrofa) and four ruminants (reindeer Rangifer tarandus, Thomson’s gazelle Eudorcas thomsonii, giraffe Giraffa camelopardalis, and sheep Ovis aries). To further explore the evolution characteristics of the AANAT gene in artiodactyls, additional 82 artiodactyl species covering all major families were added to our analyses. The results showed that one copy of AANAT gene was only observed in three families, including Camelidae, Tragulidae, and Antilopidae, whereas two adjacent AANAT copies could be identified in the vast majority of ruminants, hippopotamus, and two species of Suidae (fig. 2). Significant sequence similarity (a range of 82.8–99.2%) was found between two AANAT copies within each species. However, nonsynonymous substitution sites were also identified, most of which caused radical amino acid changes and were located within or adjacent to important functional domains, and may lead to functional divergence. For example, pairwise comparison of cattle AANAT sequences showed a total of 10 amino acid changes and two of them (i.e., E43K and Q71R) were located in the binding pocket of AANAT.

Furthermore, three copies of AANAT were observed in two species of Bovidae, including P. re David’s deer Elaphurus davidianus (NCBI: GCA_002443075.1) and gembok Orny gazelle (GCA_006410575.1). However, only two copies were identified in P. re David’s deer when using another independently sequenced genome (http://animal.nwsuaf.edu.cn/code/index.php/RGD, last accessed May 21, 2021), which left open the possibility that the different copy numbers may be sequencing errors or artifacts arising from genome assembly. Furthermore, PCR amplification and sequencing further corroborated that P. re David’s deer had only two copies of the AANAT gene. Furthermore, the premature stop codons were detected in five artiodactyls (chiru Pantholops hodgsonii, okapi Okapia johnstoni, hirola Beatragus hunteri, Bohar reedbuck Redunca redunca, and bush duiker Sylvicapra grimmia), rendering one of the two AANAT copies putatively nonfunctional, while these genes did not exhibit any shared inactivation mutations and their closely related species retained two functional copies.

Three Independent Duplications of the AANAT Gene in Artiodactyls

As AANAT gene duplications were identified specifically in artiodactyl species, we focused on the evolutionary relationships of artiodactyl AANAT genes to infer duplication events. Similar tree topologies were generated when phylogenetic trees of the AANAT gene were reconstructed using the maximum likelihood (ML) and Bayesian inference (BI) (fig. 3). Phylogenetic reconstructions for coding and noncoding sequences both showed that AANAT copies in pig and hippopotamus formed their respective clades, supporting the independent duplications in these two clades (fig. 3). However, the coding and noncoding trees presented different patterns for AANAT evolution in pereons. In the phylogenetic tree based on coding sequences, duplicated AANAT copies formed respective monophyletic clades within...
multiple species (fig. 3A and supplementary fig. S2, Supplementary Material online), implying gene duplications after pecoran species divergence. By contrast, in the noncoding tree, pecoran AANAT copies were clearly divided into two distinct groups (fig. 3B and supplementary fig. S3, Supplementary Material online), suggesting one single-gene duplication event occurred in the common ancestor of all pecoran ruminants. Actually, the findings based on noncoding sequences were corroborated by the phylogenetic analyses based on concatenated coding and noncoding sequences (supplementary fig. S4, Supplementary Material online), which split pecoran AANAT1/AANAT2 copies into two separate groups with high nodal supports (BS/PP = 100/1). Furthermore, we also detected that gene conversion might have affected the coding regions of AANAT copies in most artiodactyls using the software RDP4 (supplementary table S4, Supplementary Material online), which could have led to the phylogenetic confusion of some lineages in the coding analyses. Thus, according to the phylogenetic trees of noncoding sequences and concatenated coding and noncoding sequences, three independent duplications were clearly identified in artiodactyls, that is, the ancestral lineages of Suidae (i.e., Entelodontidae), Hippopotamidae (i.e., Anthracotheriidae), and pecorans (i.e., Gelocidae).

The ancestral character of the copy number at the relevant nodes was estimated using the ML approach (fig. 4 and supplementary fig. S5, Supplementary Material online). For the node of Artiofabula, as well as the node of Whippomorpha and the ancestral node for cetacean clades, the estimated copy numbers were 1.357 (95% CI: 0.695–2.019), 1.387 (95% CI: 0.741–2.033), and 1.107 (95% CI: 0.537 - 1.678), respectively, implying that no AANAT duplications had occurred in these ancestral nodes. Notably, the earliest increased copy number was recorded in the ancestral node of Pecora, with
Fig. 2. The repertoires and synteny of AANAT genes in cetartiodactyl species with available genome sequences. Topology of the tree is based on previous studies (Hassanin et al. 2012; Bibi 2013; Chen et al. 2019). AANAT genes synteny is schematically shown by different colored polygons pointing out the orientation in the genome, and the conserved RHBD2 upstream and UBE2O downstream genes are shown in gray. Unfilled polygons indicate target region could not be obtained completely. Distances between genes or gene length are not drawn to scale.
an ancestral character estimate of 1.818 (95% CI: 1.412–2.225), implying that expansions of the AANAT copy number occurred before the radiation of pecoran ruminants.

Additionally, we estimated the ages of divergence between AANAT copies using BI implemented in the software package BEAST (supplementary fig. S7, Supplementary Material online).
online). It was indicated that the divergence times for AANAT sequences in Suidae and Pecora were 46.94 and 47.60 Ma, respectively, which were congruent with the appearance of their ancestors during the Middle Eocene (37.2–48.6 Ma), whereas the age for the divergence of hippopotamus AANAT copies was estimated to be 22.77 Ma, which was in the early Miocene (15.9–23 Ma).

Inactivations of the AANAT Gene in All Four Pineal-Deficient Mammalian Lineages

Our genome alignments revealed that one or more inactivating mutations (premature stop codons, frameshift indels, altered start codons, or splice site mutations) were presented in all placental taxa without detectable pineal gland (i.e., cetaceans, sireniants, pangolins, and xenarthrans) (fig. 1). Inactivation mutations were found to be located in the first exon of the AANAT gene (three coding exons, 207 residue protein in humans), which may have led to a functional loss in these lineages. For cetaceans, all species had one nonsense mutation at the end of exon 1 that led to premature stop codons, and additional inactivation mutations were also detected in the most toothed whales, with even the sperm whale Physeter macrocephalus showing the deletion of entire exon 1 (supplementary fig. S8, Supplementary Material online). These disruptive mutations in AANAT genes were further confirmed by polymerase chain reaction (PCR) amplification and sequencing of AANAT genes from six cetacean species. Similar results were also found in two species of Sirenia. For example, the Florida manatee Trichechus manatus had an ACG initiation codon mutation and a 13-bp frameshift deletion in exon 3, and Steller’s sea cow Hydrodamalis gigas had a 22-bp frameshift deletion in exon 1 (supplementary fig. S9, Supplementary Material online). For Pholidota, two premature stop codons were detected in all three pangolins, of which one premature stop codon was located at positions 28–30 in exon 1 and another was located at positions 211–213 in exon 2 (supplementary fig. S10, Supplementary Material online). Furthermore, the largest number of frameshift mutations were found in xenarthran species, including the nine-banded armadillo Dasypus novemcinctus (13), the three-banded armadillo Tolypeutes matacus (9), the Hoffmann’s two-fingered sloth Choloepus hoffmanni (7), the Brown-throated sloth Bradypus variegatus (5), and the southern two-toed sloth Choloepus didactylus (4). These mutations spread across all exons of AANAT and inactivated its reading frame (supplementary fig. S11, Supplementary Material online). Finally, Blast searches failed to show any evidence for the presence of AANAT in two anteaters (giant anteater Myrmecophaga tridactyla and southern tamandua T. tetradactyla), whereas the entire sequence for conserved flanking genes (RHBDF2, UBE2O) could be recovered.

Next, selection analyses were performed to test whether AANAT evolved neutrally in the lineages with inactivating mutations (table 1). Indeed, an \( \omega \) value of 0.382 for all four gene-loss clades was significantly higher than the \( \omega \) value of 0.170 observed for other background branches with intact sequences (\( P < 0.001 \)), suggesting that AANAT evolved under relaxed selection in these pseudogenized branches. The values for crown Cetacea and crown Xenartha were 0.840 (\( P = 0.820 \)) and 1.504 (\( P = 0.700 \)), respectively, which were both close to the expected value of 1.000 for the neutrally evolving in crown branches subsequent to inactivation of AANAT gene on the stem branches. The crown Pholidota had an \( \omega \) value of 0.471 that was significantly higher relative to the background value (\( P < 0.001 \)) but was different from the expected value for neutral evolution (\( P = 0.046 \)). As for crown Sirenia, the slightly elevated \( \omega \) of 0.155 was not significantly different from the background value of 0.148 (\( P = 0.485 \)), suggesting inactivation of AANAT gene was a young event.

We further estimated the inactivation time for AANAT in each lineage (supplementary table S5, Supplementary Material online). The results revealed that the cetacean inactivation of AANAT occurred around 42.2–45.5 Ma in the common ancestor of the crown Cetacea. For crown Xenartha, the loss was estimated to have occurred in the ancestor of xenarthrans that lived 60–72 Ma. Within Pholidota, we estimate that AANAT was inactivated around 27.3–30.4 Ma. By contrast, inactivation dates for crown Sirenia were estimated in the more recent range of 0.5–0.6 Ma, allowing for the possibility that there has been a lag time between the cessation of purifying selection and the accumulation of the first inactivating mutation in this lineage.

Enzyme Assay for AANAT Genes in Different Mammalian Lineages

To determine whether different evolutionary characteristics of AANAT were accompanied by functional changes, we performed in vitro enzymatic assays for nine species from three categories, including those with single AANAT genes (C3H/HeJ mice C3, dromedary Camelus dromedarius, and Java mouse-deer Tragulus javanicus), two AANAT copies (pig, hippopotamus, cattle, and reindeer), and mutant AANAT genes (minke whale Balenoptera acutorostrata and CS7BL/6 mice B6) (fig. 6). The kinetic constants of the recombinant AANATs were determined using serotonin as a substrate and calculated using the Michaelis–Menten equation (supplementary fig. S12 and table S6, Supplementary Material online). We found that the AANATs from the four artiodactyls with two AANAT copies showed significantly higher activity than those from the three species with a single copy of AANAT (\( P = 0.036 \)). For AANATs from species with the two copies, AANAT2s were found to exhibit significantly higher activity than AANAT1s (\( P = 0.029 \)). In addition, enzyme activities were significantly different among the four artiodactyl species with AANAT duplicates, with reindeer showing the strongest activities for acetylating substrates, whereas hippopotamus showed the weakest activities. Furthermore, in the three species with single-copy AANAT, dromedary, and lesser mouse-deer showed higher catalytic capacities than C3 mice (\( P = 0.033 \)). In contrast, the minke whale, which has a pseudogenized AANAT gene, showed nearly undetectable AANAT activity, whereas B6 mice with a truncated AANAT protein had a relatively weak activity.
Table 1. LRTs of Various Models on the Selective Pressures on AANAT.

| Models | \( \Omega \) | \(-\ln L\) | np | Models Compared | 2\( \Delta(\ln L) \) | P Values |
|--------|-------------|--------|----|----------------|-----------------|----------|
| Data set I | A: all branches have one \( \omega \) | 0.175 | 20,201.823 | 320 |        |        |
|         | B: all branches have one \( \omega = 1 \) | 1.000 | 21,458.178 | 319 | A vs. B | 2,512.710 | 0 |
|         | C: all branches with pseudogenized AANAT \( \omega_2 = 0.170, \omega_1 = 0.382 \) | 0.000 | 20,192.811 | 321 | A vs. C | 18.024 | 2.18137E−05 |
|         | D: all branches with pseudogenized AANAT \( \omega_2 = 0.169, \omega_1 = 1.000 \) | 0.000 | 20,203.316 | 320 | C vs. D | 21.101 | 4.56892E−06 |
|         | E: each branch has its own \( \omega \) |        |        |        |        |        |
|         | Variable \( \omega \) | 19,922.015 | 318 | C vs. E | 541.592 | 4.6171E−117 |
| Data set II | A: all branches have one \( \omega \) | 0.159 | 17,906.143 | 300 |        |        |
|         | B: all branches have one \( \omega = 1 \) | 1.000 | 19,149.248 | 299 | A vs. B | 2,486.210 | 0 |
|         | C: the crown Cetacea has \( \omega_2 = 0.158, \omega_1 = 0.840 \) | 0.000 | 17,902.075 | 301 | A vs. C | 8.136 | 0.004 |
|         | D: the crown Cetacea has \( \omega_2 = 1, \omega_1 = 0.000 \) |        |        |        |        |        |
|         | Variable \( \omega \) | 17,902.101 | 300 | C vs. D | 0.052 | 0.820 |
| Data set III | A: all branches have one \( \omega \) | 0.148 | 16,542.044 | 252 |        |        |
|         | B: all branches have one \( \omega = 1 \) | 1.000 | 17,811.403 | 251 | A vs. B | 2,538.718 | 0 |
|         | C: the crown Sirenia branch has \( \omega_2 = 0.148, \omega_1 = 0.000 \) | 0.000 | 16,541.800 | 253 | A vs. C | 0.488 | 0.485 |
|         | D: the crown Sirenia branch has \( \omega_2 = 1, \omega_1 = 0.000 \) |        |        |        |        |        |
|         | Variable \( \omega \) | 16,551.960 | 252 | C vs. D | 20.320 | 6.55122E−06 |
| Data set IV | A: all branches have one \( \omega \) | 0.154 | 16,877.636 | 254 |        |        |
|         | B: all branches have one \( \omega = 1 \) | 1.000 | 18,113.721 | 253 | A vs. B | 2,472.170 | 0 |
|         | C: the crown Pholidota has \( \omega_2 = 0.152, \omega_1 = 0.471 \) | 0.000 | 16,871.939 | 255 | A vs. C | 11.394 | 0.000736817 |
|         | D: the crown Pholidota has \( \omega_2 = 1, \omega_1 = 0.000 \) |        |        |        |        |        |
|         | Variable \( \omega \) | 16,873.920 | 254 | C vs. D | 3.962 | 0.046 |
| Data set V | A: all branches have one \( \omega \) | 0.158 | 17,621.820 | 258 |        |        |
|         | B: all branches have one \( \omega = 1 \) | 1.000 | 18,865.766 | 257 | A vs. B | 2,487.892 | 0 |
|         | C: the crown Xenarthra has \( \omega_2 = 0.156, \omega_1 = 1.504 \) | 0.000 | 17,616.249 | 259 | A vs. C | 11.142 | 0.000843948 |
|         | D: the crown Xenarthra has \( \omega_2 = 1, \omega_1 = 0.000 \) |        |        |        |        |        |
|         | Variable \( \omega \) | 17,616.323 | 258 | C vs. D | 0.148 | 0.700 |

Discussion

In the present study, we scanned the AANAT gene against a total of 256 mammalian species and provided a comprehensive evolutionary analysis of this gene across the entire mammalian phylogeny. Most mammalian lineages have only one copy of AANAT, however, besides previously identified AANAT duplication events in hippopotamus, cattle, and goat, novel duplications in three clades (i.e., Suidae, Hippiopotamidae, and Pecora) of Artiodactyla were also identified in the present study. And intriguingly, besides cetaceans and sirenians, pangolins and xenarthrans also possessed inactivation mutations in AANAT genes. These four mammalian taxa with inactivated AANAT genes all lack pineal glands and show particular biological rhythms. Enzyme assays further corroborated the functional diversification of AANAT across mammalian lineages. Our study provides some novel insights into gene duplication and loss in mammals driven by rhythmic regulation to adapt to seasonal climate changes and specific ecological niches.

Three Independent Duplications of AANAT in Artiodactyls in Response to Seasonal Climate Changes

A complicated evolutionary pattern of AANAT was identified in artiodactyl species, with two copies found in pig, hippopotamus, and pecoran lineages, whereas other lineages had one copy. Phylogenetic trees and copy number of ancestral state reconstructions in artiodactyl AANAT sequences suggested that three independent duplication events had occurred along the ancestral lineages, respectively, of Suidae (i.e., Entelodontidae), Hippiopotamidae (i.e., Anthracotheriidae), and pecorans (i.e., Gelocidae). Fossil records revealed that three ancestors with duplicated AANAT all appeared in the Middle Eocene (Prothero and Foss 2007; Vislobokova 2008; Hackmann and Spain 2010; Scherler et al. 2019), a period characterized by global cooling and increased climate seasonality (Agusti et al. 2003; Mosbrugger et al. 2005). This was supported by the present estimates of time for the AANAT gene duplication in Suidae and Pecora which well coincided with the period of Middle Eocene. By contrast, the divergence of hippopotamus AANAT copies was estimated to occur 22.77 Ma, not in the Middle Eocene but in the Early Miocene, a period with gradually increased climate seasonality (Bruch et al. 2007, 2011). The appearance of seasonal climates drove the spread of deciduous forests and thus brought about the beginning of seasonal differences in the availability and abundance of vegetation (Janis 1989; Woodburne 2004). Consequently, herbivores were faced with big challenges to survive during the annual food shortage periods. Through detecting and transducing photoperiodic alterations, the timing mechanisms mediated by melatonin allow for triggering physiological behaviors at
the appropriate period, providing the organisms with time to prepare and survive in advance of climate changes (Kumar 2017). Furthermore, accurate anticipation is crucial for determining favorable breeding time to optimize the survival of offspring. For instance, the timing of artiodactyl parturition generally coincides with seasonal plant growth (Grzimek 2003; Martin et al. 2004). Therefore, the expansion of the AANAT gene, together with enhanced regulation of photoperiodic responses may thus confer advantages in dealing with the predictable climate changes to conquer unfavorable resource conditions. This is congruent with the hypothesis that duplications of the AANAT gene contributed to better seasonal adaptation as revealed in insects and teleosts (García-Allegue et al. 2001; Barberà et al. 2013).

The AANATs of artiodactyls showed significantly higher activities than those of mice, and furthermore, duplicated AANAT copies (i.e., AANAT1 and AANAT2) showed enhanced activity relative to single-copy AANAT genes (fig. 5). In general, AANAT activity exhibited a close positive relationship with the amount of produced melatonin (Ebihara et al. 1997; Klein 2006; Byeon and Back 2016). This increased enzyme activity of AANAT duplicates strongly suggested that artiodactyls might have an enhanced capacity for melanin secretion and improved their ability to cope with seasonal changes. In light of previous studies that gene conversion often promoted the beneficial increase in dosage of genes that mediate the interaction between organism and environment (Innan and Kondrashov 2010; Meslin et al. 2011), the extensive gene conversion events identified in the evolution of artiodactyl AANAT genes could further supported a potentially important roles of an increased amount of AANAT gene expression in artiodactyles. In addition, it was also noticed that AANAT activity differed between artiodactyles with two copies, such as reindeer, which showed the highest AANAT activity in this study, whereas hippopotamus showing the lowest. Although reindeers live in polar photic conditions, they can still determine seasonal timing accurately by monitoring the indiscernible photoperiod changes in their environments (Skogland 1989; Stokkan et al. 2007), which was supported, to some degree, by the substantially high activity of their two AANAT copies that would be beneficial for quickly adjusting melatonin level. In contrast, the hippopotamus, an example of species inhabiting the tropics with a less marked seasonality, mainly used the rainfall as a more reliable seasonal marker (Estes 2012; Ogutu et al. 2014), which might have led to decreased activities of the reserved two copies.

Inactivation of AANAT and Pineal-Deficiency Contribute to Escaping from the Circadian Regulation

Pineal gland, the primary source of circulating melatonin, is a major endocrine component in the regulation of circadian behaviors. Surgical removal of the pineal gland in animals can abolish the circadian behavior and reduce sleep duration (Kocher et al. 2006; Trivedi et al. 2016). During the
mammalian evolution, four taxa (cetaceans, sirenians, pangolins, and xenarthrans) have been reported to lack of discernable pineal glands (Ralph 1975; Kappers 1983), and correspondingly, independent inactivating mutations or missing of AANAT locus were identified in these four clades in the present study. This result may have been associated with their particular features in circadian rhythm to adapt to specific ecological niches. Cetaceans and sirenians, as fully aquatic marine mammals, have evolved an irregular circadian organization of activity and sleep. Most obviously, they adopt a unique unihemispheric slow-wave sleep, allowing one cerebral hemisphere to remain awake enough while the other one still benefits from sleep (Lyamin et al. 2008; Aulsebrook et al. 2016). This is favorable for maintaining active states in water to achieve the necessity to breathe at the surface and more efficient monitoring of the environment. The inactivation mutations of cetacean AANAT and its undetectable enzyme activity indicated the functional loss of the AANAT gene in this lineage. Knockout of the AANAT gene in zebrafish significantly reduces almost half of their nighttime sleep (Gandhi et al. 2015). The inactivation of AANAT and the loss of pineal glands in cetaceans and sirenians disrupt the production of melatonin that can promote sleep, thus contributing to increased wakefulness to adapt to completely aquatic environments. In contrast, xenarthrans and pangolins display markedly convergent traits, featuring predominantly by nocturnality, imperfect temperature regulatory systems, and low, labile body temperatures (McNab 1980; Gilmore et al. 2001). They have synchronized their activity rhythm primarily to fluctuations in ambient temperature rather than photoperiodic changes to achieve thermal balance (Ciné et al. 2015; Attias et al. 2018). Besides modulating biological rhythms, the increased circulating levels of melatonin have been shown to decrease core body temperature through enhancing distal heat loss (Saarela and Reiter 1994; Kräuchi et al. 2006). Thus, the absence of AANAT and pineal glands may be associated with adaptive changes in circadian rhythmicity, affecting activity patterns and body temperature maintenance (Cagnacci et al. 1992). Overall, the disruptions of AANAT function together with a deficiency of pineal glands would be accordingly advantageous for escaping from the circadian regulation of daily activity to accommodate adaptations for specific ecological niches and life history traits.

Conclusion

The present study demonstrated a more complex and intriguing evolutionary pattern of mammalian AANAT genes than previously recognized, which yielded some new insights into the evolution and function of this gene. Remarkably, we identified and characterized new AANAT gene sequences from artiodactyls based on sequence analyses and functional activity assays. Our findings further revealed independent gene duplications in three ancestral lineages of artiodactyls, including Suinae, Hipposaltidae, and Pecora. The coincidence between gene duplication and the enhanced seasonal climate during the Middle Eocene, in combination with the increased activity of duplicated AANAT copies, suggested that AANAT gene duplication is an advantage for artiodactyls to evolve photoperiodic response to adapt to seasonal climate changes. Additionally, it is interesting to find gene deletion or inactivation in multiple mammalian lineages that have lost pineal glands and exhibit irregular circadian features, including fully aquatic cetaceans and sirenians, pangolins, and xenarthrans. This study could provide some novel insights into mammalian AANAT evolution as driven by or associated with rhythmic adaptations in mammals.

Materials and Methods

Gene Identification

Using previously published intact AANAT gene sequences from human, mouse, panda as queries, we conducted Blastn (version 2.5.0) (Camacho et al. 2009) searches in local databases constructed using 174 mammalian genome sequences, including 12 artiodactyl species, retrieved from the National Center for Biotechnology Information (NCBI) website (https://www.ncbi. nlm.nih.gov/, last accessed May 21, 2021). The coding sequences of AANAT were determined manually after removing all introns following the canonical AG/GT rule for splicing (Burset et al. 2000). To further explore the evolution of AANAT in Artiodactyla, we then identified the AANAT gene sequences from 82 additional artiodactyl genomes retrieved from the NCBI and Ruminant Genome Database (http://animal.nwsuaf.edu.cn/code/index.php/RGD, last accessed May 21, 2021) using the camel AANAT gene sequence (XM_010983838.2) as a query. Detailed information concerning each genome assembly is described in supplementary table S1, Supplementary Material online. All the identified AANAT genes were classified into three categories as follows: intact genes, partial genes, and pseudogenes, according to amino acid alignment and Blast results. The Blast hits with a length of <200 bp were discarded. If a target gene was not found, the genomic region between the four flanking genes (SPHK1, UBE2O, RHBDL2, and CYGB) was inspected, manually to ensure the continuity or disruption of each genomic region (with presence or absence of Ns). We failed to detect AANAT sequences and recover the continuous flanking sequences in three artiodactyls, which may have resulted from either incomplete genome sequencing or poor genome assembly. The unavailable AANAT fragments from 14 artiodactyl species were excluded from subsequent analysis (supplementary table S1, Supplementary Material online). Phylogenetic relationships and divergence times of species were obtained from the Timetree website (www.timetree. org, last accessed May 21, 2021) (Kumar et al. 2017) and multiple references (Boissiere et al. 2005; McGowen et al. 2009; Meredith et al. 2011; Song et al. 2012; Janis and Theodor 2014; Upham et al. 2019; Zurano et al. 2019).

To validate the inactivation mutations of AANAT in cetaceans and the copy number of AANAT genes in artiodactyls, we additionally sequenced the AANAT gene from six cetacean species (one mysticetes and five odontocetes) and three artiodactyl species: minke whale, baiji Lipotes vexillifer, finless porpoise Neophocaena asiaeorientalis, Indopacific hump-backed dolphin Sousa chinensis, striped dolphin Stenella...
The estimated ancestral states were mapped on the phylogeny multiple times. Each assembled sequence was counted as variations, aiming to avoid inclusion of the same considering both Taq polymerase-derived errors and allelic bled at 99% similarity level to allow for 1% mutations by universal primer pair. The resulting sequences were assembled in the pMD19-T vector (TaKaRa). For each PCR product, at least species, the purified PCR products were cloned into the species for estimating the divergent time of two species set as 38–50 Ma (Hassanin et al. 2012; Bibi 2013), the split between dog Canis familiaris and giant panda Ailuropoda melanoleuca set at 42–48 Ma (Sato et al. 2009; dos Reis et al. 2012), the split between large flying fox Pteropus vampyrus and greater horseshoe bat Rhinolophus ferrumequinum set at 55.6–60.5 Ma (Eick et al. 2005; Yu et al. 2014), and that between human Homo sapiens and Rhesus monkey Macaca mulatta set at 27.95–31.35 Ma (Wildman et al. 2003; Yang and Rannala 2006).

Phylogenetic Reconstruction

Coding, noncoding, and combined phylogenetic trees were constructed through the ML approach with IQ-TREE (Trifinopoulos et al. 2016) and Bayesian methods implemented in MrBayes v.3.2.3 (Ronquist et al. 2012). The upstream/downstream regions with a length of <1,000 bp were discarded in the noncoding and combined analyses. The optimal model of sequence substitution was determined by MrModeltest 2.3 (Nylander 2004) using the Akaike Information Criterion. BI analysis was run for 30,000,000 iterations of a Markov chain Monte Carlo (MCMC) algorithm, with six simultaneous chains, and trees were sampled every 1,000 generations. Support for the nodes and parameter estimates were derived from a majority rule consensus of the last 15,000 trees sampled after convergence. For the ML analyses, the best tree was reconstructed with 10,000 bootstrap replications. Gene conversion events were also detected using five methods implemented in the RDP4 software package (Martin et al. 2015).

The ancestral states for the number of AANAT gene copies were reconstructed based on the ML approach using the fastAnc function in the phytools R package (Revell 2012). The estimated ancestral states were mapped on the phylogeny using the contMap function in phytools (supplementary fig. S5, Supplementary Material online), while the 95% confidence intervals (CIs) for point estimates of ancestral states were plotted on a traitgram using the fancyTree function from the same package (supplementary fig. S6, Supplementary Material online).

The noncoding data were used for estimating the divergence times of two AANAT copies using Bayesian method implemented in BEAST v1.10.4 (Suchard et al. 2018). In BEAUti (from BEAST package), nucleotide substitution model was set to “HKY,” and the clock model was set to “uncorrelated relaxed molecular clock” (Drummond et al. 2006) with “lognormal” distribution and “Yule Process” of speciation (Gernhard 2008) was set as Tree prior. MCMC was run independently for five times with each run for ten million generations with sampling for every 1,000 generations. LogCombiner v1.10.4 was used for removing a burnin of 2,000 from each of the five tree files and for generating a combined tree file. A best tree was generated with all the annotations incorporated using TreeAnnotator v1.10.4. The combined log file was imported to Tracer v1.7.1 for assessing the effective sample size (ESS). It was found to be >200 for all the parameters. The final tree and divergence times were visualized using FigTree v1.4.4. Four calibration points were used in the present study: the divergent time of Ruminantia set as 38–50 Ma (Hassanin et al. 2012; Bibi 2013), the split between dog Canis familiaris and giant panda Ailuropoda melanoleuca set at 42–48 Ma (Sato et al. 2009; dos Reis et al. 2012), the split between large flying fox Pteropus vampyrus and greater horseshoe bat Rhinolophus ferrumequinum set at 55.6–60.5 Ma (Eick et al. 2005; Yu et al. 2014), and that between human Homo sapiens and Rhesus monkey Macaca mulatta set at 27.95–31.35 Ma (Wildman et al. 2003; Yang and Rannala 2006).

Selective Pressure Analyses

The rate ratios (ω) of nonsynonymous (dN) to synonymous (dS) substitutions were calculated using codon-based ML models implemented in the CODEML program of PAML 4.8a (Yang 2007). All frameshift insertions in the alignments were discarded, and frameshift deletions and nonsense mutations were recoded as missing data prior to analysis. First, we used one-ratio (A) model that assumed a uniform ω ratio for all branches. Next, we tested two-ratio (C) model that allowed two ω ratios that differed between the background branches and the target branches. Then, one-ratio (B) model with a fixed ω = 1.0 for all branches and two-ratio (D) model with a fixed ω = 1 on the branches of interest were used as null hypotheses. Finally, we tested a free-ratio (E) model that allowed an independent ω value for each branch. A likelihood ratio test (LRT) with a χ2 distribution was used to evaluate which models were statistically different from the null model at a threshold of P < 0.05.

Estimation of Inactivation Times

To date the loss of AANAT in different lineages, we used the methods described in previous studies (Sharma, Hecker, et al. 2018; Sharma, Lehmann, et al. 2018). The inactivation time (Tn) was estimated by the following equation:

\[ T_n = T \times \frac{K - K_i}{1 - K_i}, \]

where T represents the time since the split from the last common ancestor. We used the lower and upper bound of the CI for the species divergence time T from TimeTree
(Kumar et al. 2017). The Ka/Ks value estimated for pseudogenic branches was referred to as K, whereas the Ka/Ks value for mammals with a functional AANAT was referred to as Ks. Ka is the number of nonsynonymous substitutions per non-synonymous site, whereas Ks is the number of synonymous substitutions per synonymous site. The Ka/Ks values were determined by the CODEML program.

Recombinant Protein Preparation
The recombinant enzymes from seven cetartiodactyl species, including cattle, hippopotamus, pig, reindeer, dromedary, Java mouse-deer, and minke whale were expressed and purified in the present study. Mice are extensively used in transgenic models and biology studies including sleep, circadian rhythm, and melatonin research (Uz and Manev 2001; Sharma, Sahota, et al. 2018). Moreover, it has been demonstrated that there are two kinds of laboratory mouse strains, melatonin-deficient (B6) and melatonin-proficient strains (C3). B6 mice have a naturally truncated AANAT protein as a result of aberrant splicing whereas C3 mice retain an intact and functional AANAT gene (Roseboom et al. 1998). Thus, we chose AANAT genes from these two mouse strains as control groups in this study. Coding regions were inserted into the bacterial expression vector pGEX4T1 and verified by Sanger sequencing. This allowed for the construction of AANAT proteins fused to glutathione S-transferase. The way of expression, production, and purification of AANATs has been described previously (Pavlicek et al. 2008; Falcón et al. 2014).

Enzymatic Activity Assays
The colorimetric assay for the AANAT enzyme activity was based on the quantification of Coenzyme A generated during acetyl transfer (Falcón et al. 2014; Paulin et al. 2015). Typically, reactions were performed in 96-well plates, where each well contained 0.5 μg of AANAT enzyme in a 100 μl final volume of phosphate buffer (0.1 M, pH 6.8) solution including acetyl-CoA (3). B6 mice have a naturally truncated AANAT protein as a result of aberrant splicing whereas C3 mice retain an intact and functional AANAT gene (Roseboom et al. 1998). Thus, we chose AANAT genes from these two mouse strains as control groups in this study. Coding regions were inserted into the bacterial expression vector pGEX4T1 and verified by Sanger sequencing. This allowed for the construction of AANAT proteins fused to glutathione S-transferase. The way of expression, production, and purification of AANATs has been described previously (Pavlicek et al. 2008; Falcón et al. 2014).

Enzymatic Activity Assays
The colorimetric assay for the AANAT enzyme activity was based on the quantification of Coenzyme A generated during acetyl transfer (Falcón et al. 2014; Paulin et al. 2015). Typically, reactions were performed in 96-well plates, where each well contained 0.5 μg of AANAT enzyme in a 100 μl final volume of phosphate buffer (0.1 M, pH 6.8) solution including acetyl-CoA (3). B6 mice have a naturally truncated AANAT protein as a result of aberrant splicing whereas C3 mice retain an intact and functional AANAT gene (Roseboom et al. 1998). Thus, we chose AANAT genes from these two mouse strains as control groups in this study. Coding regions were inserted into the bacterial expression vector pGEX4T1 and verified by Sanger sequencing. This allowed for the construction of AANAT proteins fused to glutathione S-transferase. The way of expression, production, and purification of AANATs has been described previously (Pavlicek et al. 2008; Falcón et al. 2014).

Supplementary Material
Supplementary data are available at Molecular Biology and Evolution online.

Acknowledgments
This work was supported by the National Natural Science Foundation of China (NSFC) (grant nos. 32030011 to G.Y., 31772448 to S.X., 31630071 to G.Y., and 32070409 to S.X.), the Qinglan Project of Jiangsu Province to S.X., National Key Programme of Research and Development, Ministry of Science and Technology (grant no. 2016YFC0503200 to G.Y. and S.X.), and the Priority Academic Program Development of Jiangsu Higher Education Institutions to G.Y. and S.X.

Author Contributions
G.Y. and S.X. conceptualized and supervised the study. D.Y., M.Y., and Y.C. collected the data and conducted the bioinformatics analyses. D.Y. and R.Z. performed AANAT activity experiments. D.Y. prepared the original draft, and G.Y. and S.X. revised the manuscript. All authors read and approved the final manuscript.

Data Availability
There are no data to be archived. All data were included in supplementary materials for readers to validate.

References
Agustí J, De Siria AS, Garcés M. 2003. Explaining the end of the hominoid experiment in Europe. J Hum Evol. 45(2):145–153.
Attias N, Oliveira-Santos LGR, Fagan WF, Mourão G. 2018. Effects of air temperature on habitat selection and activity patterns of two tropical imperfect homoeotherms. Anim Behav. 140:129–140.
Aulsebrook AE, Jones TM, Rattenborg NC, Roth IIT, Lesku JA. 2016. Sleep ecophysiology: integrating neuroscience and ecology. Trends Ecol Evol. 31(8):590–599.
Barberà M, Menglau B, Collantes-Alegre JM, Cortés T, González A, Martínez-Torres D. 2013. Identification, characterization and analysis of expression of genes encoding arylalkylamine N-acetyltansferases in the pea aphid Acyrthosiphon pisum. Insect Mol Biol. 22(6):623–634.
Bartness TJ, Goldman BD. 1989. Mammalian pineal melatonin: a clock for all seasons. Experimentia 45(10):939–945.
Bibi F. 2013. A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. BMC Evol Biol. 13(1):166–115.
Boissière J-R, Lihoreau F, Brunet M. 2005. The position of Hippopotomidae within Cetartiodactyla. Proc Natl Acad Sci U S A. 102(5):1537–1541.
Bradshaw WE, Holzapfel CM. 2007. Evolution of animal photoperiodism. Annu Rev Ecol Evol Syst. 38(1):1–25.
Bruch AA, Uhl D, Mosbrugger V. 2007. Miocene climate in Europe: patterns and evolution: a first synthesis of NELCIME. Palaeogeogr Palaeoc. 253(1–2):1–7.
Bruch AA, Utescher T, Mosbrugger V. 2011. Precipitation patterns in the Miocene of Central Europe and the development of continentality. Palaeogeogr Palaeoc. 304(3–4):202–211.
Burset M, Seledtsov IA, Solovyev VV. 2000. Analysis of canonical and non-canonical splice sites in mammalian genomes. Nucleic Acids Res. 28(21):4364–4375.
Byeon Y, Back K. 2016. Low melatonin production by suppression of either serotonin N-acetyltansferase or N-acetylserotonin methyl-transferase in rice causes seedling growth retardation with yield penalty, abiotic stress susceptibility, and enhanced coleoptile growth under anoxic conditions. J Pineal Res. 60(3):348–359.
Cagnacci M, Elliott JA, Yen SS. 1992. Melatonin: a major regulator of the circadian rhythm of core temperature in humans. J Clin Endocrinol Metab. 75(2):447–452.
Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10(1):421–429.
Chen L, Qu Q, Jiang Y, Wang K, Lin Z, Li Z, Bibi F, Yang Y, Wang J, Nie W, et al. 2019. Large-scale ruminant genome sequencing provides...
