Article

An Ancient Mutation in the *TPH1* Gene is Consistent with the Changes in Mammalian Reproductive Rhythm

Chenhui Liu [1,2,†], Xunping Jiang [1,2,†], Guiqiong Liu [2,*], Teketay Wassie [1,2] and Shishay Girmay [1,2]

[1] Laboratory of Small Ruminant Genetics, Breeding and Reproduction, College of Animal Science and Technology, Huazhong Agricultural University, Wuhan 430070, China; liu890621@webmail.hzau.edu.cn (C.L.); xpjiang@mail.hzau.edu.cn (X.J.); teketay@webmail.hzau.edu.cn (T.W.); shishay@webmail.hzau.edu.cn (S.G.)

[2] Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of the Ministry of Education, Wuhan 430070, China

* Correspondence: liuguiqiong@mail.hzau.edu.cn; Tel./Fax: +86-27-87585120

† These authors contributed equally to this work.

check for updates

Abstract: The reproductive rhythm undergoes several changes during the evolution of mammals to adapt to local environmental changes. Although the critical roles of melatonin (MLT) in the formation of reproductive rhythm have been well established, the genetic basis for the changes of reproductive rhythm remains uncertain. Here, we constructed the phylogenetic trees of 13 melatonin synthesis, metabolism and receptor genes, estimated their divergence times, and calculated their selection pressures. Then, we evaluated the effect of positively selected and functionally related mutations on protein activity. Our results showed that there were significant positive selection sites in the three major genes, including tryptophan hydroxylase 1 (*TPH1*), tryptophan hydroxylase 2 (*TPH2*) and indoleamine-2,3-dioxygenase 1 (*IDO1*) that are involved in melatonin synthesis, metabolism and function. At the protein level, amino acids at the 442nd site of TPH1 protein and the 194th, 286th, 315th and 404th sites of IDO1 protein were under positive selection, and the variants of the amino acid in these sites might lead to the changes in protein function. Remarkably, the 442nd site of these positive selection sites is in the tetramerization domain of TPH1 protein, and it is proline or leucine. At this site, 89.5% of the amino acid of non-seasonal reproducing mammals was proline, while that of 88.9% of seasonal reproducing mammals was leucine. This variation of the amino acid was derived from the T/C polymorphism at the 1325th site of the *TPH1* gene coding sequence, which significantly altered the TPH1 activity ($p < 0.01$). Interestingly, the predicted age of the allele C in the mammalian genome appeared about 126.6 million years ago, and allele T appeared about 212.6 million years ago, indicating that the evolution of the *TPH1* gene was affected by the two mammalian split events and the K-T extinction event. In conclusion, the T/C polymorphism at the 1325th site in the *TPH1* gene coding sequence altered TPH1 activity, suggesting that this polymorphism is consistent with the reproductive rhythm of mammals.

Keywords: melatonin-related genes; positive selection; functional diversification; reproductive rhythm

1. Introduction

Seasonal reproduction is a strategy for mammals to adapt to environmental changes. It allows mammals to reproduce in the most favorable period of the year for the survival and growth of their

offspring. The optical signal is a significant factor in controlling the seasonal reproduction of animals [1]. Melatonin (MLT), secreted by the pineal gland, is the key to synchronizing external optical signals and neuroendocrine signals [2].

Melatonin is synthesized and secreted by the pineal gland, retina, salivary gland, thymus, thyroid, kidney and liver [3], of which the pineal gland is the most important synthetic and secretory organ. There are four steps in the melatonin synthesis pathway in vivo. Firstly, tryptophan is converted to 5-hydroxytryptamine (5-HTP) by tryptophan hydroxylase (TPH) and decarboxylated by aromatic-L-amino acid decarboxylase (AAAD) to produce 5-hydroxytryptamine (5-HT), which is also known as serotonin. Then, serotonin is acetylated with aromatic alkyl-N-acetyltransferase (AANAT) and methylated with hydroxyl indole-o-methyl transferase (HIOMT) to form melatonin. Among these, TPH is the rate-limiting enzyme in serotonin synthesis [4]. AANAT and HIOMT are the rate-limiting enzymes in melatonin synthesis [5]. Notably, there are two *TPH* genes in the mammalian genome, including *TPH1* and *TPH2* [6]. The expression products of the two genes have similar structures and functional domains, and thus their functions are similar [7]. However, both genes were expressed in different organs: the *TPH1* is mainly expressed in the pineal gland, retina, kidney, lung, kidney, duodenum, liver and adrenal gland, while the *TPH2* is mainly expressed in the brain stem [7,8].

Melatonin is released into the cell membrane and binds to specific receptors to exert hormone function [9]. There are two melatonin receptors in mammals, namely melatonin receptor 1A (MTNR1A) and melatonin receptor 1B (MTNR1B). These two receptors are the members of G protein-coupled receptors, and MTNR1A is more common than MTNR1B in mammals [4].

The metabolic processes of melatonin are widely found in various tissues, such as the liver, plasma, lung, spleen, heart, adrenal gland and brain, among which the liver and brain are the main ones [10]. Melatonin is mainly converted to 6-hydroxymelatonin (6-HMT) and N-acetylserotonin (NAS) by cytochrome P450, which is further metabolized to form sulfate or glucoside compounds and then excreted through urine. It is documented that four cytochrome P450 proteins, including CYP1A1, CYP1A2, CYP1B1 and CYP2C19, are involved in the metabolism of melatonin, of which CYP1A1 and CYP1A2 are the main enzymes involved in melatonin metabolism in the liver, and the metabolite is 6-HMT. At the same time, a small amount of melatonin was metabolized by CYP1A2 and CYP2C19, and the metabolite was NAS [11]. In other tissues, especially in the intestine and cerebral cortex, melatonin metabolism is mainly catalyzed by CYP1B1 and its metabolite is 6-HMT [12]. In the brain, melatonin can be catalyzed by indoleamine-2,3-dioxygenase (IDO) to form N^1-acetyl-N^2-formyl-5-methoxyknuramine (AFMK), which is then converted to a more stable *N*-acetyl-5-methoxyknuramine (AMK) [13] by arylamide–formylase. There are two IDO proteins in mammals, namely IDO1 and IDO2, whose structures, functions, and expression regions are similar in vivo, but the expression process is regulated by different cytokines [14].

Variations in the sequences and expression of melatonin-related genes are strictly related to the human mood [15], behavior [16] and diseases [17]. Among these genes, polymorphisms of the *MTNR1A* gene encoding melatonin receptor are associated with the reproductive rhythm in sheep [18]. However, the association between genetic variations of other melatonin-related genes and mammalian reproductive rhythms remains unclear. Therefore, the purpose of this study was to detect: the allelic variations associated with the reproductive rhythm in the above melatonin-related genes and their evolutionary history.

2. Results

2.1. Evolution of Melatonin-Related Genes

The positive selection analysis of the 13 melatonin-related genes is shown in Table S1A, S1B and S1C. The ω values changed significantly among the codons of these genes as the likelihood ratio test (LRT) values of the M0–M3 model are significant in the Chi-square test ($p < 0.05$). Eight of these 13 melatonin-related genes have positively selected sites because their LRT values in the M7–M8 model

are significant ($p < 0.05$). Further Bayesian Empirical Bayesian (*BEB)* analysis showed that *TPH1*, *TPH2* and *IDO1* of the eight genes had significant positive selection sites ($p < 0.05$).

The positive selection analysis of *TPH1*, *TPH2*, and *IDO1* genes are shown in Table 1. For the *TPH1* gene, the LRT value of the M7–M8 model was 36.742 ($p < 0.01$), and 1.8% of *TPH1* codons were positively selected. However, the LRT value of the M1a–M2a model was 0 ($p > 0.05$). The *BEB* analysis showed that there were two significant positive selection sites ($p < 0.01$) at the protein level, which are the 97th and 442nd sites.

The results of the *TPH2* gene were similar to that of the *TPH1* gene. The LRT value of M7–M8 model was 47.952 ($p < 0.01$), and 1.7% of the codons were positively selected. However, the LRT value of the M1a–M2a model was 0 ($p > 0.05$). Besides, the *BEB* analysis exhibited two significant positive selection sites at protein level ($p < 0.01$), 47th and 66th sites.

For the *IDO1* gene, the LRT values of M1a–M2a and M7–M8 models were 23.524 and 52.290, respectively ($p < 0.01$). The M1a-M2a model showed that 0.4% of the codons were positively selected, while 25% of the codons were positively selected in the M7–M8 model. The *BEB* analysis in this model found ten significant positive selection sites at the protein level ($p < 0.05$): they are the 194th, 197th, 205th, 219th, 229th, 286th, 315th, 336th, 400th and 404th sites.

Table 1. Likelihood values and parameter estimates for tryptophan hydroxylase (TPH) genes in site models.

Gene	Model [a]	ln *L*	Estimates of Parameters [a]	LRT [b]	Positive Sites [c]
TPH1	M2a	−6828.237	$p_1 = 0.953$, $p_2 = 0.047$, $p_3 = 0.000$ $\omega_1 = 0.075$, $\omega_2 = 1.000$, $\omega_3 = 9.527$	0.000	97, 442
	M1a	−6828.237	$p_1 = 0.953$, $p_2 = 0.047$ $\omega_1 = 0.075$, $\omega_2 = 1.000$		
	M8	−6817.136	$p_0 = 0.982$, $p = 0.472$, $q = 4.095$ ($p_1 = 0.018$), $\omega = 1.391$	36.742 *	
	M7	−6835.507	$P = 0.348$, $q = 2.304$		
TPH2	M2a	−8015.897	$p_1 = 0.949$, $p_2 = 0.043$, $p_3 = 0.008$ $\omega_1 = 0.034$, $\omega_2 = 1.000$, $\omega_3 = 1.000$	0.000	47, 66
	M1a	−8015.897	$p_1 = 0.949$, $p_2 = 0.051$ $\omega_1 = 0.034$ $\omega_2 = 1.000$		
	M8	−7982.384	$p_0 = 0.983$, $p = 0.044$, $q = 0.337$ ($p_1 = 0.017$), $\omega = 1.228$	47.952 *	
	M7	−8006.360	$p = 0.109$, $q = 0.929$		
IDO1	M2a	−8015.897	$p_1 = 0.666$, $p_2 = 0.293$, $p_3 = 0.004$ $\omega_1 = 0.125$, $\omega_2 = 1.000$, $\omega_3 = 2.447$	23.524 *	194, 197, 205, 219, 229, 286, 315, 336, 400, 404
	M1a	−8015.897	$p_1 = 0.673$, $p_2 = 0.327$ $\omega_1 = 0.120$, $\omega_2 = 1.000$		
	M8	−7982.384	$p_0 = 0.750$, $p = 0.816$, $q = 3.690$ ($p_1 = 0.250$), $\omega = 1.204$	52.290 *	
	M7	−8006.360	$P = 0.078$, $q = 0.098$		

[a] Two pairs of nested models (M2a–M1a; M8–M7) were showed. In M1a and M2a, p_1, p_2 and p_3 were the proportions of sites with $0 < \omega < 1$, $\omega = 1$ and $\omega > 1$. In M7 and M8, p and q were the β distribution parameters, and p_0 and p_1 were the proportions of sites with $0 < \omega \leq 1$ and $\omega > 1$. Sequences were translated from nucleotides to amino acids. The F3×4 CodonFreq model tested the average nucleotide frequencies at three codon positions.; [b] Likelihood ratio test (LRT) values were followed the chi-square distribution with 4 (M3 and M0), 2 (M2a and M1a) and 2 (M8 and M7) degrees of freedom, respectively. * indicates $p < 0.05$.; [c] Amino acid residues were determined with sheep protein sequences. Only significant positive selection sites were listed.

The adaptive branching-site random effect likelihood (aBSREL) model was used to test the positively selected branches of *TPH1*, *TPH2* and *IDO1* genes, respectively, and the results are shown

in Figure 1. The *TPH1* gene was significantly positive selected on the horse and the ancestor of Green monkey and Sumatran orangutan ($p < 0.05$), with 3.6% and 1.2% of codons positively selected, respectively. The *TPH2* gene showed significant positive selection on golden hamster and crab-eating macaque ($p < 0.05$), and 0.76% and 0.44% of the codons of the expression products were positively selected in these two species, respectively. The *IDO1* gene had a significant positive selection on the ancient Boreoeutheria ($p < 0.01$), and 17% of the codons in this branch were positively selected.
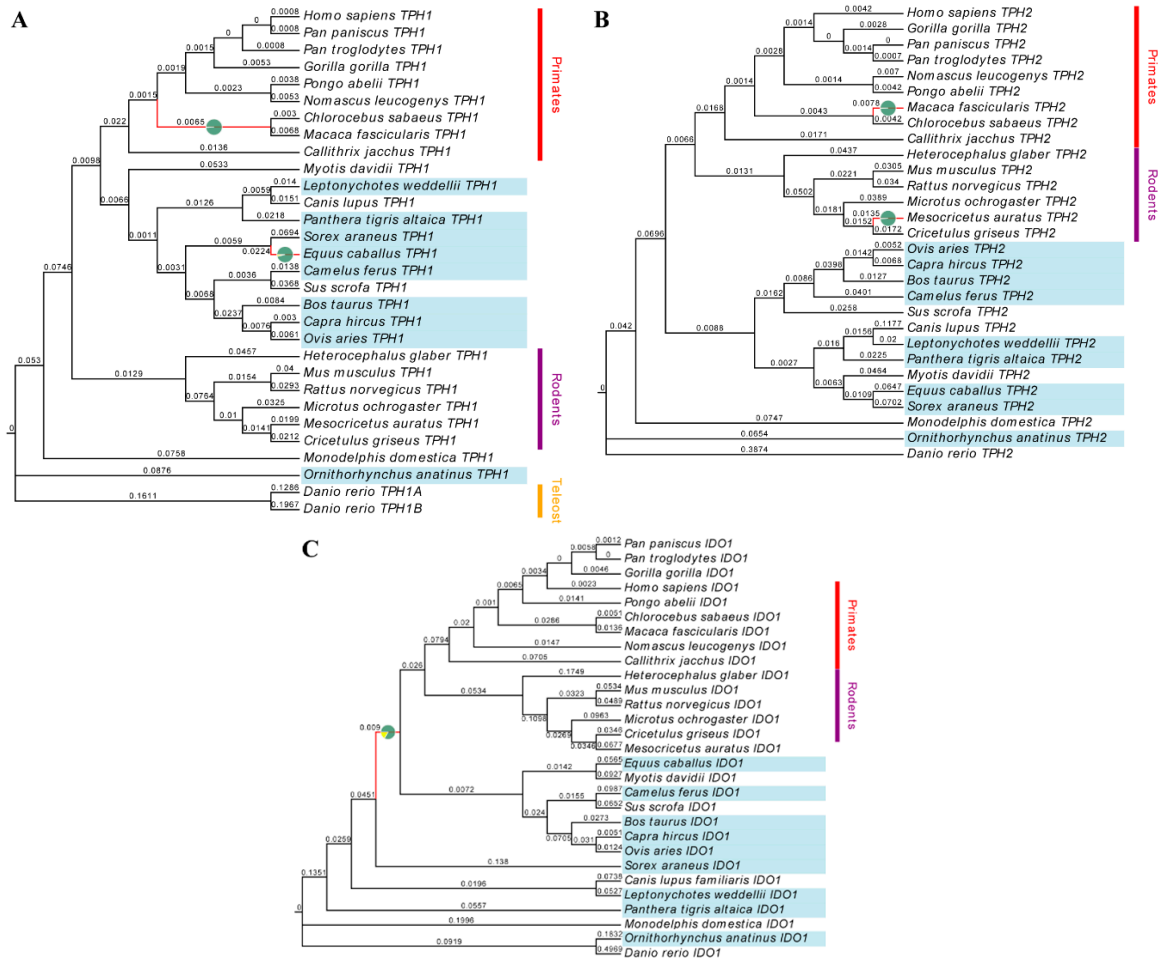

Figure 1. Phylogenetic trees of *TPH1*, *TPH2* and *IDO1* genes. The phylogenetic trees of *TPH1*, *TPH2* and *IDO1* genes are shown in (**A**), (**B**) and (**C**), respectively. The number on each branch represents the relative length of the branch. Colored bars and text describe the brief classification of species. The species with the blue background are seasonal breeders, while those without background are annual breeders or are not well defined. The red branches show that these lineages have a significant positive selection ($p < 0.05$). The pies of red branches show the proportion of positive selection sites (the yellow part) and non-positive selection sites (the green part) over the entire sites.

2.2. Functional Divergence of Melatonin-Related Proteins

The functional diversification analysis examined whether the functions of subtype proteins diverged, including TPH1 and TPH2, IDO1 and IDO2, and four cytochrome P450 isoforms, respectively. Then, it was determined which amino acid variants caused their functional divergence (Table 2). In 28 mammals, the LRT values of TPH1/TPH2, IDO1/IDO2 were 42.837 and 91.682 ($p < 0.01$), respectively, indicating that the structure and function of TPH1 and TPH2, IDO1 and IDO2 had visible divergence in these mammals. For the six combinations of the four cytochrome P450 protein isoforms, the LRT value of CYP1A1/CYP1A2 was 0.176 ($p > 0.05$), indicating that the structure and function

of CYP1A1 and CYP1A2 proteins did not diverge significantly in these mammals, while the protein divergence of the other five combinations was extremely significant ($p < 0.01$).

Table 2. Functional divergence estimated in melatonin-related genes.

Comparison (Cluster 1/Cluster 2)	θ [a]	SE [b]	LRT [c]	N (0.9) [d]
TPH1/TPH2	0.513	0.061	42.837 *	5
IDO1/IDO2	0.560	0.080	91.682 *	16
MTNR1A/MTNR1B	0.554	0.070	45.038 *	5
CYP1A1/CYP1A2	−0.019	0.093	0.176	0
CYP1A1/CYP1B1	0.802	0.102	60.455 *	16
CYP1A1/CYP2C19	0.660	0.116	34.390 *	4
CYP1A2/CYP1B1	0.766	0.100	62.899 *	13
CYP1A2/CYP2C19	0.892	0.118	57.428 *	22
CYP1B1/CYP2C19	0.839	0.100	77.931 *	51

[a] θ: the coefficient of functional divergence.; [b] SE: standard error.; [c] LRT is a likelihood ratio test. The value approximately follows a chi-square distribution with one degree of freedom. * indicated $p < 0.05$.; [d] $N_{(0.9)}$ means the numbers of divergent residues when the cut-off value $Q_{(k)}$ is over 0.9. The specific amino acids related to protein functional divergence are listed in Table S2.

The results in Table S2 indicated that the 442nd site of TPH1, which was positively selected, might be related to the functional divergence of TPH1 and TPH2. For IDO1, four positively selected sites—the 194th, 286th, 315th and 404th—might be related to the functional divergence of IDO1 and IDO2.

2.3. The T/C Polymorphism at the 1325th Site in TPH1 Gene Change Enzyme Activity

There is a total of five amino acid sites on the TPH1 and IDO1 proteins that are both positively selected and related to functional divergence. The 442nd site of TPH1 is located in the tetramer-binding domain (Figure 2B), and three of the four positive selected sites of IDO1 are in the pyridoxal phosphate-dependent transferase domain, including the 194th, 286th and 315th sites. There was no polymorphism at the 442nd site of TPH1 within the species included in the Ensemble database. For IDO1, two synonymous variants at the 315th and 404th sites can be detected in human and house mouse, respectively, and the 194th and 286th sites showed no polymorphism within the species included in the database. Therefore, we aligned the amino acid sequences of 28 mammals and found that it was leucine (L) or proline (P) at 442nd site of TPH1, which was skewed distributed in seasonal or non-seasonal reproducing mammals (Figure 2A). The amino acid at the 442nd site was proline in 89.5% of non-seasonal reproducing mammals, while it was leucine in 88.9% of seasonal reproducing mammals.

The nucleotide sequence alignment of the *TPH1* gene in 28 mammals showed that the variation of the 442nd amino acid originated from the T/C mutation of the 1325th nucleotide site in the coding sequence. The allelic distribution of this site was consistent with the distribution of the amino acid in seasonal or non-seasonal reproducing mammals. The allele C was in 89.5% of non-seasonal reproducing mammals, while allele T was in 88.9% of seasonal reproducing mammals. The allele C appeared in bats 83.0 million years ago (mya) and spread in rodents 52.4 mya. The allele T appeared 212.6 mya and spread in primates 25.1 mya.

The enzyme assay of TPH1-1325T and TPH1-1325C proteins revealed that the T/C polymorphism of the 1325th site could significantly change the activity of TPH1 protein (Figure 2C). The activity of TPH1-1325T protein was 0.020 U/mg and that of TPH1-1325C was 0.004 U/mg. The activity of TPH1-1325T was five times higher than that of TPH1-1325C, and their difference was highly significant ($p < 0.01$).
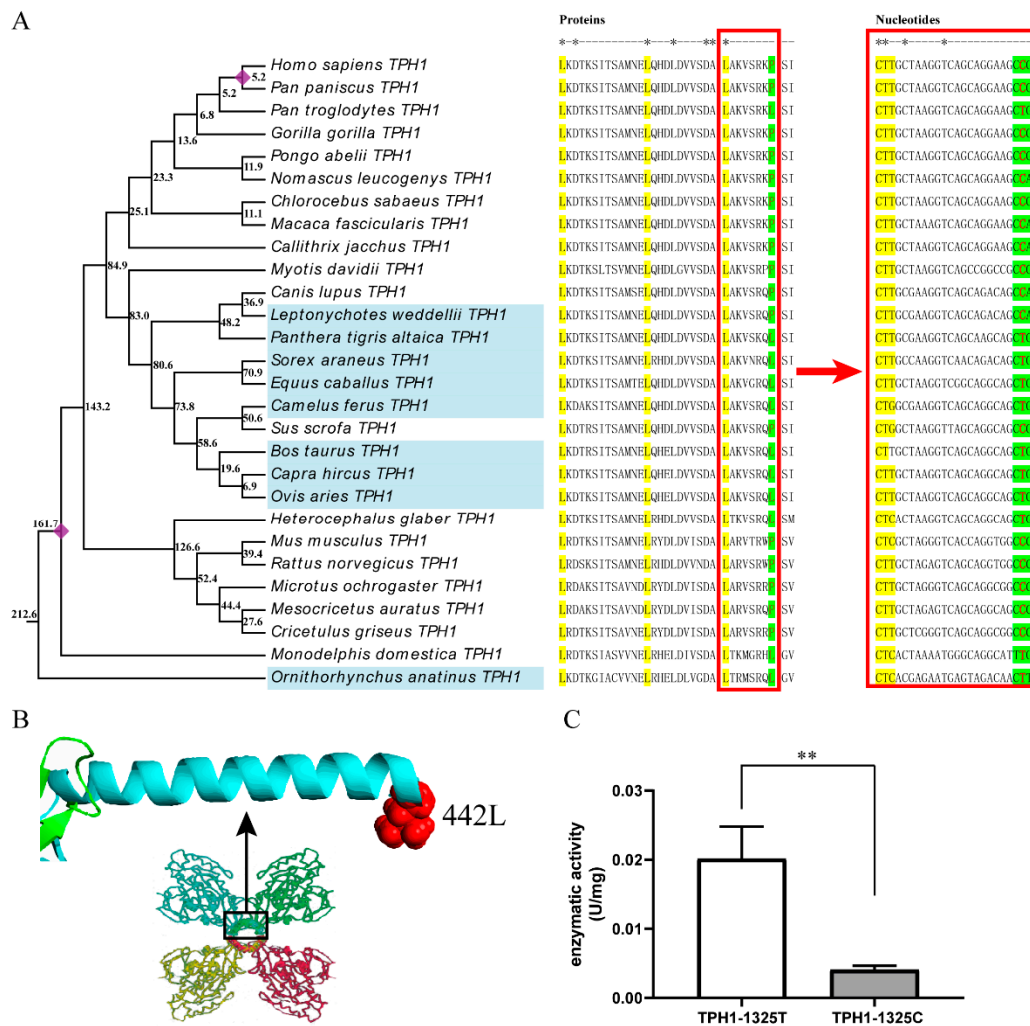
Figure 2. T1325C polymorphism of the *TPH1* gene and its effect on enzyme activity. (**A**) The variation schematic map of *TPH1* genes in mammalian evolution. The phylogenetic tree of *TPH1* genes is shown on the left. The species with the blue background are seasonal breeders, while those without background are annual breeders or are not well defined. The number of nodes indicated the divergence time of the *TPH1* gene. The nodes labeled with a purple diamond are two calibration constraints. The amino acid and nucleotide sequences of the tetramerization domains of these species are shown on the right. There are four leucine in the tetramerization domain. The first three leucine have no difference among species, and the amino acids and corresponding nucleotides are labeled with the yellow background. The last leucine is at the 442nd site, and the corresponding nucleotides are colored with green. The 1325th site in nucleotide sequences was labeled with red. (**B**) The partial three-dimensional structure, positive selection sites and tetramer of TPH1 proteins. The red residue represented the 442nd site. The tetramerization domain is an alpha-helix at the C-terminal of protein and marked in cyan. The picture below the structure of the alpha-helix indicates the tetramer of TPH1. Each monomer is in a different color. The tetramer is anchored by the C-terminal tetramerization domain and additional inter-subunit hydrophobic interactions between the regulatory domains. (**C**) Enzymatic activity of TPH1-1325T and TPH1-1325C proteins. ** indicates significant differences ($p < 0.01$).

3. Discussion

3.1. TPH1, TPH2 and IDO1 Genes are Positively Selected in Mammals

Calculating the ratio of the number of non-synonymous substitution sites to that of synonymous substitution sites is a standard method for the positive selection test. However, this method has

been criticized for producing excessive false-positives and for being sensitive to various disturbing factors [19].

Eight of the 13 melatonin-related genes studied here had a significant difference in *LRT* values ($p < 0.05$) in the M7–M8 model. As for the eight genes, only the *IDO1* gene had different *LRT* values in both M1a–M2a and M7–M8 models ($p < 0.05$). For the other seven genes, including *TPH1*, *TPH2*, *DDC*, *MTNR1B*, *IDO2*, *CYP1A1* and *CYP1B1* genes, the *LRT* values of the M1a–M2a model were not significantly different ($p > 0.05$) while those of the M7–M8 model showed significant differences ($p < 0.05$). Two of these seven genes, *TPH1* and *TPH2*, had significant positive selection sites ($p < 0.05$) in *BEB* analysis. The study of Twyman showed similar results in detecting the selection pressure of the *CYP2J19* gene [20]. Therefore, we accepted the results of M7–M8 that positive selection sites existed in these seven genes.

Berlin and Smith pointed out that the M7–M8 model may overestimate non-positive selection sites as positive selection sites, leading to false-positive results [21]. To avoid misjudging false-positive sites, we used the aBSREL branch length-site model to detect the selection of *TPH1* and *TPH2* gene. The M7–M8 model in the codeml program assumes the ω values in different branches are fixed. However, the aBSREL program assumes that the ω values are variable between codons and in different branches. Therefore, the model can detect not only positive selection sites but also positive selection branches during evolution [22], thus complementing the results of codeml. The results of the aBSREL showed that the *TPH1* gene was significantly positively selected in two branches of the phylogenetic tree (i.e., the horse branch and the ancient branch of the green monkey-Sumatran orangutan). The *TPH2* gene was positively selected in the branches of golden hamster and crab-eating macaque, respectively. The results were consistent with those of the M7–M8 model in the codeml program, indicating that our results of *TPH1* and *TPH2* genes were credible. Although, the gene has multiple transcript variants in a specific species in GenBank, we only chose one sequence which is most similar to the human sequence for research. Obviously, this method will actually lose some polymorphic information, which will lead to a reduction in the accuracy of selection pressure on different branches. Therefore, the more accurate selection pressure of each branch needs to be further studied in specific species.

3.2. Effect of T1325C Polymorphism of the TPH1 Gene on Seasonal Reproduction in Mammals

In mammals, the expression tissues of the *TPH1* gene and the synthesis tissues of melatonin are similar. Even if the expression product of the *TPH1* gene was not regarded as a rate-limiting enzyme in the melatonin synthesis pathway [23], recent studies still believed that it could regulate the synthesis of serotonin and melatonin in the pineal gland and peripheral tissues [6,24].

Positive selection and functional divergence analysis showed that 442L was not only a significant positive selection site but also related to functional divergence. This site was in the tetramer region of TPH1 protein, which is an alpha-helix structure at the C-terminal. In seasonal reproducing mammals, the amino acid at 442nd site was mainly leucine (88.9%), while the amino acid was mainly proline (89.5%) in non-seasonal reproducing mammals. However, the amino acid here was proline in Weddell's seal, which is a seasonal reproducing mammal—probably because Weddell's seal is an aquatic mammal.

Comparing the nucleotide sequences and protein sequences, we found that the variation of TPH1 protein L442P originated from the T1325C mutation of the *TPH1* gene. When the T1325C site of the *TPH1* gene mutated from T to C, the 442nd amino acid changed from L to P, thus reducing the number of leucine in the tetramer region. Our results in the activity assay showed a higher activity of TPH1-1325T (four leucine in the tetramer region) and lower activity of TPH1-1325C (three leucine in the tetramer region). The four leucine in this region formed a leucine zipper. The leucine zipper maintained TPH1 tetramers that are stable and active. The decrease of the leucine amount results in dimmers or monomers of the TPH1 that are unstable or inactive [25]. The lower active of TPH1 with 442P in our study supports Marshall's study.

TPH1 protein catalyzes tryptophan to produce serotonin, which provides a substrate for the synthesis of melatonin. Whether in the retina or pineal, the expression of the *TPH1* gene showed a trend

of low during the day and high during the night [26], which was consistent with the rule of melatonin synthesis [27]. However, serotonin in vivo is the highest at midnight and noon and the lowest at dawn and dusk [28]. We suppose that serotonin may be synthesized before melatonin as a substrate at night. Therefore, the increased activity of TPH1 can advance the nighttime peak of melatonin by improving the synthesis efficiency of serotonin, leading to the increased sensitivity of animals to diurnal changes, which provides a basis for seasonal reproduction. On the contrary, the decrease of TPH1 activity may lead to a decline in melatonin synthesis efficiency. Thus, the sensitivity of animals to optical signals would be reduced and eventually lead to the weakening or disappearance of seasonal reproduction of mammals.

However, there are exceptions to the correlation of the 1325th site of the TPH1 gene coding sequence to the seasonal reproduction of mammals. For example, in the *TPH1* gene of gray short-tailed opossum and chimpanzee, the allele at the T1325C site is T, but these two species were not seasonal breeders. It is inferred that this may be due to the complex process of seasonal reproduction controlled by multiple genes. The effect of T1325C mutation on melatonin synthesis in these species may be offset by mutations in other genes that are also involved in this pathway, especially AANAT and HIOMT speed-limiting enzymes.

3.3. Seasonal Reproduction Improved the Adaptability of Mammals to Cretaceous–Tertiary Extinction Events

The estimation of the divergence time of the *TPH1* gene in 28 mammals showed that the allele T at the 1325th site has existed for over 210 million years, and allele C probably first appeared about 83.0 million years ago (mya). We mapped the TPH1 evolutionary history from 210 mya to the present in Figure 3.
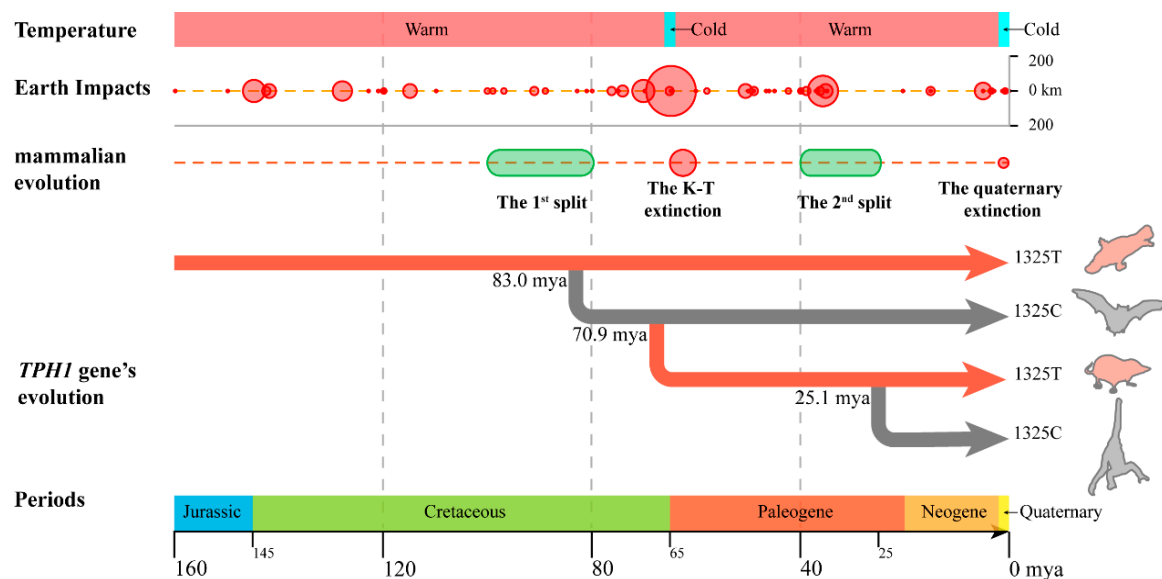

Figure 3. The scheme of *TPH1* gene evolution in mammals.

The platypus is the oldest and the most ancient mammal that still exists today; it appeared at the end and the warmest time of the Cretaceous (145.0–165.0 mya) [29]. After millions of years, it has neither gone extinct nor evolved significantly. The platypus is a seasonal reproducing mammal. The base at 1325th site of the *TPH1* coding sequence is T in the platypus, suggesting the base at this site in mammals may be T at first and the T base may control seasonal reproduction in platypus. The base at 1325th site changed from T to C and spread in ancestors of rodents (mainly non-seasonal reproduction animals) about 83.0 mya (Figure 2A), which coincided with the time of the first major split events in mammals (100.0–85.0 mya) [30]. During this period, the first split event of mammalian species occurred, and the ancestors of many existing species appeared, such as hamsters, rabbits, goats and

sheep (all about 96.0 mya). The disappearance of seasonal reproductive characteristics is conducive to more offspring and a rapid expansion of population size.

Later, allele T spread again about 70.9 mya, which was close to the time of Cretaceous–Tertiary (K-T) extinction (about 65.5 million years ago) [31]. The K-T extinction event was a mass extinction event in the history of the Earth which led to the extinction of most of the animals and plants at that time and is famous due to the rise of mammals [31]. The most likely cause of this extinction is multiple impact events [32] or long-term volcanic eruptions [33], which threw a large amount of dust into the atmosphere. The dust blocked the sunlight, causing the temperature to drop sharply [29], and resulting in large-scale plant deaths due to reduced photosynthesis, thus resulting in a shortage of animal food sources, leading to the extinction of a variety of animals. We hypothesize that seasonal reproduction improves the adaptability of some mammals to this extreme environmental pressure. For non-seasonal reproducing mammals, such as rodent ancestors, their small size and cave-dwelling characteristics helped them to survive and reproduce with less food, and it was easy to find suitable habitats for population continuation. In the genome of these mammals, allele C is retained. For seasonal reproducing mammals with the allele T at the 1325th site, such as the ancestors of existing cattle, horse, goat and sheep, they can change their reproductive rhythm by sensing light changes and produce offspring when food is relatively adequate, thus ensuring the survival of offspring and population continuity. Seasonal reproduction improved their adaptability to the surging environmental pressure, making individuals with allele T dominant in natural selection.

After the K-T extinction, the Earth entered the Paleocene (65.0–55.0 mya). During this period, the light returned to normal, global temperature increased by 5–8 °C, and the Paleocene–Eocene Thermal Maximum (PETM or IETM) occurred. As a result, environmental pressure relaxed [34]. Under this background, a second mammalian split event occurred (45.0–25.0 mya). Non-seasonal reproduction enabled mammals to have more offspring to gain an advantage in competing for the ecological space once occupied by extinct species [35]. In this context, the individuals with allele C predominate in natural selection. The second spread of allele C in the mammalian genome occurs.

We inferred that the evolution of the *TPH1* gene in mammals might be influenced by two mammalian split events and the K-T extinction event. The T1325C site under positive selection could be evolved under temperature-associated selection. The environmental changes play a dominant role in the spread of the *TPH1* T1325C site and the change of seasonal reproduction in mammals. When light and temperature decreased, seasonal reproduction could improve the adaptability of populations to extreme environmental pressures, which caused allele C to spread. Once light increased and the climate warmed up, non-seasonal reproducing populations could gain more offspring under relaxed environmental pressures. Thus, individuals with allele T predominated in natural selection.

We mapped the mammalian evolutionary history from 160 million years ago (mya) to the present. Five periods were mainly experienced: Jurassic (200.0–145.0 mya), Cretaceous (145.0–65.0 mya), Paleogene (65.0–23.3 mya), Neogene (23.3–2.6 mya) and Quaternary (2.6 mya–now). During these periods, the Earth was mainly in the Triassic–Tertiary interglacial period (200.0–2.0 mya), with a warm climate. However, about 65 million years ago, there was a cold period due to multiple impacts (with a maximum radius of about 150 km) and long-term volcanic eruptions. After this intense period, the global temperature rose rapidly. During the Quaternary, the Earth entered the Quaternary Ice Age, the second cold period occurred, and the temperature decreased sharply.

During these 160 million years, mammals experienced two splits and two extinctions. The first split occurred about 100–85 mya, which was the warmest time in the whole Cretaceous. Subsequently, the K-T extinction happened about 65.5 mya. After the K-T extinction, mammals entered the second major split (about 40–25 million years ago). The second extinction occurred during the Quaternary Ice Age, during which many mammals were extinct.

The evolutionary history of the *TPH1* gene is highly consistent with the change of reproductive rhythm in mammals. Based on the phylogenetic tree of 29 species, with the platypus (red, seasonal reproduction) as a model, the three changes of reproductive rhythm and the 1325th site of *TPH1*

gene were found at the same time in bat (gray, non-seasonal reproduction), European shrew (red, seasonal reproduction) and marmoset (gray, non-seasonal reproduction), respectively. Therefore, these four species (arrange from top to bottom) were chosen to describe the evolution of the *TPH1* gene in mammals.

4. Materials and Methods

4.1. Data Collection

A total of 259 coding sequences for 13 melatonin-related genes from 29 species were collected from the NCBI database. Of the 29 species, 28 of them are mammals, and one is zebrafish as an outgroup. The Latin and English names of the species and the genes with accession numbers are shown in Table S3. If there are multiple transcript variants of a gene in a specific species, only one sequence which is most similar to the human sequence will be selected for further research.

4.2. Phylogenetic Trees and Divergence Time

To construct the evolutionary history of melatonin-related genes, the coding sequences of 13 genes were aligned with MEGA-X [36] and their phylogenetic trees were constructed by maximum likelihood method (ML) with 1000 bootstrap values [37]. Pairwise deletion was used to delete the sites with less than 95% coverage among species. The relative substitution rates of nucleotides were calculated by Kimura's two-parameter model [38]. The EvolView online program (http://www.evolgenius.info/evolview/) was used to modify the phylogenetic trees [39].

The maximum likelihood method and Kimura two-parameter model were used to infer the ancestral sequences for these TPH proteins in 29 species [40]. The divergence times of 13 genes were calculated based on the divergence time [41–43] of the human and gray short-tailed opossum (155.0–166.0 million years ago), and human and gorilla (5.1–11.8 million years ago) by the clocks process based on the RelTime method in MEGA X [44].

4.3. Positive Selection Assessment

Three pairs of nested models (M0–M3, M1a–M2a, and M7–M8) were employed in the site model of the codeml program in the Phylogenetic analysis by maximum likelihood (PAML) package to predict the potential positively selected sites in melatonin-related genes. The selection pressure of genes was indicated by the ratio (ω) of non-synonymous substitution and synonymous substitution number (d_N/d_S). When $\omega = 1$, the gene is neutrally selected; when $\omega < 1$, the gene is purified; when $\omega > 1$, the gene is positively selected [45].

In three pairs of nested models (M0–M3, M1a–M2a, M7–M8), the M0 model assumes that all loci have the same ω values, while the M3 model assumes that the ω values of each codon are independent and have three possibilities: more than 1, equal to 1, and less than 1. The M0–M3 model is not used for the positive selection test, but for detecting whether the ω values between loci are consistent or not. Both M1a–M2a and M7–M8 models consisted of null hypothesis models (M1a and M7), which assume that all codons are under near-neutral selection, and alternative hypothesis models (M2a and M8) assume that positive selection exists. The difference between the two pairs of models was that the M7–M8 models used β distribution to fit the value of ω [46]. Compared with the M7–M8 model, the M1a–M2a model is more stringent [46], while the M7–M8 model is more sensitive and can detect weaker positive selection [21].

For each pair of models, the likelihood ratio test (LRT) was used to evaluate the fitting degree of nested models with the likelihood logarithm (ln*L*) of the paired model. Significance was evaluated by the chi-square test with 4, 2 and 2 degrees of freedom for M0–M3, M1a–M2a and M7–M8 models, respectively. Naive Bayesian (NEB) and Bayesian Empirical Bayesian (*BEB*) methods in M2a and M8 were used to calculate the posterior probability of each site selected by codons and to judge the site

selection from the significant test results. According to the suggestions given in the codeml instructions, we ignored *NEB* and used *BEB* results directly [47].

An adaptive branching-site random effect likelihood model (aBSREL) was used to calculate the selection pressure (ω) on the codons of melatonin-related genes in various branches of phylogenetic trees. In this model, branches are divided into foreground branches (branches to be detected) and background branches in advance. A likelihood ratio test was used to evaluate the fitting degree of foreground branches to near-neutral selection (null hypothesis) and positive selection (alternative hypothesis) and then to determine whether positive selection has taken place [22]. The calculation program of the model is provided by the DataMonkey server (http://www.datamonkey.org/).

4.4. Functional Diversification Analysis

The structural divergence of melatonin-related proteins between different branches of the phylogenetic tree was examined, and their amino acid sites related to functional divergence were predicted with the Gu99 and Type-I functional divergence in the Diverge v.3.0 program. After comparing the amino acid sequences of all proteins, the structural divergence of proteins between different branches was calculated by the likelihood ratio test (LRT), and their functional divergence was inferred. Type-I functional divergence (site-specific rate shift) refers to the evolutionary process resulting in site-specific rate shifts after gene duplication. It identifies amino acid residues highly conserved in one gene copy and highly variable in the other. The probability of a residue being under Type-I divergence is denoted as θ. $Q_{(k)}$ is the site (k)-specific score corresponding to the posterior probability that site k is related to type-I functional divergence. Based on Thompson and his colleagues' report, 0.9 was chosen as the critical value of $Q_{(k)}$, and the site of $Q_{(k)} > 0.9$ was regarded as an important amino acid leading to functional divergence [45].

The polymorphism of sites which are the positive selected and related to functional divergence was detected with the variant tables (Table S4) of the Ensemble database.

4.5. Structural Analysis and Homology Modeling

The InterPro server (http://www.ebi.ac.uk/interpro/scan.html) was used to predict the conservative domain of melatonin-related proteins [48]. Three-dimensional structures of melatonin-related proteins were predicted with the I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) online server [49]. The PyMOL v1.6.x (San Carlos, California, USA) was used to visualize the 3D structure as well as the positioning of conservative domains and positively selected sites.

4.6. Cell Expression and Enzyme Activity Determinations

The effect of sites that were positively selected and the functional divergence associated with enzyme activity were examined. The wild-type *TPH1* gene was synthesized based on the coding sequence of sheep *TPH1* gene (NC_019459.1) in NCBI, and its expression product was named as TPH1-1325T. Then, the coding sequence of the T1325C mutated *TPH1* gene was obtained by replacing the 1325th nucleotide on wild-type *TPH1* gene with cytosine from thymine, and the expressed product was named as TPH1-1325C. These two sequences were artificially synthesized and subcloned into pVAX1-asd vector by adding *EcoRI* and *NotI* digestion sites at 5′ and 3′ ends, respectively. Subsequently, DNA sequencing was used to detect the accuracy of the two inserts.

To express the protein from the cloned sequences, HEK293T cells were cultured at 37 °C with 5% CO_2 and 95% humidity for 24 h in Dulbecco's modified Eagle medium (DMEM, high glucose 4.5 g/L, Gibco, Invitrogen) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Gibco), and 1% L-glutamine, 0.5% penicillin-streptomycin. Then, cells were transfected with a lipofectamine 2000 (Invitrogen) transfection kit. After 48 h of incubation, the total proteins were extracted with the Tryptophan Hydroxylase assay kit (Haling biological technology CO., LTD., Shanghai, China) according to the instructions. The concentration of total protein was determined by the BCA protein assay kit, and the relative concentration of TPH protein was determined by the Western blot. Finally,

according to the steps provided by the tryptophan hydroxylase assay kit instructions, the two TPH1 proteins of equal amounts were added to the reaction solution, and their enzyme activities were measured at 750 nm.

5. Conclusions

It is concluded that the T/C polymorphism at 1325th site in the *TPH1* gene coding sequence changed the TPH1 enzyme activity and that this polymorphism is consistent with the reproductive rhythm of mammals.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/20/23/6065/s1.

Author Contributions: G.L. and X.J. conceived and designed the experiments. C.L. performed the experiments. S.G. and T.W. were involved in the methodology and revised the manuscript. X.J. and G.L. performed the supervision of the article and contributed reagents/materials/analysis tools. All authors read and approved the final manuscript.

Conflicts of Interest: The authors declare that no conflict of interest could be perceived as distorting the impartiality of the research report.

References

1. Lewis, P.D. A review of lighting for broiler breeders. *Br. Poult. Sci.* **2006**, *47*, 393–404. [CrossRef] [PubMed]
2. Rosenthal, N.E.; Sack, D.A.; Jacobsen, F.M.; James, S.P.; Parry, B.L.; Arendt, J.; Tamarkin, L.; Wehr, T.A. Melatonin in seasonal affective disorder and phototherapy. *J. Neural Transm. Suppl.* **1986**, *21*, 257–267. [PubMed]
3. Kvetnoy, I. Extrapineal melatonin in pathology: New perspectives for diagnosis, prognosis and treatment of illness. *Neuro. Endocrinol. Lett.* **2002**, *23*, 92–96. [PubMed]
4. ViviD, D.; Bentley, G. Seasonal reproduction in vertebrates: Melatonin synthesis, binding, and functionality using Tinbergen's four questions. *Molecules* **2018**, *23*, 652. [CrossRef] [PubMed]
5. Marchesani, R.B. Metaphorical thinking on art, education and therapy: Late night conversations with bernard brenner. *Psychother. Patient* **2003**, *12*, 17–44. [CrossRef]
6. McKinney, J.; Knappskog, P.M.; Haavik, J. Different properties of the central and peripheral forms of human tryptophan hydroxylase. *J. Neurochem.* **2005**, *92*, 311–320. [CrossRef]
7. Zill, P.; Büttner, A.; Eisenmenger, W.; Bondy, B.; Ackenheil, M. Regional mRNA expression of a second tryptophan hydroxylase isoform in postmortem tissue samples of two human brains. *Eur. Neuropsychopharmacol.* **2004**, *14*, 282–284. [CrossRef]
8. Patel, P.D.; Pontrello, C.; Burke, S. Robust and tissue-specific expression of TPH2 versus TPH1 in rat raphe and pineal gland. *Biol. Psychiatry.* **2004**, *55*, 428–433. [CrossRef]
9. Dores, R.M. Observations on the evolution of the melanocortin receptor gene family: Distinctive features of the melanocortin-2 receptor. *Front. Neurosci.* **2013**, *7*, 28. [CrossRef]
10. Deprés-Brummer, P.; Metzger, G.; Morin, D.; Urien, S.; Touitou, Y.; Tillement, J.P.; Claustrat, B.; Lévi, F. Pharmacokinetically guided melatonin scheduling in rats with circadian system suppression. *Eur. J. Pharmacol.* **1996**, *312*, 171–178. [CrossRef]
11. Li, C.; Li, G.; Tan, D.X.; Li, F.; Ma, X. A novel enzyme-dependent melatonin metabolite in humans. *J. Pineal Res.* **2012**, *54*, 100–106. [CrossRef] [PubMed]
12. Ma, X.; Idle, J.R.; Krausz, K.W.; Gonzalez, F.J. Metabolism of melatonin by human cytochromes p450. *Drug Metab. Dispos.* **2005**, *33*, 489–494. [CrossRef] [PubMed]
13. Dun-Xian, T.; Lucien, M.; Eduardo, E.Z.; Zhou, Z.; Russel, R. Melatonin as a potent and inducible endogenous antioxidant: Synthesis and metabolism. *Molecules* **2015**, *20*, 18886–18906.
14. Maiwald, S.; Wehner, R.; Schmitz, M.; Bornhäuser, M.; Loeb, S.; Wassmuth, R. Ido1 and ido2 gene expression analysis by quantitative polymerase chain reaction. *Tissue Antigens* **2011**, *77*, 136–142. [CrossRef] [PubMed]

15. Brown, S.M.; Peet, E.; Manuck, S.B.; Williamson, D.E.; Dahl, R.E.; Ferrell, R.E.; Hariri, A.R. A regulatory variant of the human tryptophan hydroxylase-2 gene biases amygdala reactivity. *Mol. Psychiatry* **2005**, *10*, 884–888. [CrossRef] [PubMed]

16. Rotondo, A.; Schuebel, K.; Bergen, A.; Aragon, R.; Virkkunen, M.; Linnoila, M.; Goldman, D.; Nielsen, D. Identification of four variants in the tryptophan hydroxylase promoter and association to behavior. *Mol. Psychiatry* **1999**, *4*, 360–368. [CrossRef] [PubMed]

17. Tao, S.; Chattun, M.R.; Yan, R.; Geng, J.; Zhu, R.; Shao, J.; Lu, Q.; Yao, Z. TPH-2 gene polymorphism in major depressive disorder patients with early-wakening symptom. *Front. Neurosci.* **2018**, *12*, 827. [CrossRef]

18. Mura, M.C.; Luridiana, S.; Pulinas, L.; Di Stefano, M.V.; Carcangiu, V. Reproductive response to male joining with ewes with different allelic variants of the mtnr1a gene. *Anim. Reprod. Sci.* **2019**, *200*, 67–74. [CrossRef]

19. Suzuki, Y. Statistical methods for detecting natural selection from genomic data. *Genes Genet. Syst.* **2010**, *85*, 359–376. [CrossRef]

20. Twyman, H.; Andersson, S.; Mundy, N.I. Evolution of CYP2J19, a gene involved in colour vision and red coloration in birds: Positive selection in the face of conservation and pleiotropy. *BMC Evol. Biol.* **2018**, *1*, 22. [CrossRef]

21. Berlin, S.; Smith, N.G. Testing for adaptive evolution of the female reproductive protein ZPC in mammals, birds and fishes reveals problems with the M7-M8 likelihood ratio test. *BMC Evol. Biol.* **2005**, *5*, 65. [CrossRef] [PubMed]

22. Smith, M.D.; Wertheim, J.O.; Weaver, S.; Murrell, B.; Scheffler, K.; Kosakovsky Pond, S.L. Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* **2015**, *32*, 1342–1353. [CrossRef] [PubMed]

23. Boularand, S.; Darmon, M.C.; Ravassard, P.; Mallet, J. Characterization of the human tryptophan hydroxylase gene promoter transcriptional regulation by camp requires a new motif distinct from the camp-responsive element. *J. Biol. Chem.* **1995**, *27*, 3757–3764. [CrossRef] [PubMed]

24. Walther, D.J.; Peter, J.U.; Bashammakh, S.; Hörtnagl, H.; Voits, M.; Fink, H.; Bader, M. Synthesis of serotonin by a second tryptophan hydroxylase isoform. *Science* **2003**, *299*, 76. [CrossRef] [PubMed]

25. Marshall, A.M.; Hernandez, L.L.; Horseman, N.D. Serotonin and serotonin transport in the regulation of lactation. *J. Mammary Gland Biol. Neoplasia* **2014**, *19*, 139–146. [CrossRef]

26. Bernard, M.; Guerlotté, J.; Grève, P.; Gréchez-Cassiau, A.; Iuvone, M.P.; Zatz, M.; Chong, N.W.; Klein, D.C.; Voisin, P. Melatonin synthesis pathway: Circadian regulation of the genes encoding the key enzymes in the chicken pineal gland and retina. *Reprod. Nutr. Dev.* **1999**, *39*, 325–334. [CrossRef]

27. Weems, P.W.; Goodman, R.L.; Lehman, M.N. Neural mechanisms controlling seasonal reproduction: Principles derived from the sheep model and its comparison with hamsters. *Front. Neuroendocrinol.* **2015**, *37*, 43–51. [CrossRef]

28. Qian, S.W.; Yang, X.Z.; Li, H.Y.; Li, H.; Zhang, S.S.; Lu, Y.; Wang, J.P.; Zhang, Z.X.; Xu, J. Analysis on time queue data of circadian rhythms of 5-hydroxytryptamine. *CJTCMP* **2016**, *31*, 93–95.

29. Keller, G.; Stinnesbeck, W.; Adatte, T.; Stüben, D. Multiple impacts across the Cretaceous–Tertiary boundary. *Earth-Sci. Rev.* **2003**, *62*, 327–363. [CrossRef]

30. Bininda-Emonds, O.R.; Cardillo, M.; Jones, K.E.; MacPhee, R.D.; Beck, R.M.; Grenyer, R.; Price, S.A.; Vos, R.A.; Gittleman, J.L.; Purvis, A. The delayed rise of present-day mammals. *Nature* **2007**, *446*, 507–512. [CrossRef]

31. MacLeod, N.; Rawson, P.F.; Forey, P.L.; Banner, F.T.; Boudagher-Fadel, M.K.; Bown, P.R.; Burnett, J.A.; Chambers, P.; Culver, S.; Evans, S.E.; et al. The Cretaceous-tertiary biotic transition. *J. Geol. Soc.* **1997**, *154*, 265–292. [CrossRef]

32. Sleep, N.H.; Lowe, D.R. Physics of crustal fracturing and chert dike formation triggered by asteroid impact, ~3.26 ga, barberton greenstone belt, south africa. *Geochem. Geophys. Geosyst.* **2014**, *15*, 1054–1070. [CrossRef]

33. Keller, G. The Cretaceous–Tertiary mass extinction, Chicxulub impact, and Deccan volcanism. In *Earth and Life*; Springer Netherlands: Dordrecht, The Netherlands, 2012; pp. 759–793.

34. Westerhold, T.; Röhl, U.; Raffi, I.; Fornaciari, E.; Monechi, S.; Reale, V.; Bowles, J.; Evans, H.F. Astronomical calibration of the Paleocene time. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **2008**, *257*, 377–403. [CrossRef]

35. Sahney, S.; Benton, M.J.; Ferry, P.A. Links between global taxonomic diversity, ecological diversity and the expansion of vertebrates on land. *Biol. Lett.* **2010**, *6*, 544–547. [CrossRef] [PubMed]

36. Thompson, J.D.; Gibson, T.J.; Higgins, D.G. Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinform.* **2002**. [CrossRef] [PubMed]

37. Liang, G.M.; Jiang, X.P. Positive selection drives lactoferrin evolution in mammals. *Genetica* **2010**, *138*, 757–762. [CrossRef]

38. Xu, S.; Chen, Y.; Cheng, Y.; Yang, D.; Zhou, X.; Xu, J.; Zhou, K.; Yang, G. Positive selection at the ASPM gene coincides with brain size enlargements in cetaceans. *Proc. Biol. Sci.* **2012**, *279*, 4433–4440. [CrossRef]

39. Zhang, H.; Gao, S.; Lercher, M.J.; Hu, S.; Chen, W.H. EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res.* **2012**, *40*, W569–W572. [CrossRef]

40. Mika, K.M.; Lynch, V.J. An ancient fecundability-associated polymorphism switches a repressor into an enhancer of endometrial tap2 expression. *Am. J. Hum. Genet.* **2016**, *99*, 1059–1071. [CrossRef]

41. Wang, Y.; Gu, X. Evolutionary patterns of gene families generated in the early stage of vertebrates. *J. Mol. Evol.* **2000**, *51*, 88–96. [CrossRef]

42. Wang, Y.; Gu, X. Functional divergence in the caspase gene family and altered functional constraints: Statistical analysis and prediction. *Genetics* **2001**, *158*, 1311–1320. [PubMed]

43. Koda, Y.; Tachida, H.; Soejima, M.; Takenaka, O.; Kimura, H. Ancient origin of the null allele se 428 of the human ABO-secretor locus (FUT2). *J. Mol. Evol.* **2000**, *50*, 243–248. [CrossRef] [PubMed]

44. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef] [PubMed]

45. Thompson, C.E.; Freitas, L.B.; Salzano, F.M. Molecular evolution and functional divergence of alcohol dehydrogenases in animals, fungi and plants. *Genet. Mol. Biol.* **2018**, *41*, 341–354. [CrossRef]

46. Yang, Z. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **1997**, *13*, 555–556. [CrossRef]

47. Nielsen, R.; Bustamante, C.; Clark, A.G.; Glanowski, S.; Sackton, T.B.; Hubisz, M.J.; Fledel-Alon, A.; Tanenbaum, D.M.; Civello, D.; White, T.J.; et al. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **2005**, *3*, e170. [CrossRef]

48. Zdobnov, E.M.; Apweiler, R. InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **2001**, *17*, 847–848. [CrossRef] [PubMed]

49. Kelley, L.A.; Sternberg, M.J. Protein structure prediction on the Web: A case study using the Phyre server. *Nat. Protoc.* **2009**, *4*, 363–371. [CrossRef]