# Functional and regulatory diversification of *Period* genes responsible for circadian rhythm in vertebrates

**Jun Soung Kwak*[1], M. Ángel León-Tapia*[2], Celian Diblasi[1], Domniki Manousi[1], Lars Grønvold[1], Guro Katrine Sandvik[1], Marie Saitou[1].**

***co-1st authors**

**Funding:**

The study was supported by The Research Council of Norway (grant no. 325874).

**Affiliations**

[1]Centre for Integrative Genetics (CIGENE), Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Norway

[2] Colección Nacional de Mamíferos, Pabellón Nacional de la Biodiversidad, Instituto de Biología, Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico

Corresponding Author: marie.saitou@nmbu.no

Email addresses
M. Ángel León <mal@st.ib.unam.mx>, Junsoung Kwak <junsoung.kwak@nmbu.no>, Guro Katrine Sandvik <guro.sandvik@nmbu.no>, Lars Grønvold <lars.gronvold@nmbu.no>, Domniki Manousi <domniki.manousi@nmbu.no>, Célian Diblasi <celian.diblasi@nmbu.no>

**Author Contribution:**

Junsoung Kwak: Conceptualization, Data curation, Investigation, Writing – Original Draft Preparation

Miguel Ángel León-Tapia: Conceptualization, Data curation, Investigation, Writing – Original Draft Preparation

Celian Dilasi: Investigation, Software

Domniki Manousi: Software

Lars Grønvold: Investigation, Software

Guro Katrine Sandvik: Conceptualization

Marie Saitou: Conceptualization, Investigation, Writing – Original Draft Preparation Writing – Review & Editing

## ***Abstract*:**

The Period genes (*Per*) play essential roles in modulating the molecular circadian clock timing in a broad range of species, which regulates the physiological and cellular through the transcription-translation feedback loop. While the *Period* gene paralogs are widely observed among vertebrates, the evolutionary history and the functional diversification of *Per* genes across vertebrates are not well known. In this study, we comprehensively investigated the evolution of *Per* genes, including de novo binding motif discovery by comparative genomics. We also determined the lineage-specific transcriptome landscape across tissues and developmental stages and phenotypic effects in public RNA-seq data sets of model species. We observed multiple lineage-specific gain and loss events of *Per* genes, though no simple association was observed between ecological factors and *Per* gene numbers in each species. Among salmonid fish species, the *per3* gene has been lost in the majority, whereas those retaining the per3 gene exhibit not a signature of relaxed selective constraint but rather a signature of intensified selection. We also determined the signature of adaptive diversification of the CRY-binding region in *Per1* and *Per3*, which modulates the circadian rhythm. We also discovered putative regulatory sequences, which are lineage-specific, suggesting that these cis-regulatory elements may have evolved rapidly and divergently across different lineages. Collectively, our findings revealed the evolution of *Per* genes and their fine-tuned contribution to the plastic and precise regulation of circadian rhythms in various vertebrate taxa.

**Keywords**: Circadian rhythm, gene family, molecular evolution, gene expression, development, natural selection

## Significance

The *Period* (*Per*) genes play essential roles in the circadian rhythm in animals. In this study, we comprehensively investigated the evolutionary diversification of the three types of *Period* genes in vertebrates. As a result, we observed a rapid evolution and sub-functionalization of these genes, especially adaptive diversification signatures in the protein-binding region, which plays a crucial role in regulating circadian rhythms. This underscores the fine-tuned contribution of *Per* genes in the biological clock's precision and adaptability across various vertebrate taxa.

## Introduction

The circadian clock regulates the physiological and cellular timing of organisms, adapting to the external environment with the 24-h light/darkness cycle of the solar day at the molecular level. The system is essential for an organism's rhythmic behavior and physiological responses, including sleep/wake cycle and foraging. The circadian clock is thus fundamental to the survival and adaptation of organisms in their respective habitats (Cermakian & Sassone-Corsi 2000; Schultz & Kay 2003; Roenneberg & Merrow 2016; Challet 2019). The biological rhythm can efficiently interact with the dynamic and temporal external resources and environments by adjusting its physiological processes (Patke et al. 2020; Umemura & Yagita 2020).

The first genetic factor of circadian rhythm regulation, the *Period* gene, was discovered in *Drosophila* mutant strains that showed unusually long or short biological rhythm patterns (Konopka & Benzer 1971; Bargiello et al. 1984; Zehring et al. 1984). The Period gene has also been reported in other invertebrates, such as insects, German cockroaches (Lin et al. 2002), gastropods (Aplysia and Bulla)(Siwicki et al. 1989), and crustaceans (American lobster) (Grabek & Chabot 2012). Mammalian homologs of the *Period* gene were also characterized soon after that (Tei et al. 1997), and the genetics, molecular mechanism, and behavioral biology of *Period* genes have been extensively studied in model organisms (Wager-Smith & Kay 2000; Cermakian et al. 2001; Pendergast & Friday 2009). These studies revealed that transcription-translation feedback loops regulate the critical role of the molecular circadian clock, core clock genes with changes in external cues, in which *Period* genes (*Per*) play the key role, at least in teleost fish and mammals (Langmesser et al. 2008; Relógio et al. 2011; Idda et al. 2012; Jensen et al. 2012; Lande-Diner et al. 2013; Ono et al. 2017). When there is enough light in the daytime, E-box, the DNA element is bounded by the transcription factors CLOCK-BMAL1, and the level of the PER and Cryptochrome (CRY) proteins are suppressed (Cox & Takahashi 2019). When night comes, and once PER and CRY levels have sufficiently dropped in the dark condition, a new cycle of CLOCK–BMAL1 begins to the *Per* and *Cry* transcription (Vallone et al. 2004; Patke et al. 2020).

Humans and mice have three copies of *PER* genes, *PER1*, *PER2*, and *PER3*, likely due to the two rounds of ancient genome duplications (von Schantz et al. 2006), a prime example of evolutionary processes by ancient gene duplications (Ohno 1970, 1999), highlighting the evolutionary significance of these genes. *Per* paralogues show some sequence similarity and overlapping functions, and specific functions for each gene. In mice, *Per1* and *Per2* genes are responsible for the circadian rhythm, while *Per3* did not indicate an observable, crucial role in circadian rhythm (Bae et al. 2001). On the contrary, a genetic variant in *PER3* in humans is associated with sleep disorders (Zhang et al. 2016).  In ray-finned fish, Wang (Wang 2008) highlighted that additional teleost per gene copies originated from a teleost-specific genome duplication, with different species retaining various per duplicates. In that study, the absence of *per3* in sticklebacks (Gasterosteidae) is also reported, indicating a lineage-specific evolutionary pattern of per gene family. Moreover, relaxed selection on *per1b* in Tinaja cavefish was reported (Idda et al. 2012; Mack et al. 2021), while Mack et al. (2021) demonstrated positive selection on per genes in *Squalius* freshwater fish species with population specificity. These studies suggest that the evolution of the *per* gene has diversified in response to ecological niches, contributing to variation in circadian behavior in the biological clock. The rapid turnover of circadian genes in amphibians also indicates the adaptive role of these genes (Stanton et al. 2022). However, despite recent advancements in genomic data, the comprehensive characterization and understanding of the evolutionary history and diversification process of *Per* genes in most non-model vertebrates remain a significant gap in our knowledge.

At the molecular and physiological levels, it has been thought that the *Per* genes are exclusively expressed in the suprachiasmatic nucleus, a brain region in the hypothalamus, which is responsible for controlling circadian rhythms (Takumi et al. 1998; Reppert & Weaver 2002). Contrary to the historical notion, studies revealed that *Per* genes are also expressed in cell lines and non-brain tissues (Balsalobre 2002), sometimes with circadian rhythmic expression patterns, such as in mouse liver (Siepka et al. 2007) cell lines (Ramanathan et al. 2014), and human cell lines (Gabriel et al. 2021). Currently, it is believed that the circadian clock is controlled by the master clock located in the suprachiasmatic nucleus with core neuron gene expressions and is hierarchically conveyed to peripheral clocks in tissues all over the body (Patke et al. 2020; Umemura & Yagita 2020). This revelation highlights the complexity and ubiquity of the circadian clock beyond the central nervous system, indicating a systemic regulation of circadian rhythms. Interestingly, studies suggested that circadian rhythm starts functioning even before birth in various species (Landgraf et al. 2015). For example, differential regulation of *per2* and *per3* expression during zebrafish embryogenesis is reported (Delaunay et al. 2003). It is reasonable that circadian genes are expressed in oviparous embryos with a transparent eggshell, which is exposed to light in the day and darkness in the night cycles. Studies indicated that circadian rhythm genes are expressed even in mammal embryos that cannot directly sense the day/night cycle by light (Shimomura et al. 2001; Yagita et al. 2010), and it is speculated that the mother transmits the fetus information about the day/night cycle so that the fetus can calibrate the circadian rhythms (Seron-Ferre et al. 2007; Landgraf et al. 2015). In addition, *Per* genes have non-circadian functions, such as DNA damage response (Fu et al. 2002) and embryonic neuron development (Nagata et al. 2019) underscores their multifaceted roles in biological processes. Despite numerous studies on *Per* expression in various conditions, the simultaneous investigation of *Per* gene paralogs remains unexplored. Thus, the understanding of their shared and species-specific roles is still yet to be investigated. This paper aims to bridge this gap by comprehensively investigating the evolutionary history and the functional diversification of *Per* genes in vertebrates.

## Results and Discussion

### Lineage-specific gene gain and loss of Period genes in vertebrates

First, we curated the *Per* genes in vertebrates (**Fig 1A**) using Ensembl Compara v106 and then visualized the evolutionary history of *Per* gene gain and loss in the species-based phylogenetic tree. After genome quality filterings (**Materials and Methods**), we included 154 species in total and obtained the sequences of 173 *Per1* genes, 212 *Per2* genes, and 150 *Per3* genes. There is no correlation observed between the reported *Per* gene number and the genome quality metrics (N50, **Supplementary Table1**) (rho= -0.00038, $p > 0.05$, Spearman correlation), excluding the possibility that the genome quality could influence the observed patterns of *Per* gene numbers. The phylogenetic signal (Pagel's lambda λ) (Pagel

1997) was high in the three *Per* genes reflecting a pattern where closely related species tend to be more similar in *Per* number than the more distantly related species. *Per1* and *Per 3* showed the strongest phylogenetic signal, and *Per2* also showed a high phylogenetic signal: *Per1* λ= 1.00, *Per2* λ= 0.94, and *Per3* λ= 1.00. Most mammalian and amphibian species have three paralogs (*Per1-3*) (**Fig 1A**), one for each family, yet we observed several lineages with marked gene gain/loss. Most reptiles lost the *Per1* gene, while birds lost the *Per3* gene, respectively **(Fig 1B**). The *Per2* number is the most stable among the three genes across taxa, and most species maintain the *Per2* gene, suggesting that *Per2* plays a major, universal role. The consistent presence of the *Per2* gene in the genomes across different taxa underscores its indispensable and universal significance. Cartilaginous fishes have been inferred to be basal extant jawed vertebrates, and their subclass *Elasmobranchii*, including Elephant shark (*Callorhinchus milii*), possess only one *per2* gene. In contrast, in most teleost fish species, two *per2* genes were observed, likely due to the teleost fish-specific whole genome duplication event (Ts3R) (Near et al. 2012; Berthelot et al. 2014; Inoue et al. 2015). In addition, there was a *Cyprinidae*-specific whole genome duplication (Cs4R) (Xu et al. 2019) and a *Salmonidae*-specific whole genome duplication (Ss4R) (Lien et al. 2016), which led to additional duplications of *Per* genes in each lineage **(Fig 1C**). Conversely, *per3* gene loss is observed in multiple distant fish lineages, including pike (*Esox lucius*), most salmonid species (Bolton et al. 2021), and cod (*Gadus morhua*), (**Fig 1C**). Although the loss of another circadian gene and retrogression of circadian rhythm is reported in cavefish (*Astyanax mexicanus*) (Mack et al. 2021), there is no specific trend observed in the *per* gene copy number in cavefish. We observed that *per3* genes are lost in most salmonid species, except for Atlantic salmon *(Salmo salar)*, and brown trout *(Salmo trutta)*, consistent with a previous study (Bolton et al. 2021). To test if *per3* gene is under relaxed selection in these two species, we used a codon-based statistical test to detect relaxed selection, RELAX (Wertheim et al. 2015), using the *per3* genes in these three species with retained *per3*, compared to the Acanthomorphata group. As a result, we observed a signature for intensified positive selection in the *per3* genes in these species with retained *per3*, compared to the control group (k=1.28, p=0.023) (**Fig S1**). This implies that the *per3* gene maintains indispensable roles in salmonid species that have this gene, which is contrary to the current view that *Per is the least important, based on earlier knock-out mouse studies, and thus, least studied among the paralogs* (Bae et al. 2001).

We also reconstructed ancestral *Per* states calculating for each gene states transitions based on an equal rate (ER), all-rates-different (AR), and symmetrical (SYM) using the hidden-rate model (Beaulieu et al. 2013). The best model of ancestral state transitions for *Per1* was ER (logL = -47.46), ARD for *Per2* (logL = -34.10), and SYM for *Per3* (logL = -45.26). Ancestral state estimations showed a similar pattern for the three *Per* genes in the vertebrate's phylogeny, in which the current observed gene number was derived from a single ancestral copy (**Fig 2**). Dynamic gain/loss pattern changes were more recent clades in *Per1* **(Fig 2A**) and *Per3* (**Fig 2C**) than the duplication pattern noticed in *Per2* (**Fig 2B**). Disparity Through Time (DTT) (Harmon et al. 2003) plots for each *Per* gene revealed differences in the patterns of evolutionary history (**Fig 2D**). *Per1* showed a major increase in the among-clade disparity while *Per2* showed stability, likely driven mainly by the duplication in the whole fish clade, with some slight peaks of within-clade disparity led by the diversification in Cyprinidae, indicating that diversification was high later in the evolutionary history of the group. *Per3* had a pattern of gradual increase in the among-clade disparity.

The evolution of circadian rhythm has been discussed in the context of ecological adaptation and behavioral strategies in both terrestrial and aquatic species, including temperature, diet, habitat, reproduction, and migration; although the effect of these environmental and ecological factors on each species is complex, mutually influential, and variable (Horton 2001; Bloch et al. 2013; Oliveira et al. 2016; Hewitt & Shaikh 2021). We hypothesized that some environmental factors may have driven the evolution of *Per* gene gain and loss (**Material and Methods**). To investigate the evolutionary association between each *Per* gene number and in each species, we conducted the Phylogenetic Generalized Least Squares test (Symonds & Blomberg 2014) between per gene copy number and environmental/ecological factors of 138 species with data available. A total of ten environmental/ecological factors we tested included elevation, temperature, precipitation seasonality, solar radiation, windspeed, water vapor pressure, and net primary productivity, and ecological values included migration, body mass, and habitat (**Materials and Methods**). To avoid noise due to taxa-specific factors, we separated each taxon (mammals, birds and reptiles, amphibians, and fish) in the analysis (**Materials and Methods**). As a result, we did not observe any statistically significant association between *Per* gene numbers and each environmental or ecological variable (Phylogenetic Generalized Least Squares test, *p*-value less than 0.05 was considered to be significant). We assume that influence of environmental and ecological factors on circadian rhythm is based on a more intricate system. For instance, as a result of a spatiotemporal regulation of the enzymatic activity, rather than the simple *Per* gene copy number. In mammals, compelling evidence underscores the presence of circadian synchrony between predator-prey activity patterns (Caravaggi et al. 2018). These intricate interspecific relationships have emerged as pivotal factors, propelling organisms towards physiological and behavioral adaptations (Bennie et al. 2014). These adaptations, in turn, may wield significant influence as selective pressures in vertebrates, instigating genetic changes within species and shaping their capacity to exploit available resources effectively.

We also examined if the copy gain and loss pattern of each *Per* gene family deviated from other gene families available at Ensembl v104 compara gene trees, as previously reported  (Gundappa et al. 2022). We calculated the coefficient of variation of gene numbers of each gene, and found three gene groups, with (A) low variation, (B) medium variation, and (C) high variation. *Per* genes fell in the medium variation cluster. Genes with very low variation included highly conserved essential genes such as Kinesin family-like genes and genes with high variation included dynamically evolving genes, including olfactory receptors (**Fig S2, Table S5**).

## Evolutionary conservation and divergence in *Period* genes function across taxa

Previous studies with intensive genetic engineering and behavioral experiments, primarily in mice, suggested that the *Per3* gene has no intrinsic effect on circadian rhythms compared to the *Per1* and *Per2* genes (Bae et al. 2001), and *Per3* is often excluded from later investigations and discussions probably for this reason (Chiou et al. 2016; Cox & Takahashi 2019). However, this makes it difficult to reveal the comprehensive dynamics of how *Per* gene paralogs share roles, and thus how conserved these roles are across species is not known. To investigate if each *Per* gene function is shared or specific and how much they are evolutionarily conserved, we curated and re-analyzed published phenotype datasets. We first inferred that vertebrate *Per* genes are likely to have a single ancestor by constructing a phylogenetic tree based on the amino acid sequences of Per from vertebrate species, with the *Drosophila* Per gene sequence. The *Drosophila* Per sequence was clustered with the Per3 group, suggesting that *Per3* is the most ancestral form of the *Per* gene family (**Fig 3A**). It is speculated that the ancestral *Per3* gene was duplicated twice due to the first and second vertebrate whole genome duplications and further duplicated by the teleost-specific whole genome duplication corresponding with previous studies (Wang 2008). The phylogenetic analysis also supported that lineage-specific whole genome duplications in common carp (*Cyprinus carpio*) and salmon (*Salmo salar*) have also likely contributed to their additional *per* gene duplications.

We next investigated the regulatory and functional differentiation between *Per* paralogs by analyzing public gene-phenotype data sets of multiple species. We examined the genome-wide association study hits of variants in the three *PER* genes in humans that are curated and summarized at GWASATLAS (Watanabe et al. 2019). The gene-level p-value of association with traits was computed by MAGMA (de Leeuw et al. 2015), a computational method based on a multiple linear principal components regression model to aggregate the effects of variants on traits in a gene. We extracted traits that showed significant association (*p*-value smaller than $10^{-9}$) in any of *PER* genes in the human GWAS database (**Fig 3B**). Human *PER1* showed the highest statistical significance with height and medium significance with chronotype and blood-related traits, such as hematocrit, and the percentage of red cells in the blood. *PER2* is also moderately associated with chronotype. *PER3* showed the highest statistical significance with chronotype and also showed some association with height and another blood-related trait, platelet volume. A deletion (54 bases) polymorphism was reported in the genomAD paper (Collins et al. 2020), which is associated with sleep (Archer et al. 2018). This variant is not reported in the 1000 Genomes phase 3 dataset (The 1000 Genomes Project Consortium et al. 2015) and the updated 1000 genome dataset (Byrska-Bishop et al. 2022) as it falls in a complex repetitive region, and little is known about the molecular function and evolutionary significance of the repeats.

It is reported that single *Per1* or *Per2* knockout mice have shorter circadian periods with reduced precision and stability, and both *Per1* and *Per2* knockout mice are completely arrhythmic in constant conditions (Cermakian et al. 2001; Zheng et al. 2001; Bae et al. 2001). Interestingly, *Per3* knockout mice have relatively moderate changes in circadian rhythm, observed light-dependent phenotypes, and reduced body size, growth, and metabolism (Archer et al. 2018; Pereira et al. 2014). In our meta-analysis, using data from Mouse Genome Database (MGD), 2021 (Blake et al. 2020), *Per2* gene has a broad effect on various phenotypes, while *Per1* and *Per3* are only involved in several neurological traits (**Fig 3C**). In zebrafish (*Danio rerio*), based on the Zebrafish Information Network (zfin.org), 2022 (Bradford et al. 2022), it is reported that *per1b* or *per2* deficiency has effects on locomotor, photoperiod, neural development, and neutrophil migration, a blood-related trait similar to the human case, interestingly. The *per1b* gene has specifically reported phenotypic effects such as memory, and *per2* has specifically reported phenotypic effects such as eye synapse morphology. Meanwhile, there are no marked effects reported for *per1a* and *per3*, except for some expression changes of circadian-related genes (**Fig 3D**). These observations indicate the flexibility of each *Per* gene's roles between species.

## Dynamic evolution of *Period* gene expression pattern between paralogs and species

To understand the regulatory differentiation that may underline such diversity in phenotypic effects of *Per* genes across species, we searched for the *Per* gene expression in various tissues in vertebrates. We compared adult *PER* gene expressions of humans (*Homo sapiens*), from the GTEx portal (GTEx Consortium et al. 2017), mouse (*Mus musculus*) expression from the mouse Gene Expression Database (GXD) (Baldarelli et al. 2021), and African clawed frog (*Xenopus laevis)* Xenbase (Fortriede et al. 2020) **(Fig 4A-C).** In humans, *PER1* showed the highest expressions among *PER* genes overall except for Cerebellum. Notably, *PER3* shows a marked expression level at the Cerebellum among other organs, which is strongly influenced by the sleep-wake cycle (Canto et al. 2017) while it does not show such high expression in other parts of the brain. The *Per2* gene shows mostly low expressions in both the human and the clawed frog, but in the mouse, the *Per2* gene shows a moderate expression level and its expression pattern across species is markedly similar to the *Per3* gene expression. Two *Per3* gene duplicates in the clawed frog, which experienced allotetraploidization around 17–18 Mya (Session et al. 2016), showed similar expression patterns each other across tissues **(Fig 4C)**. The orthologous genes, such as *PER2* in humans and *Per2* in mice, did not necessarily show a strong correlation in their expression pattern between species **(**r=-0.07, Spearman's correlation, **Fig 4D)**, suggesting the dynamic evolutionary plasticity of the expression pattern. Rapid transition of tissue-specificity of gene expression has been reported (Fukushima & Pollock 2020), in particular between the ovary and testis, as we also observed in *PER1* in the human and *Per1* in the mouse **(Fig 4AB).**

To further investigate the evolutionary trajectory of this gene expression regulation, we compared the expression pattern of relatively new *Per3* gene duplicates in the African clawed frog, *Xenopus laevis*, which experienced allotetraploidization around 17–18 Mya (Session et al. 2016) and the Japanese medaka (*Oryzias latipes*) which has likely duplicated the *per2* gene because of the teleost genome duplication around 310 Mya (Davesne et al. 2021). The newly duplicated two *Per3* genes in *Xenopus laevis*, showed a very similar expression pattern both across tissues and across developmental stages Data from Xenbase (Fortriede et al. 2020) (**Fig 4E**), such as high expression in the ovary, liver, brain, and lung, and stage 8 and 9, showing their regulatory mechanism is still very similarly conserved. On the contrary, the two Japanese medaka *per2* genes were already differentiated at the regulation level at Medaka Omics Data Portal (Li et al. 2020) (**Fig 4F**). Collectively, the *Per* gene expression pattern in tissues is not necessarily conserved across species and is dynamically diversifying, probably due to the rapid evolution of gene regulation (Rifkin et al. 2005; Hill et al. 2021). We further asked about the developmental pattern of *PER* gene in different parts of the human brain using the Brainspan dataset (https://www.brainspan.org/) (Johnson et al. 2009; Kang et al. 2011) **(Fig S3)**. Overall, *PER1* showed the highest expression across tissues, especially in the 37 post-conception weeks fetus, except for the cerebellar cortex. *PER3* shows high expression in the cerebellar cortex after birth when humans are exposed to the external light and dark cycle. *PER3* also showed high expression in the thalamus, which acts as a sensory hub (Hwang et al. 2017). Neither the suprachiasmatic nucleus nor the broader region, the hypothalamus, which is responsible for controlling circadian rhythms (Takumi et al. 1998; Reppert &

Weaver 2002), are reported in Brainspan. In considering the function of *Period* genes, we need to consider their circadian rhythms. Although there have not been many reports on the rhythmicity of all the *Per* genes in one study, West et al. (West et al. 2020) measured the expression of circadian gene paralogues in Atlantic salmon during development in various tissues by using JTK cycle. We extracted the rhythmicity of *Period* gene expression from West et al. (West et al. 2020) (**Fig 4G**). In the Optic tectum, all the expressed *per* genes showed rhythmicity, while in saccus vasculosus, three of seven salmon *Period* genes showed rhythmicity. In gill, only *per1* on chr4 showed moderate rhythmicity.

### Evolutionary conservation and diversification at the sequence level.

*Per1*, *Per2*, and *Per3* do not necessarily show the same phenotype and expression pattern between species, rather, these genes appear to be evolving their expression patterns rapidly. To investigate the genomic basis of this dynamic evolution of the *Per* gene regulation, we investigated how the coding sequences and putative regulatory regions of *Per* genes evolved.

We first explored if there are any sites under diversifying selection on *Per1*, *Per2*, and *Per3* vertebrate genes independently, with MEME (Mixed Effects Model of Evolution) (Murrell et al. 2012) on Datamonkey v2.0 (Weaver et al. 2018) (**Table S6**). MEME estimates a site-wise synonymous and a two-category mixture of non-synonymous rates and uses a likelihood ratio test to determine the signal from episodic diversification, a combination of strength of selection and the proportion of the lineages affected. As a result, we observed signatures of diversifying selection in the CRY-binding sites in *Per1* (site “2566” in **Table S6** for all species and site “1583” in **Fig 5B** with the selected species) and *Per3* (site “2102” in **Table S6** for all species and site “1393” in **Fig 5B** with the selected species), as shown in **Fig 5AB** and **Fig S4,** contrary to the general notion that protein binding sites in a gene are evolutionarily conserved (Guharoy & Chakrabarti 2010). As summarized in the introduction section, the PER and Cryptochrome CRY proteins orchestrate the circadian feedback loop with the CLOCK–BMAL1 (Vallone et al. 2004; Patke et al. 2020). CRY-binding region of the *Per* gene was characterized in rats (*Rattus norvegicus*) (Miyazaki et al. 2001). The site under selection in CRY-binding region is conserved, but after that, the latter binding region shows a dramatic decline in conservation (**Fig 5A and 5B)**, which may indicate sub-functionalization between paralogs, and the molecular basis of functional diversification of *Per* genes between species. *Per2* gene showed a marked signature of diversifying selection around the site 263 in the alignment (**Fig S4, TableS6),** while we did not find known functional annotation there but “Polar residues, Compositional bias” in Uniprot (corresponding with the 46th “N” in NE**N**CS in the human sequence, O15055 · PER2_HUMAN at UNIProt, https://www.uniprot.org/uniprotkb/O15055/entry, last accessed, July 3rd 2023 (Deutsch et al. 2023).

We also observed variations in gene lengths, which are presented in **Table S7**. While a subset of genes appeared short, establishing a definitive cutoff for functional loss based solely on gene length proved challenging. Consequently, we expanded our investigation to include domain analyses to identify the presence of conserved functional domains as indicators of potential gene activity. Interestingly, some genes, such as the per3 gene in sheep (ENSOARG00020020217) without recognizable domains reported (**Table S8**), still showed rhythmic expression patterns, suggesting functionality (Varcoe et al. 2014). This underscores the complexity of inferring gene pseudogenization from domain absence or gene length alone.

We further investigated the conservation and putative regulatory regions of *Per* genes. There have been multiple studies that conducted a comparative analysis of the *Per* gene regulatory region. Nakahata et al. (Nakahata et al. 2008) proved that two tandem E-box-like sequences in tandem with a six-base-pair interval are necessary for cell-autonomous oscillation in the mammal and are conserved in various mammalian species. Here, we explored the conservation of E-box elements, expanding their investigations. At the same time, Paquet et al. (Paquet et al. 2008) proposed a model to predict the CLOCK-controlled cis-elements with E-boxes that regulate *Per* gene expression in various species, from *Drosophila* to mammals; Paquet et al (Paquet et al. 2008). also indicated that the E-box sequence is highly conserved between fish and mammals.

As general Blast search is not fine-tuned for searching such motif sequences with short query sequences with very rapid turnovers, we used Jbrowse (Buels et al. 2016; Diesh et al. 2022) -based SalmoBase (Samy et al. 2017) (https://salmobase.org/, last accessed, January 5th 2023) to investigate the promoter sequences, allowing ambiguous spacer sequences around the *per* genes in Atlantic salmon. We searched for the three mammalian sequences reported in Nakahata et al. (Nakahata et al. 2008), allowing the six-base spacer sequences for against 100K upstream and downstream within each *Per* gene region. Of the combination of three query sequences and seven salmon *per* genes, we only found one *per1-regulatory-like* sequence, CACGTGatctggTGTGTG, near the *per3*, but over 80K apart from the gene, and in a repeat region (**Fig 5C**). In Ensembl Comparative Region Comparison Analysis using the 65 fish EPO-Extended, the sequence (Ssal_v3.1, 22:25993822-25993839) was observed only in the Atlantic salmon genome and Brown trout (*Salmo trutta*) genome. It is thus plausible that the mammalian *Per1*-regularoty-like sequence has arisen independently from mammals in this lineage, and the ancestral sequences underwent rapid turnover and are not detectable. The sequence that Paquet et al. (Paquet et al. 2008) estimated corresponds to Japanese medaka HdrR (ASM223467v1) 5:28606354-28606380 region, near the *per3* gene, while it is located near the *Per2* sequence in mammalian species (Paquet et al. 2008). In Ensembl 65 fish EPO-Extended, this region is widely observed in *Percomorpha* (**Fig 5D**) as represented in Paquet et al. (Paquet et al. 2008). However, the zebrafish (GRCz11, 2: 48,352,711-48,352,738) sequence is located near the *per2* gene, and this region was not detectable in other species. This implies that at least the zebrafish sequence evolved to become similar to the regulatory sequence in its lineage, independent from other lineages. These observations of salmon and zebrafish sequences resemble *per* regulatory regions, suggesting the fast rise and fall of the regulatory sequences concordant with previous observations on the evolution of regulatory elements in general (Stone & Wray 2001; Neme & Tautz 2016).

To uncover potential cis-regulatory elements associated with the per1, per2, and per3 genes, we conducted a comprehensive, unbiased *de novo* motif search for 1000 base upstream sequences of per1, per2, and per3 genes using the MEME suite (Bailey & Elkan 1994; Bailey et al. 2009, 2015). Our analysis aimed to identify enriched sequences that could potentially serve as cis-regulatory elements. While MEME identified 10 sequences with the lowest *p*-values, we also observed that some of the sequences were merely repetitive sequences, which are likely false positives and unlikely to be cis-regulatory mechanisms. Therefore, we chose to focus on the sequences with the highest bit scores per gene. The putative regulatory sequences we identified are as follows (**Figure 5E, Table S9**). *Per1*: CCGCCCCCGGATAAAATTAGCC (found in *Eutheria*), *Per2*: CTTGTATAACCGTTTAGATAAC (found in *Teleostei*), *Per3*: GGARGCCTAAATATAGGAGGCG (found in *Boreoeutheria*). These results highlight the lineage specificity of these sequences, suggesting that these cis-regulatory elements may have evolved rapidly and divergently across different lineages. Collectively, the rapid sequence and regulatory evolution that we observed, despite the relatively stable gene copy number, are likely to contribute to the dynamic effect of the *Per* gene family on the phenotype, both circadian and noncircadian effects in various species.

## Conclusion

One of the major questions in evolutionary genomics is how duplicated genes develop the diverse phenotypes in various species by fine-tuning their original functions. This study concentrated on the evolutionary trajectory and functional diversification of the *Per* gene family in vertebrates, key regulators of circadian rhythms. We unveiled how duplicated copies of *Per* genes have adapted their functions across species. While *Per1*, *Per2*, and *Per3* paralogous genes were duplicated in the common ancestor of vertebrates. Notably, most birds have lost the *Per1* gene, whereas the majority of reptiles have shed the *Per3* gene. The *Per2* gene, however, has been retained across most species, suggesting its crucial, universal role in circadian regulation. Contradicting the prevailing perception that *Per3* is of lesser importance (Bae et al. 2001), our analysis indicates intensified selection of *Per3* in certain salmonids, as opposed to relaxed selection. This challenges the current understanding and underscores the gene's significant yet overlooked role. Through comparative analysis of public genome, phenotype, and transcriptome datasets, we determined that the phenotypic impacts and expression patterns of *Per* gene paralogs exhibit variability across species and paralogs. This variability underscores the rapid evolution of sub-functionalization within the *Per* gene family, indicating a dynamic evolutionary response to diverse environmental and physiological demands.

We postulate that as more transcriptome and gene-phenotype datasets of a variety of species become available, not only genome datasets but the evolution of gene function and regulation will be more comprehensive in the context of how genome evolution contributes to cross-species phenotypic diversity through molecular mechanisms. Although knockout/knock-in studies are a powerful way to understand gene function in a controlled environment, the effects of each *Per* gene are often not investigated within one study. Therefore, when we encounter a lack of known phenotype report of a gene compared to another, it is difficult to know if this is because that gene was "not investigated" or "was examined, but no observable changes were found/thus not reported”, which remains a limitation to investigating the cross-species effects of the paralogs. Our study represents the first step in integrating multiple angles of -omics datasets to understand the functional effect of each *Per* gene in multiple species. It is important to note that although there are interesting lineage-specific gene loss trends in reptiles and birds at the genomic level, currently, there are not many comprehensive cross-tissue transcriptome data and gene-phenotype data available in these taxa. Also, in most transcriptomic studies, the sampling time is not considered and reported, which may raise limitations for comparative circadian rhythm gene investigation. While large-scale international genome sequencing projects for various non-model organisms are underway, the next step for evolutionary genomics is to connect insights from genome information to their functional roles on molecular and macroscopic phenotypes and environmental conditions.

## Material and Methods

### *Gene Sequences*

We have retrieved all available *Per1*, *Per2*, and *Per3* sequences from Ensembl Compara v106, via the Wasabi app (http://wasabiapp.org/). We initially obtained 200 species. To filter out species with potential low-quality genomes, we removed 18 species with sequences containing “N” (script available in Supplementary Material). We also excluded 25 species that have any *Per* genes not starting from methionine code, which suggests patchy gene annotations. We further removed 10 species with inconsistent reports, which had zero *Per1*/*Per2*/*Per3* genes in Ensembl Compara but with *Per1*/*Per2*/*Per3* existing in NCBI orthologs (last accessed, August 15t$^{h}$ 2022) and two species with projection-built genomes. We tested Spearman’s correlation between the observed Per gene number and the genome quality metrics (N50, obtained from NCBI). We examined the copy gain and loss pattern of *Per* gene in relation to the other gene families, calculating the coefficient of variation for each one using cmstatr package (https://doi.org/10.21105/joss.02265) in R software v4.2.2 (https://www.R-project.org/).

To inspect likely pseudogenes, we removed branches of gene trees that were significantly longer than other genes (likely because they accumulated mutations at a faster rate than others, which is a sign of pseudogenization) by using TreeShrink (Mai & Mirarab 2018), a software that detects branches that significantly increase the tree diameter. We set the chance of false positives to 10 percent. We thus identified two putative pseudogenes for per3 (ENSAMEG00000030567 from giant panda & ENSCJAG00000000288 from White-tufted-ear marmoset), that were removed from further analysis.

Lastly, we removed six species that are not available at TimeTree5 (Sudhir Kumar et al. 2022) and four non-vertebrate species, which additionally four do not have geographic range data on Map of Life/IUCN (last accessed, August 15th 2022). A total of 133 vertebrate species were used for subsequent analyses, species lists and *Per* gene numbers are available in **Table S1**. Amino acid alignments were performed using MAFFT v7 (Katoh et al. 2002). For tree construction, the neighbor-joining distance-based method was applied (Saitou & Nei 1987), focusing on conserved sites across 363 amino acids. The nucleotide substitution model employed was JTT (Jones et al. 1992). Bootstrap analysis was conducted to assess the robustness of the tree, with the number of resampling events set to 100.

### Evolutionary history of gene gain and loss of *Period* genes in vertebrates

We employed a phylogenetic framework that incorporated the species of interest, selected from the TimeTree5 dataset (Sudhir Kumar et al. 2022) (**Supplementary file S2**) accessible via www.timetree.org. The comprehensive tree is the product of a compilation of published molecular trees, the estimation of divergence times, and the construction of a supertree using the hierarchical average linkage method (HAL) of clade pairs; methodological details are available (Hedges et al. 2015). This tree was used to infer the evolution of each *Per* gene in vertebrates. We reconstruct ancestral *Per* states calculating for each gene state transitions based on equal rates (ER), all-rates-different (AR), and symmetrical (SYM) using the hidden-rate-model (Beaulieu et al. 2013) and one category for each level of the trait. The best model was selected according to the Akaike Information Criterion and plotted into the tree, all analyses were performed using ape v5.6 (Paradis & Schliep 2019) and phytools v1.5.1 (Revell 2012) packages. We estimated the phylogenetic signal for each *Per* gene based on Pagel’s lambda λ parameter and the best rate transition for each gene. To examine how *Per* genes changed through time in vertebrates, we used a disparity-through-time (DTT) approach, this analysis quantifies the disparity for the whole clade and each node in the phylogeny (subclade). The relative disparity is calculated by dividing each subclade disparity value by the overall disparity of the clade, whereas the mean relative disparity is calculated for each subclade present at each divergence point. Values close to zero suggest that the variation is partitioned among the distinct subclades in the tree (slowdown in trait diversification), while values near one indicate that a major proportion of the total variation is contained by the subclades in the tree (rapid trait diversification) (Harmon et al. 2003), we performed these analyses in geiger v2.0.10 package (Pennell et al. 2014).

### ***Association between environmental and ecological factors and Period gene numbers***

#### *Selection of environmental variables*

Besides internal mechanisms, circadian rhythm oscillations in vertebrates are driven by abiotic environmental changes (Zheng et al. 2021). Some of the main environmental factors include the light-dark cycle as the primary factor (Junko et al. 2019), the temperature (Vera et al. 2023), food availability (Vinod Kumar et al. 2022) and oxygen consumption (Adamovich et al. 2022) among others. Based on this information, we chose a set of terrestrial and marine world raster layers as likely factors to predict the evolution of *Per* genes. For terrestrial environments, six climatic layers at 10 arc min (~340 km²) were obtained from Worldclim v2.1 (Fick & Hijmans 2017): altitude, annual temperature range, annual precipitation seasonality, annual mean solar radiation, annual mean wind speed, and annual mean water vapor pressure. Additionally, we incorporate the Net Primary Productivity of the MOD17A3 v55 layers at 0.1 degrees downloaded from NASA for 2000-2016 (Running et al.), mean value was calculated for each year, and finally, the mean value for the whole period. For marine environments, we selected bathymetry and annual mean temperature range downloaded from MARSPEC at 10 arc min (Sbrocco & Barber 2013), and annual wind speed downloaded at 5 arc min from Global Marine Environment Datasets v2.0  (Basher et al. 2018).

All raster layers were scaled to 10 arc min. The geographic range species maps were obtained from the International Union for Conservation of Nature (IUCN. 2022. The IUCN Red List of Threatened Species v2022-2. https://www.iucnredlist.org. Accessed on October 2022) and Map of Life (Jetz et al. 2012) for extracting raster values and mean values calculated for each species using raster v3.6 and exactextractr v0.9.1 (Baston) in R. The mean latitude was calculated for each species based on the geographic range maps.

#### *Selection of ecological variables*.

Circadian rhythms, shaped by cyclical environmental changes, drive activity patterns in organisms, influencing both physiological and behavioral responses. These activities may be influenced by other factors, including complex ecological factors (Caravaggi et al. 2018). To explore the potential correlation between ecological traits and the evolution of *Per* genes, we accessed ecological trait data for key taxonomic groups from publicly available databases. For fishes, migration, body mass, vision, and swimming mode were downloaded from (Froese et al. 2010) through rfishbase v4.0.0 R package (Boettiger et al. 2012). For amphibians, body mass, habitat, and migration were downloaded from AmphiBio (Oliveira et al. 2017). For birds, body mass, habitat, and migration were downloaded from AVONET (Tobias et al. 2022). For mammals, activity cycle, body mass, and habitat breadth were downloaded from PanTHERIA (Jones et al. 2009).

### *Comparative statistical analyses*

To investigate the nature of the evolutionary association between predictors and *Per* gene number we used a Phylogenetic Generalized Least-Squares analysis (PGLS; (Symonds & Blomberg 2014). Analyses were performed for five taxonomic groups/datasets “Mammals” “Amphibians/Reptiles/birds”, “All fishes”, “Continental fishes”, and “Marine fishes”. A violation of the assumptions about the residuals or model instability can severely affect the conclusions drawn, and hence, it is of crucial importance that these are thoroughly checked and an assessment is made about how much the models can be trusted (Mundry 2014). Continuous variables were log10 transformed, for altitude/bathymetry predictor with negative values were squared before log-transformed. The distribution of all transformed variables was checked and tested for collinearity between predictors using the Variance Inflation Factors (VIF) with car package v3.1 (Rasco 2020) before running all models. Variables with no normality and a VIF major of 5 were discarded from the final model.

PGLS models were individually computed for each Per gene using the final predictors, and a phylogenetic signal was determined through Pagel’s lambda (λ), estimated by maximum likelihood (Pagel 1997), and implemented using caper package v1.0.1 (<https://CRAN.R-project.org/package=caper>). Diagnostic plots of residuals, including outlier analysis and Q-Q plots, were thoroughly examined to ensure that assumptions were not violated. A comprehensive overview of the species and the *Per* number, environmental, and ecological data analyzed is provided in **Table S3**.

### *Gene-phenotype association datasets*

Reported phenotypes associated with *PER1*, *PER2,* and *PER3* in humans at GWAS Atlas (Watanabe et al. 2019) under PheWAS. Traits with a -$\log_{10}$ p-value of more than nine in at least one of the *PER* genes were considered to be statistically significant and included in the analysis. Traits in the *Per* deficient mouse were retrieved from the Mouse Genome Database (MGD), 2021 (Blake et al. 2020). Traits in the *per*-deficient zebrafish were retrieved from the Zebrafish Information Network 2022 (zfin.org) (Bradford et al. 2022).

### *Transcriptome datasets*

The human adult transcriptome data set was retrieved from GTEx portal v8 (last accessed: January 18th 2023) (GTEx Consortium et al. 2017). The mouse transcriptome dataset was retrieved from The mouse Gene Expression Database (GXD) (Baldarelli et al. 2021) Mouse embryo dataset: E-ERAD-401, Mouse adult datasets: E-MTAB-2801 and E-MTAB-6798. Frog (*Xenopus laevis*) transcriptome dataset was retrieved from Xenbase (Fortriede et al. 2020). For adult datasets, we selected organs with available gene expression data in all three species for comparison purposes (muscle, ovary, testis, kidney, liver, spleen, heart, lung, and brain or cerebellum). We used Spearman's rank correlation coefficient test to investigate the correlation between gene expression patterns between paralogs.  For the developmental transcriptome analysis, we used expression data sets from African clawed frog (*Xenopus laevis)* Xenbase (Fortriede et al. 2020), medaka (*Oryzias latipes*) at Medaka Omics Data Portal (Li et al. 2020), and the rhythmicity of gene data sets in Atlantic salmon (*Salmon salar*), *Optic tectum*, gill, and *saccus vasculosus*. Data is retrieved from supplementary table 8 of (West et al. 2020), except for one gene with no expression. For these analyses, we investigated the expression divergence among genes following genome duplication events. Duplicated genes, especially after whole genome duplications, need a rigorous and often species-specific approach to gene annotation. Therefore, for the target species in this analysis that have experienced duplications, we leveraged species-specific datasets for gene subtype annotations where available.

### *Molecular evolution analysis*

Intensification or relaxation of selection was estimated using RELAX (Wertheim et al. 2015) implemented in HyPhy v2.5 (Kosakovsky Pond et al. 2019). This software compares a set of “test” branches to a set of “reference” branches and measures intensification ($k > 1$) or relaxation ($k < 1$) of selection in the test group compared to the reference group. The salmonid group (Atlantic salmon, brown trout) *per3* sequences were used as a test group and compared to sequences of the outgroup (Acanthomorphata) as references.

We explored if there are any sites under diversifying selection on *Per1*, *Per2*, and *Per3* vertebrate genes independently, with MEME (Mixed Effects Model of Evolution) (Murrell et al. 2012) on Datamonkey v2.0 (Weaver et al. 2018). We analyzed each *Per* gene independently as *Per1*, *Per2*, and *Per3* paralogs are too diverged to align and thus expected to reduce statistical accuracy if analyzed simultaneously. MEME estimates a site-wise synonymous and a two-category mixture of non-synonymous rates and uses a likelihood ratio test to determine the signal from episodic diversification, a combination of strength of selection and the proportion of the lineages affected.

### *Transcription binding sites analysis*

We explored the conservation of E-box elements, expanding previously reported investigations. As general Blast search is not fine-tuned for searching such motif sequences with short query sequences with very rapid turnovers, we used Jbrowse (Buels et al. 2016; Diesh et al. 2022) -based SalmoBase (Samy et al. 2017) (https://salmobase.org/, last accessed, January 5th 2023) to investigate the promoter sequences, allowing ambiguous spacer sequences around the *per* genes in Atlantic salmon. We searched for the three mammalian sequences reported in Nakahata et al. (Nakahata et al. 2008), allowing the six-base spacer sequences for against 100K upstream and downstream within each *Per* gene region.

| | |
|---|---|
| *per1* | caggtcCACGTGcgcccgTGTGTGtgacac |
| *per2* | cgcggtCACGTTttccacTATGTGacagcg |
| *per3* | gaccggCACGCCgcgagcCTCGAGactgcg |

We also searched for the predicted CLOCK-controlled cis-elements with E-boxes in figure 5B in Paquet et al (Paquet et al. 2008):

| | |
|---|---|
| Medaka | cgtTCACGTgga-tccccatGTTACA |
| Zebratish | cggTCACCTgtt-tctccacATGCTG |

By using Ensembl Comparative Region Comparison Analysis using the 65 fish EPO-Extended chain.

We also conducted an unbiased search using the MEME v5.5.5 (Multiple Expectation maximizations for Motif Elicitation) *de novo* motif discovery results using the 1000 base sequences upstream (the script is available as a Supplementary Material, **file S1**) of each *Per* gene on MEME Suites (Bailey et al. 2015). We used the following parameters for the three *Per* genes:

```
meme PER1_5prime.fa -dna -oc . -nostatus -time 14400 -mod zoops -nmotifs 10 -minw 18 -maxw 22 -objfun classic -revcomp -markov_order 0 .
```

Our analysis aimed to identify enriched sequences that could potentially serve as cis-regulatory elements. While MEME identified 10 sequences with the lowest p-values, we observed that some of the sequences were merely repetitive sequences, which are likely false positives and unlikely to be cis-regulatory mechanisms. Then, we only considered the reported sequence with the top bits score for each gene.

## Data Availability

We used publicly available datasets, which is described in the Materials and Methods section. Scripts and Supplementary result files are available at FigShare (doi: 10.6084/m9.figshare.25924420).

## Acknowledgments

We express our gratitude to Drs. Shona Wood, Nicola Barson, and Siri Fjellheim for the discussion and comments on the early stage of this study.

## Figures and Tables

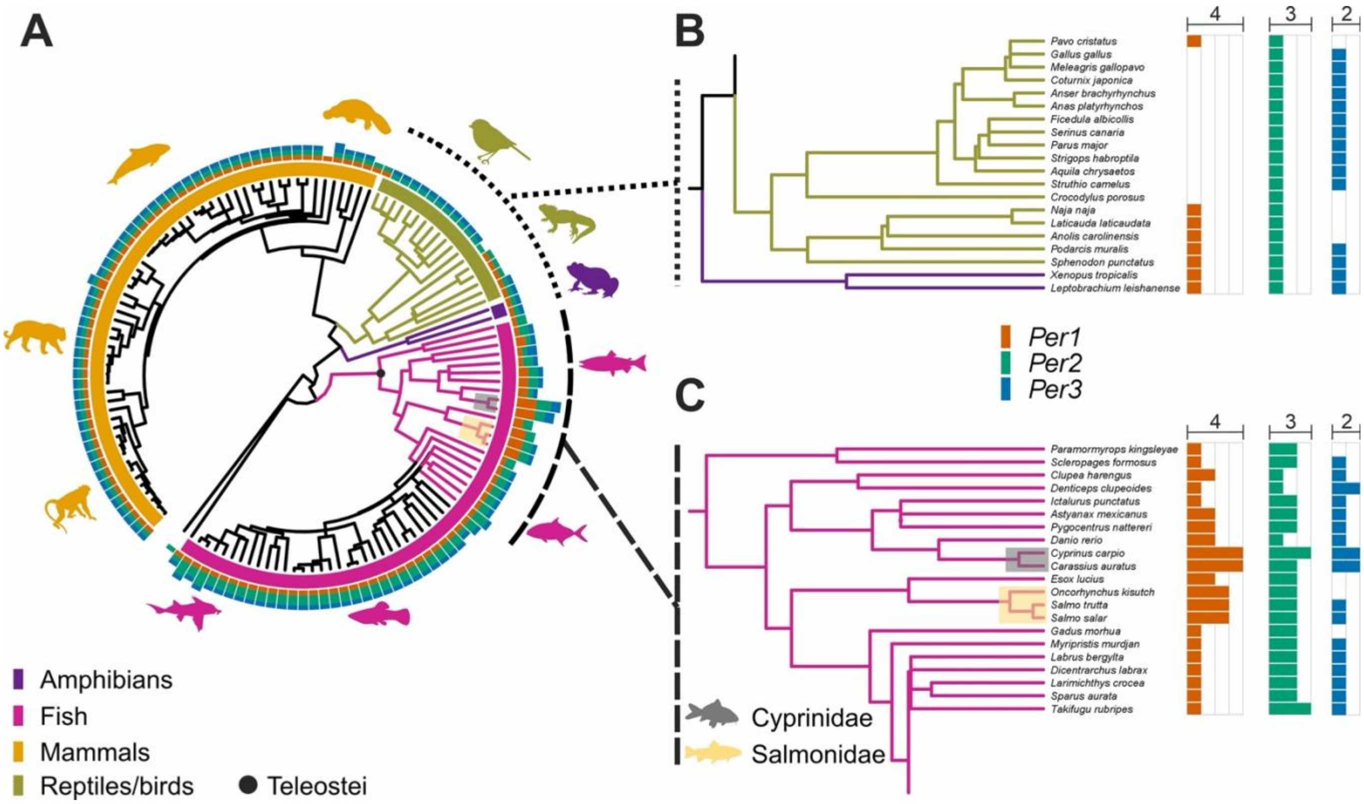


**Fig 1.** The evolutionary history of the *Period* gene (*Per*) gain and loss in vertebrates. **A**. Phylogenetic tree and numbers of *Per* paralogs in vertebrates. Bar colors around the tree denote the main vertebrate taxonomic clades; amphibians in purple, fishes in deep pink, mammals in gold, and reptiles/birds in olive green. The bar color after the tips indicates the gene number for each species; orange for *Per1*, dark green for *Per2*, and blue for *Per3*. **B**. Zoom in on amphibians, reptiles, and bird clades showing unusual gene loss patterns. **C**. Zoom in on *Salmonidae, Cyprinidae, and Chichilidae* clades that show unusual gene gain-loss patterns.

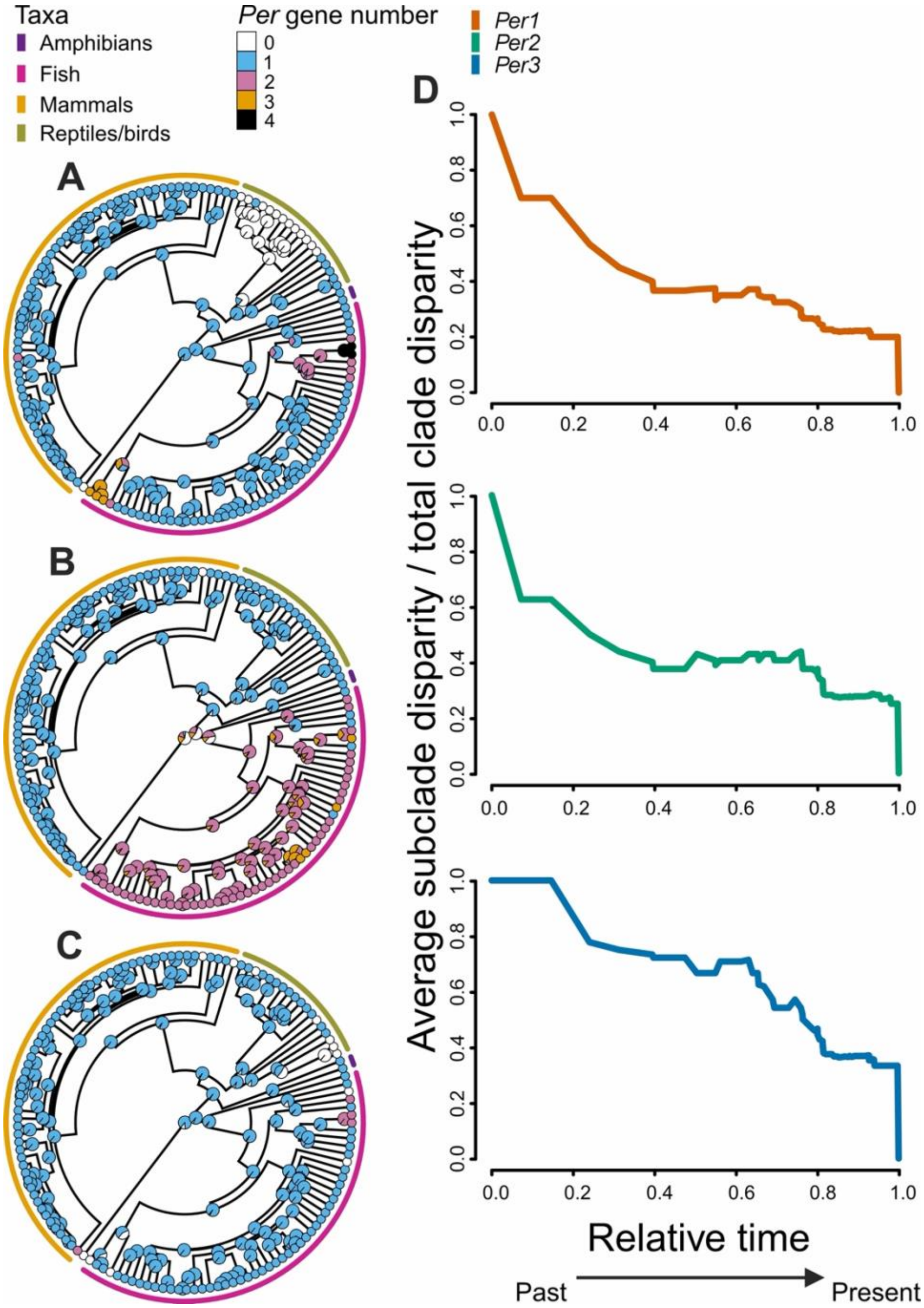


**Fig 2**. Ancestral state reconstruction for Period genes (*Per*) gain/loss numbers in vertebrates, **A**. *Per1*, **B**. *Per2*, and **C** *Per3*. **D**. Disparity Through Time (DTT) plots for each *Per* gene. The horizontal axis represents relative time values, 0.0 represents the root, and 1.0 the tips of the tree. The vertical axis represents the average subclade disparity divided by the total clade disparity and is calculated by each internal node of the tree. Higher values of disparity correspond to greater variance values within subclades relative to the disparity of the whole subclade.

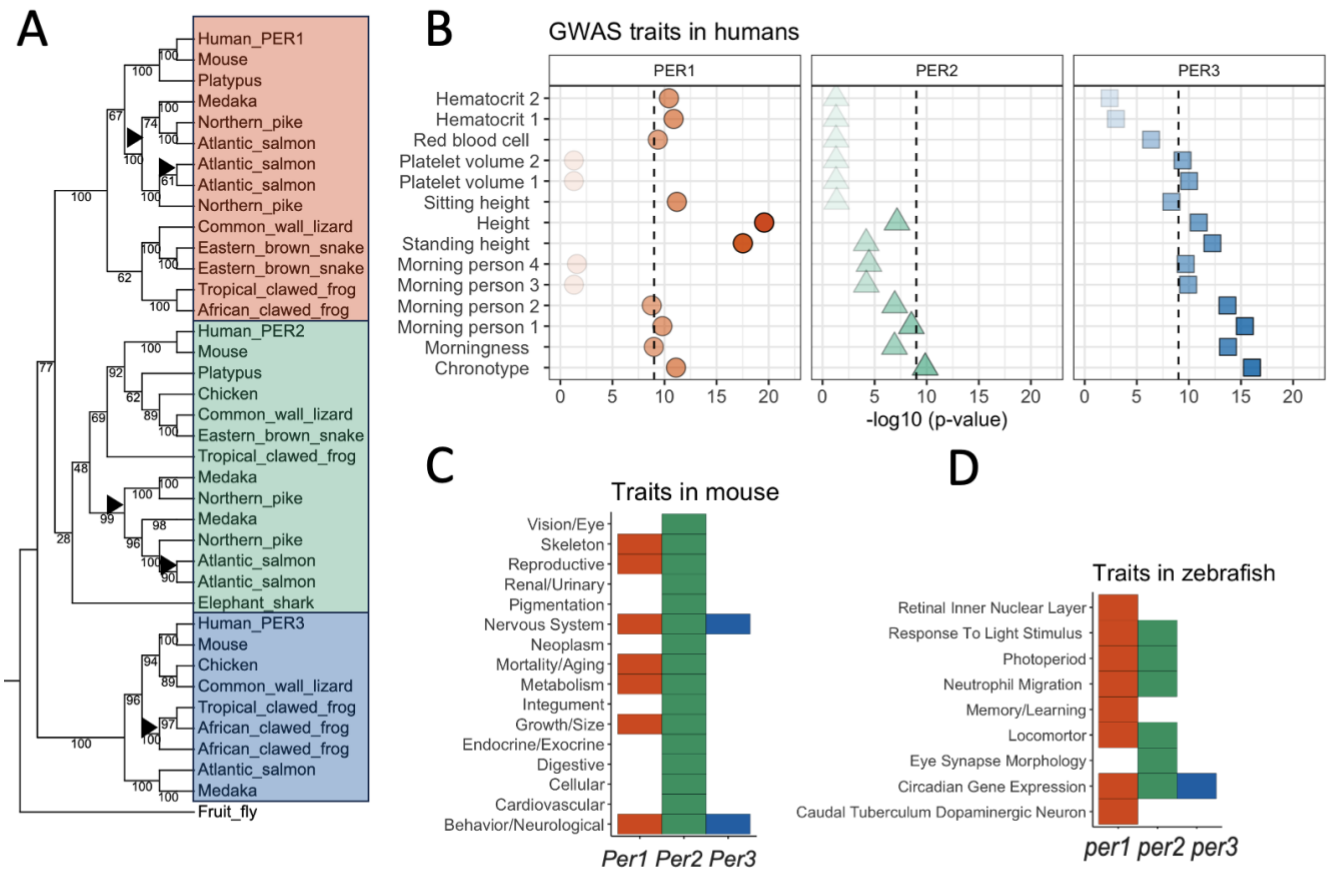


**Fig 3.** Phenotypic effect differentiation of *Period* genes (*Per*) in various vertebrate species. **A**. Amino-acid-based phylogenetic tree of vertebrates with the outlier fruit fly (*Drosophila melanogaster*) *per*. Amino acid alignments were performed using MAFFT v7 (Katoh et al. 2002). For the tree construction, the neighbor-joining method was applied (Saitou & Nei 1987), focusing on conserved sites across 363 amino acids. The nucleotide substitution model employed was the Jones-Taylor-Thornton model (Jones et al. 1992). Bootstrap analysis was conducted to assess the robustness of the tree, with the number of resampling events set to 100. Bootstrap value is shown in the tree (**file S3**). Black triangles indicate whole genome duplication events. The background color shows the gene classification (Red: *Per1*, Green: *Per2*, Blue: *Per3*). **B.** Reported phenotypes associated with *PER1*, *PER2*, and *PER3* in humans at GWAS Atlas (Watanabe et al. 2019). The -log$_{10}$ p-value of the association between genes and phenotypes is plotted on the x-axis. Only phenotypes with -log$_{10}$ p > 9 in at least one of the *PER* genes are shown. **C**. Reported traits in *Per* deficient mouse (*Mus musculus*) from Mouse Genome Database (MGD), 2021 (Blake et al.

2020). **D.** Reported traits in *per* deficient zebrafish (*Danio rerio*), the Zebrafish Information Network (zfin.org), 2022 (Bradford et al. 2022).

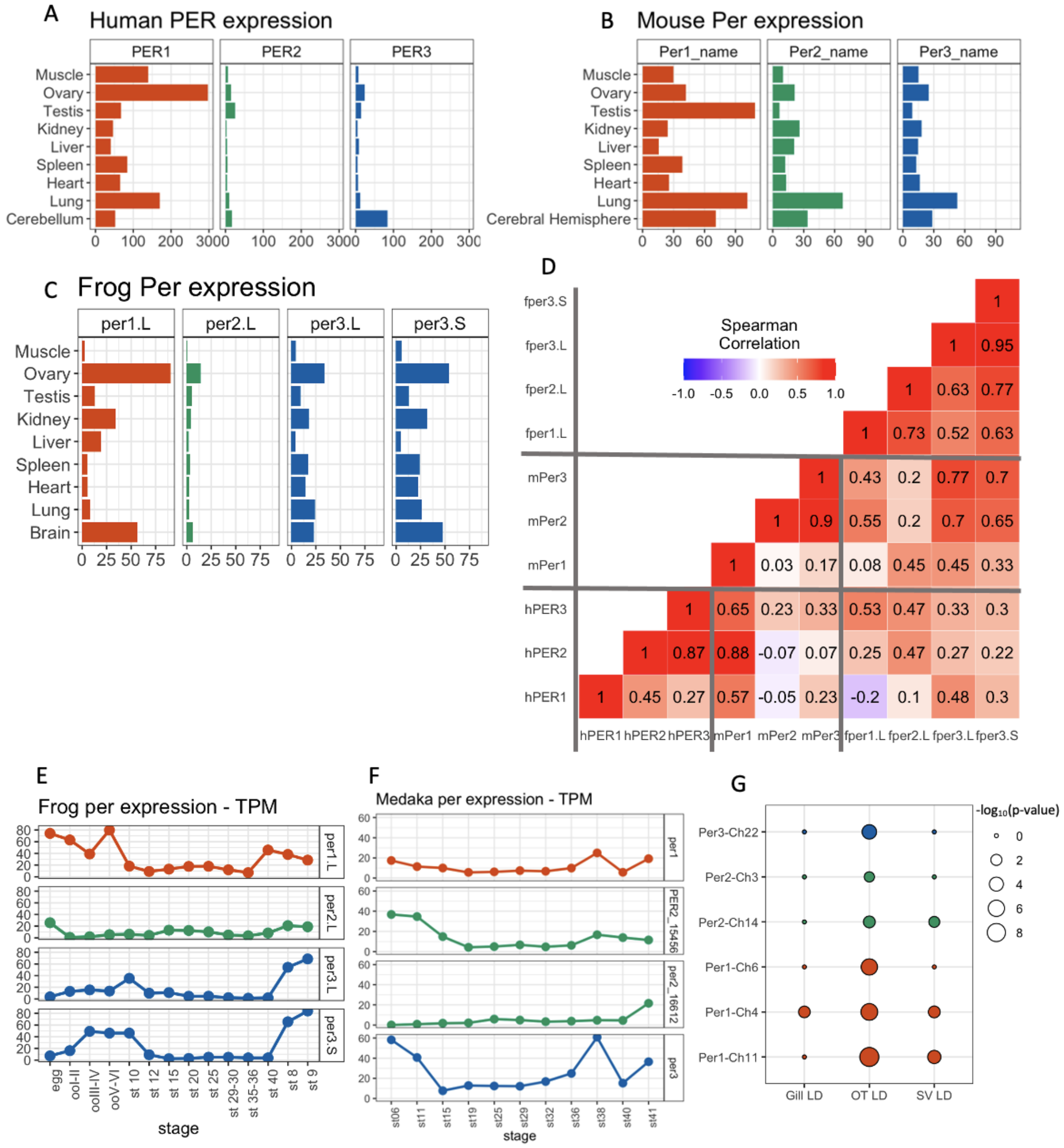


**Fig 4.** Transcriptomic differentiation of *Period* genes (*Per*) in **A**. Human, **B**, Mouse (*Mus musculus*), and **C** Frog (*Xenopus laevis*). The expression data was obtained from the Genotype-Tissue Expression

(GTEx) Portal, the Mouse Genome Database, and the Zebrafish Information Network. **D** Correlation of gene expression patterns between per genes (hPER: human, mPer: mouse, fper: frog). **E**. Developmental change of *per* gene expression in African clawed frog (*Xenopus laevis)* Xenbase (Fortriede et al. 2020) . **F**. Developmental change of *per* gene expression in medaka (*Oryzias latipes*), at Medaka Omics Data Portal (Li et al. 2020) **G**. Rhythmicity of *per* genes in tissues in Atlantic salmon (*Salmon salar*), *Optic tectum*, gill, and *saccus vasculosus*. Data is retrieved from supplementary table 8 of (West et al. 2020), one gene with no expression is not displayed.

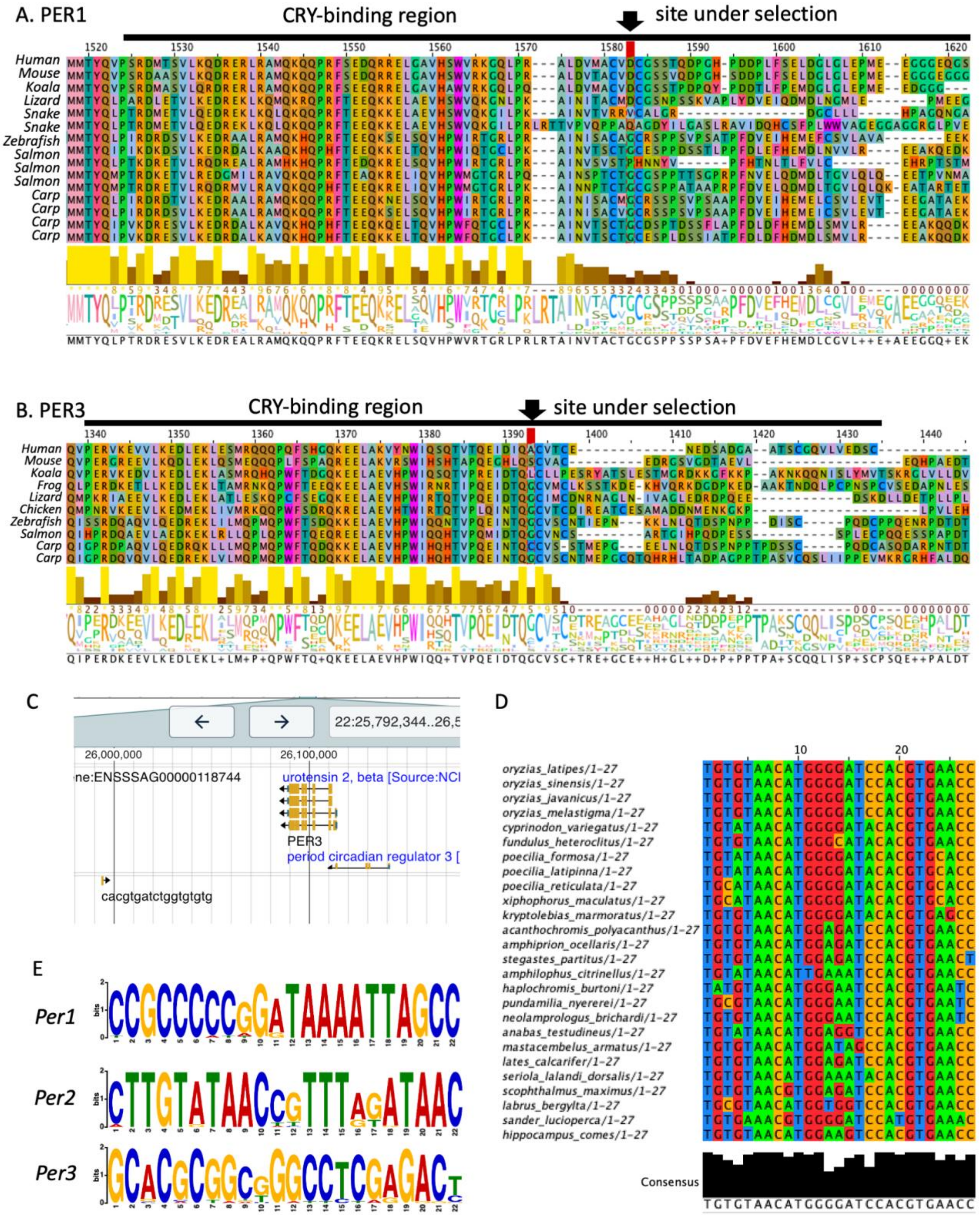


**Fig 5. AB**. The site with diversifying selection detected (arrow) by MEME (Mixed Effects Model of Evolution) (Murrell et al. 2012) on Datamonkey v2.0 (Weaver et al. 2018), which are both located in the

CRY_binding sites of *Per1* (**A**) and *Per3* (**B**). As a result, we observed signatures of diversifying selection in the CRY-binding sites in *Per1* (site "1583" in **Fig 5A**) and *Per3* (site "1393" in **Fig 5B** with the selected species). The sites in **Fig 5** and **Table S5** do not have the same numbers because only selected species are aligned and shown in **Fig 5** for a simplified visualization purpose, and **Table S5** is based on the aligned sequences of all the investigated species. **C**. A sequence, which is similar to the mammalian *PER1*-regulatory sequence, CACGTGatctggTGTGTG, near the *per3* gene in Atlantic salmon (*Salmo salar*) on Salmobase, (Ssal_v3.1, 22:25993822-25993839). **D**. Highly conserved *per2* regulatory region of Japanese medaka (*Oryzias latipes*) HdrR (ASM223467v1) 5:28606354-28606380 region in *Percomorpha.* **E.** MEME (Multiple Expectation maximizations for Motif Elicitation) *de novo* motif discovery results using the 1000 base upstream sequences of each *Per* gene on MEME Suites (Bailey et al. 2015). Putative motif sequence with the top bits score for each gene is displayed. *Per1*: CCGCCCCCGGATAAAATTAGCC (found in *Eutheria*), *Per2*: CTTGTATAACCGTTTAGATAAC (found in *Teleostei*), *Per3*: GGARGCCTAAATATAGGAGGCG (found in *Boreoeutheria*)

## Data Availability

All data used in this work are publicly available, and the accession information is described in the Materials and Methods section.

## Supplementary Materials

**Table S1. Master table of investigated species.**

**Table S2. Excluded species because of “N” in the *Period* genes.**

**Table S3. Variables in continental species.**

**Table S4. Variables in marine species.**

**Table S5. Coefficient of variation of gene numbers in gene families**

**Table S6. MEME selection detection results for *Period* genes.**

**Table S7. The length of each *Period* gene**

**Table S8. NCBI Conserved Domain search results**

**Table S9. MEME motif discovery results**

**Figure S1 RELAX result of salmonid *per3* gene.**

**Figure S2 The distribution of coefficient of gene copy number variation of each gene family across taxa**

**Figure S3 Brain *PER* gene expression in humans.**

**Figure S4 Diversifying selection signature in Period genes in vertebrates.**

**File S1. An R script for gene sequence curation and visualization and fetching upstream sequences**

**File S2. A Timetree-based species tree.**

**File S3. The phylogenetic tree of PER proteins.**

## References

Adamovich Y, Dandavate V, Asher G. 2022. Circadian clocks' interactions with oxygen sensing and signalling. Acta Physiol. (Oxf.). 234:e13770.

Archer SN, Schmidt C, Vandewalle G, Dijk D-J. 2018. Phenotyping of PER3 variants reveals widespread effects on circadian preference, sleep regulation, and health. Sleep Med. Rev. 40:109–126.

Bae K et al. 2001. Differential functions of mPer1, mPer2, and mPer3 in the SCN circadian clock. Neuron. 30:525–536.

Bailey TL et al. 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37:W202-8.

Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. 2:28–36.

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME suite. Nucleic Acids Res. 43:W39-49.

Baldarelli RM et al. 2021. The mouse Gene Expression Database (GXD): 2021 update. Nucleic Acids Res. 49:D924–D931.

Balsalobre A. 2002. Clock genes in mammalian peripheral tissues. Cell Tissue Res. 309:193–199.

Bargiello TA, Jackson FR, Young MW. 1984. Restoration of circadian behavioural rhythms by gene transfer in Drosophila. Nature. 312:752–754.

Basher Z, Bowden DA, Costello MJ. 2018. GMED: Global Marine Environment Datasets for environment visualisation and species distribution modelling. Earth Syst. Sci. Data Discuss. 1–62.

Baston D. exactextractr: Fast extraction from raster datasets using polygons. R package version 0.5. 0.

Beaulieu JM, O'Meara BC, Donoghue MJ. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. Syst. Biol. 62:725–737.

Bennie JJ, Duffy JP, Inger R, Gaston KJ. 2014. Biogeography of time partitioning in mammals. Proc. Natl. Acad. Sci. U. S. A. 111:13727–13732.

Berthelot C et al. 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat. Commun. 5:3657.

Blake JA et al. 2020. Mouse Genome Database (MGD): Knowledgebase for mouse–human comparative biology. Nucleic Acids Res. 49:D981–D987.

Bloch G, Barnes BM, Gerkema MP, Helm B. 2013. Animal activity around the clock with no overt circadian rhythms: patterns, mechanisms and adaptive value. Proc. Biol. Sci. 280:20130019.

Boettiger C, Lang DT, Wainwright PC. 2012. rfishbase: exploring, manipulating and visualizing FishBase data from R. J. Fish Biol. 81:2030–2039.

Bolton CM, Bekaert M, Eilertsen M, Helvik JV, Migaud H. 2021. Rhythmic Clock Gene Expression in Atlantic Salmon Parr Brain. Front. Physiol. 12:761109.

Bradford YM et al. 2022. Zebrafish information network, the knowledgebase for Danio rerio research. Genetics. 220. doi: 10.1093/genetics/iyac016.

Buels R et al. 2016. JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol. 17:66.

Byrska-Bishop M et al. 2022. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. Cell. 185:3426-3440.e19.

Canto CB, Onuki Y, Bruinsma B, van der Werf YD, De Zeeuw CI. 2017. The Sleeping Cerebellum. Trends Neurosci. 40:309–323.

Caravaggi A et al. 2018. Seasonal and predator-prey effects on circadian activity of free-ranging mammals revealed by camera traps. PeerJ. 6:e5827.

Cermakian N, Monaco L, Pando MP, Dierich A, Sassone-Corsi P. 2001. Altered behavioral rhythms and clock gene expression in mice with a targeted mutation in the Period1 gene. EMBO J. 20:3967–3974.

Cermakian N, Sassone-Corsi P. 2000. Multilevel regulation of the circadian clock. Nat. Rev. Mol. Cell Biol. 1:59–67.

Challet E. 2019. The circadian regulation of food intake. Nat. Rev. Endocrinol. 15:393–405.

Chiou Y-Y et al. 2016. Mammalian Period represses and de-represses transcription by displacing CLOCK–BMAL1 from promoters in a Cryptochrome-dependent manner. Proceedings of the National Academy of Sciences. 113:E6072–E6079.

Collins RL et al. 2020. A structural variation reference for medical and population genetics. Nature. 581:444–451.

Cox KH, Takahashi JS. 2019. Circadian clock genes and the transcriptional architecture of the clock mechanism. J. Mol. Endocrinol. 63:R93–R102.

Davesne D et al. 2021. Fossilized cell structures identify an ancient origin for the teleost whole-genome duplication. Proc. Natl. Acad. Sci. U. S. A. 118. doi: 10.1073/pnas.2101780118.

Delaunay F, Thisse C, Thisse B, Laudet V. 2003. Differential regulation of Period 2 and Period 3 expression during development of the zebrafish circadian clock. Gene Expr. Patterns. 3:319–324.

Deutsch EW et al. 2023. The ProteomeXchange consortium at 10 years: 2023 update. Nucleic Acids Res. 51:D1539–D1548.

Diesh C et al. 2022. JBrowse 2: A modular genome browser with views of synteny and structural variation. bioRxiv. 2022.07.28.501447. doi: 10.1101/2022.07.28.501447.

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1‐km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37:4302‐4315.

Fortriede JD et al. 2020. Xenbase: deep integration of GEO & SRA RNA-seq and ChIP-seq data in a model organism database. Nucleic Acids Res. 48:D776–D782.

Froese R, Pauly D, Others. 2010. FishBase. http://www.ices.dk/sites/pub/CM%20Doccuments/1992/L/1992_L10.pdf.

Fu L, Pelicano H, Liu J, Huang P, Lee C. 2002. The circadian gene Period2 plays an important role in tumor suppression and DNA damage response in vivo. Cell. 111:41–50.

Fukushima K, Pollock DD. 2020. Amalgamated cross-species transcriptomes reveal organ-specific propensity in gene expression evolution. Nat. Commun. 11:4459.

Gabriel CH et al. 2021. Live-cell imaging of circadian clock protein dynamics in CRISPR-generated knock-in cells. Nat. Commun. 12:3796.

Grabek KR, Chabot CC. 2012. Daily Rhythms of PERIOD protein in the eyestalk of the American lobster, Homarus americanus. Mar. Freshw. Behav. Physiol. 45:269–279.

GTEx Consortium et al. 2017. Genetic effects on gene expression across human tissues. Nature. 550:204–213.

Guharoy M, Chakrabarti P. 2010. Conserved residue clusters at protein-protein interfaces and their use in binding site identification. BMC Bioinformatics. 11:286.

Gundappa MK et al. 2022. Genome-Wide Reconstruction of Rediploidization Following Autopolyploidization across One Hundred Million Years of Salmonid Evolution. Mol. Biol. Evol. 39. doi: 10.1093/molbev/msab310.

Harmon LJ, Schulte JA, Larson A, Losos JB. 2003. Tempo and Mode of Evolutionary Radiation in Iguanian Lizards. Science. 301:961–964. doi: 10.1126/science.1084786.

Hedges SB, Marin J, Suleski M, Paymer M, Kumar S. 2015. Tree of life reveals clock-like speciation and diversification. Mol. Biol. Evol. 32:835–845.

Hewitt OH, Shaikh HM. 2021. The Rhythm of Many: Biological Rhythms in the Marine Environment, From Macro-Scale Planktonic Ecosystems to Micro-Scale Holobionts. Frontiers in Marine Science. 8. doi: 10.3389/fmars.2021.744169.

Hill MS, Vande Zande P, Wittkopp PJ. 2021. Molecular and evolutionary processes generating variation in gene expression. Nat. Rev. Genet. 22:203–215.

Horton TH. 2001. Conceptual Issues in the Ecology and Evolution of Circadian Rhythms. In: Circadian Clocks. Takahashi, JS, Turek, FW, & Moore, RY, editors. Springer US: Boston, MA pp. 45–57.

Hwang K, Bertolero MA, Liu WB, D’Esposito M. 2017. The Human Thalamus Is an Integrative Hub for Functional Brain Networks. J. Neurosci. 37:5594–5607.

Idda ML et al. 2012. Circadian clocks: lessons from fish. Prog. Brain Res. 199:41–57.

Inoue J, Sato Y, Sinclair R, Tsukamoto K, Nishida M. 2015. Rapid genome reshaping by multiple-gene loss after whole-genome duplication in teleost fish suggested by mathematical modeling. Proc. Natl. Acad. Sci. U. S. A. 112:14918–14923.

Jensen LD et al. 2012. Opposing effects of circadian clock genes bmal1 and period2 in regulation of VEGF-dependent angiogenesis in developing zebrafish. Cell Rep. 2:231–241.

Jetz W, McPherson JM, Guralnick RP. 2012. Integrating biodiversity distribution knowledge: toward a global map of life. Trends Ecol. Evol. 27:151–159.

Johnson MB et al. 2009. Functional and evolutionary insights into human brain development through global transcriptome analysis. Neuron. 62:494–509.

Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8:275–282.

Jones KE et al. 2009. PanTHERIA: a species‐level database of life history, ecology, and geography of extant and recently extinct mammals. Ecology. 90:2648–2648.

Junko I, Okamoto-Uchida Y, Nishimura A, Hirayama J. 2019. Light-dependent regulation of circadian clocks in vertebrates. In: Chronobiology [Working Title]. IntechOpen.

Kang HJ et al. 2011. Spatio-temporal transcriptome of the human brain. Nature. 478:483–489.

Katoh K, Misawa K, Kuma K-I, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.

Konopka RJ, Benzer S. 1971. Clock mutants of Drosophila melanogaster. Proc. Natl. Acad. Sci. U. S. A. 68:2112–2116.

Kosakovsky Pond SL et al. 2019. HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. Mol. Biol. Evol. 37:295–299.

Kumar Sudhir et al. 2022. TimeTree 5: An Expanded Resource for Species Divergence Times. Mol. Biol. Evol. 39. doi: 10.1093/molbev/msac174.

Kumar Vinod, Sharma A, Tripathi V. 2022. Physiological effects of food availability times in higher vertebrates. J. Exp. Biol. 225. doi: 10.1242/jeb.239004.

Lande-Diner L, Boyault C, Kim JY, Weitz CJ. 2013. A positive feedback loop links circadian clock factor CLOCK-BMAL1 to the basic transcriptional machinery. Proc. Natl. Acad. Sci. U. S. A. 110:16021–16026.

Landgraf D, Achten C, Dallmann F, Oster H. 2015. Embryonic development and maternal regulation of murine circadian clock function. Chronobiol. Int. 32:416–427.

Langmesser S, Tallone T, Bordon A, Rusconi S, Albrecht U. 2008. Interaction of circadian clock proteins PER2 and CRY with BMAL1 and CLOCK. BMC Mol. Biol. 9:41.

de Leeuw CA, Mooij JM, Heskes T, Posthuma D. 2015. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. 11:e1004219.

Li Y et al. 2020. Dynamic transcriptional and chromatin accessibility landscape of medaka embryogenesis. Genome Res. 30:924–937.

Lien S et al. 2016. The Atlantic salmon genome provides insights into rediploidization. Nature. 533:200–205.

Lin GG-H, Liou R-F, Lee H-J. 2002. The period gene of the German cockroach and its novel linking power between vertebrate and invertebrate. Chronobiol. Int. 19:1023–1040.

Mack KL et al. 2021. Repeated evolution of circadian clock dysregulation in cavefish populations. PLoS Genet. 17:e1009642.

Mai U, Mirarab S. 2018. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. BMC Genomics. 19:272.

Miyazaki K, Mesaki M, Ishida N. 2001. Nuclear entry mechanism of rat PER2 (rPER2): role of rPER2 in nuclear localization of CRY protein. Mol. Cell. Biol. 21:6651–6659.

Moreno JM, Jesus TF, Coelho MM, Sousa VC. 2021. Adaptation and convergence in circadian‐related genes in Iberian freshwater fish. BMC Ecology and Evolution. 21:1‐15.

Mundry R. 2014. Statistical Issues and Assumptions of Phylogenetic Generalized Least Squares. In: Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice. Garamszegi, LZ, editor. Springer Berlin Heidelberg: Berlin, Heidelberg pp. 131–153.

Murrell B et al. 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 8:e1002764.

Nagata K-I, Noda M, Iwamoto I, Tabata H, Ito H. 2019. Role of Per3, a circadian clock gene, in brain development. IBRO Reports. 6:S81. doi: 10.1016/j.ibror.2019.07.263.

Nakahata Y et al. 2008. A direct repeat of E-box-like elements is required for cell-autonomous circadian rhythm of clock genes. BMC Mol. Biol. 9:1.

Near TJ et al. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. Proc. Natl. Acad. Sci. U. S. A. 109:13698–13703.

Neme R, Tautz D. 2016. Fast turnover of genome transcription across evolutionary time exposes entire non-coding DNA to de novo gene emergence. Elife. 5:e09977.

Ohno S. 1970. *Evolution by Gene Duplication*. Springer Science & Business Media.

Ohno S. 1999. Gene duplication and the uniqueness of vertebrate genomes circa 1970-1999. Semin. Cell Dev. Biol. 10:517–522.

Oliveira BF, São-Pedro VA, Santos-Barrera G, Penone C, Costa GC. 2017. AmphiBIO, a global database for amphibian ecological traits. Sci Data. 4:170123.

Oliveira ML de et al. 2016. Phylogenetic signal in the circadian rhythm of morphologically convergent species of Neotropical deer. Mamm. Biol. 81:281–289.

Ono D et al. 2017. Dissociation of Per1 and Bmal1 circadian rhythms in the suprachiasmatic nucleus in parallel with behavioral outputs. Proc. Natl. Acad. Sci. U. S. A. 114:E3699–E3708.

Pagel M. 1997. Inferring evolutionary processes from phylogenies. Zool. Scr. 26:331–348.

Paquet ER, Rey G, Naef F. 2008. Modeling an evolutionary conserved circadian cis-element. PLoS Comput. Biol. 4:e38.

Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 35:526–528.

Patke A, Young MW, Axelrod S. 2020. Molecular mechanisms and physiological importance of circadian rhythms. Nat. Rev. Mol. Cell Biol. 21:67–84.

Pendergast JS, Friday RC. 2009. Endogenous rhythms in Period1 mutant suprachiasmatic nuclei in vitro do not represent circadian behavior. Journal of Neuroscience. https://www.jneurosci.org/content/29/46/14681.short.

Pennell MW et al. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics. 30:2216–2218. doi: 10.1093/bioinformatics/btu181.

Pereira DS et al. 2014. The effect of different photoperiods in circadian rhythms of per3 knockout mice. Biomed Res. Int. 2014:170795.

Ramanathan C et al. 2014. Cell type-specific functions of period genes revealed by novel adipocyte and hepatocyte circadian clock models. PLoS Genet. 10:e1004244.

Rasco D. 2020. *An R Companion for Applied Statistics I: Basic Bivariate Techniques*. SAGE Publications.

Relógio A et al. 2011. Tuning the mammalian circadian clock: robust synergy of two loops. PLoS Comput. Biol. 7:e1002309.

Reppert SM, Weaver DR. 2002. Coordination of circadian timing in mammals. Nature. 418:935–941.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution. 3:217–223. doi: 10.1111/j.2041-210x.2011.00169.x.

Rifkin SA, Houle D, Kim J, White KP. 2005. A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. Nature. 438:220–223.

Roenneberg T, Merrow M. 2016. The Circadian Clock and Human Health. Curr. Biol. 26:R432-43.

Running S, Mu Q, Zhao M. MOD16A3 MODIS/Terra Net Evapotranspiration Yearly L4 Global 500m SIN Grid V006. University of Montana and MODAPS SIPS–NASA.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. doi: 10.1093/oxfordjournals.molbev.a040454.

Samy JKA et al. 2017. SalmoBase: an integrated molecular data resource for Salmonid species. BMC Genomics. 18:482.

Sbrocco EJ, Barber PH. 2013. MARSPEC: ocean climate layers for marine spatial ecology. Ecology. 94:979–979.

von Schantz M, Jenkins A, Archer SN. 2006. Evolutionary history of the vertebrate period genes. J. Mol. Evol. 62:701–707.

Schultz TF, Kay SA. 2003. Circadian clocks in daily and seasonal control of development. Science. 301:326–328.

Seron-Ferre M, Valenzuela GJ, Torres-Farfan C. 2007. Circadian clocks during embryonic and fetal development. Birth Defects Res. C Embryo Today. 81:204–214.

Session AM et al. 2016. Genome evolution in the allotetraploid frog Xenopus laevis. Nature. 538:336–343.

Shimomura H et al. 2001. Differential daily expression of Per1 and Per2 mRNA in the suprachiasmatic nucleus of fetal and early postnatal mice. Eur. J. Neurosci. 13:687–693.

Siepka SM et al. 2007. Circadian mutant Overtime reveals F-box protein FBXL3 regulation of cryptochrome and period gene expression. Cell. 129:1011–1023.

Siwicki KK, Strack S, Rosbash M, Hall JC, Jacklet JW. 1989. An antibody to the Drosophila period protein recognizes circadian pacemaker neurons in Aplysia and Bulla. Neuron. 3:51–58.

Stanton D, Justin HS, Reitzel AM. 2022. Step in Time: Conservation of Circadian Clock Genes in Animal Evolution. Integr. Comp. Biol. 62:1503–1518.

Stone JR, Wray GA. 2001. Rapid evolution of cis-regulatory sequences via local point mutations. Mol. Biol. Evol. 18:1764–1770.

Symonds MRE, Blomberg SP. 2014. A Primer on Phylogenetic Generalised Least Squares. In: Modern

Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice. Garamszegi, LZ, editor. Springer Berlin Heidelberg: Berlin, Heidelberg pp. 105–130.

Takumi T et al. 1998. A new mammalian period gene predominantly expressed in the suprachiasmatic nucleus. Genes Cells. 3:167–176.

Tei H et al. 1997. Circadian oscillation of a mammalian homologue of the Drosophila period gene. Nature. 389:512–516.

The 1000 Genomes Project Consortium et al. 2015. A global reference for human genetic variation. Nature. 526:68.

Tobias JA et al. 2022. AVONET: morphological, ecological and geographical data for all birds. Ecol. Lett. 25:581–597.

Umemura Y, Yagita K. 2020. Development of the Circadian Core Machinery in Mammals. J. Mol. Biol. 432:3611–3617.

Vallone D, Gondi SB, Whitmore D, Foulkes NS. 2004. E-box function in a *period* gene repressed by light. Proc. Natl. Acad. Sci. U. S. A. 101:4106–4111.

Varcoe TJ et al. 2014. Rapidly alternating photoperiods disrupt central and peripheral rhythmicity and decrease plasma glucose, but do not affect glucose tolerance or insulin secretion in sheep. Exp. Physiol. 99:1214–1228.

Vera LM et al. 2023. Circadian rhythm of preferred temperature in fish: Behavioural thermoregulation linked to daily photocycles in zebrafish and Nile tilapia. J. Therm. Biol. 113:103544.

Wager-Smith K, Kay SA. 2000. Circadian rhythm genetics: from flies to mice to humans. Nat. Genet. 26:23–27.

Wang H. 2008. Comparative analysis of period genes in teleost fish genomes. J. Mol. Evol. 67:29–40.

Watanabe K et al. 2019. A global overview of pleiotropy and genetic architecture in complex traits. Nat. Genet. 51:1339–1348.

Weaver S et al. 2018. Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. Mol. Biol. Evol. 35:773–777.

Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. 2015. RELAX: detecting relaxed selection in a phylogenetic framework. Mol. Biol. Evol. 32:820–832.

West AC et al. 2020. Diversified regulation of circadian clock gene expression following whole genome duplication. PLoS Genet. 16:e1009097.

Xu P et al. 2019. The allotetraploid origin and asymmetrical genome evolution of the common carp Cyprinus carpio. Nat. Commun. 10:4625.

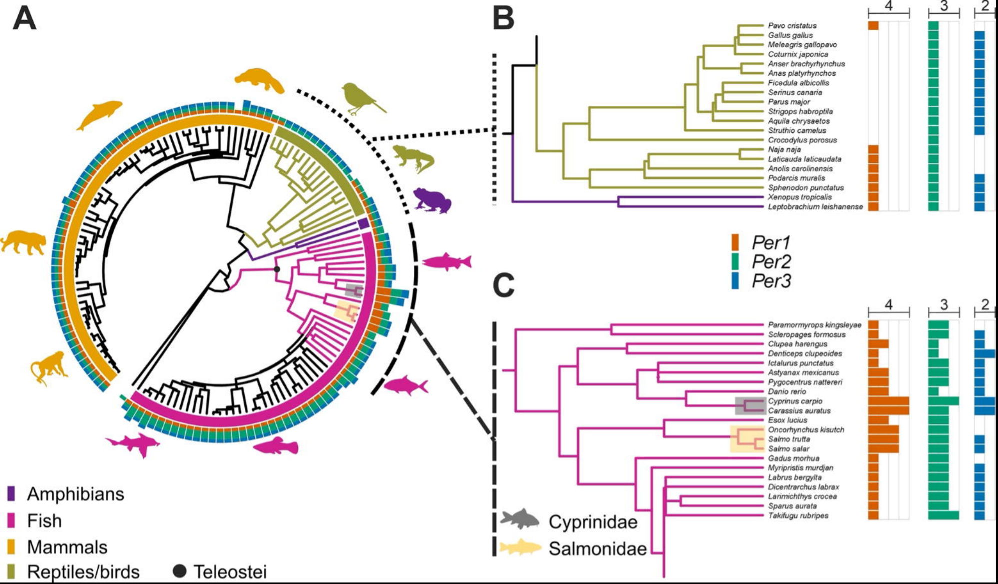
Yagita K et al. 2010. Development of the circadian oscillator during differentiation of mouse embryonic stem cells in vitro. Proc. Natl. Acad. Sci. U. S. A. 107:3846–3851.

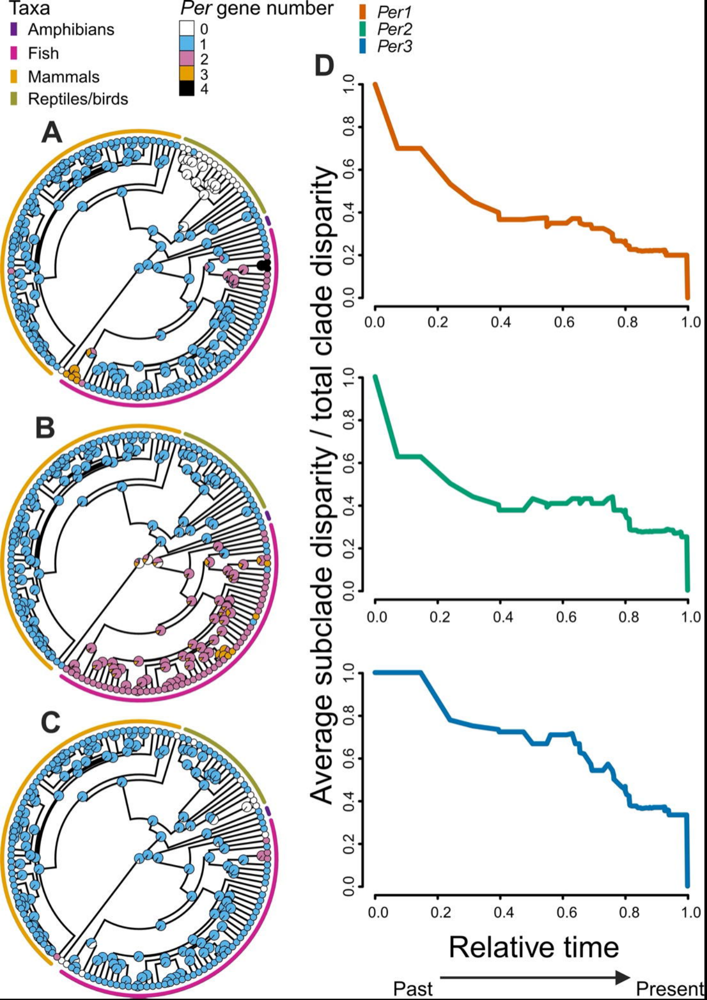
Zehring WA et al. 1984. P-element transformation with period locus DNA restores rhythmicity to mutant, arrhythmic Drosophila melanogaster. Cell. 39:369–376.

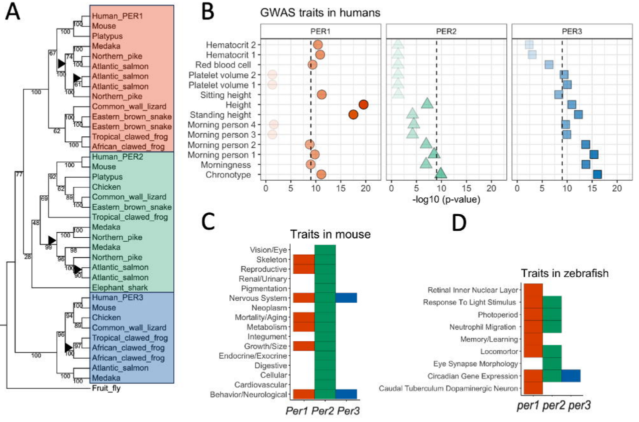
Zhang L et al. 2016. A *PERIOD3* variant causes a circadian phenotype and is associated with a seasonal mood trait. Proc. Natl. Acad. Sci. U. S. A. 113:E1536-44.

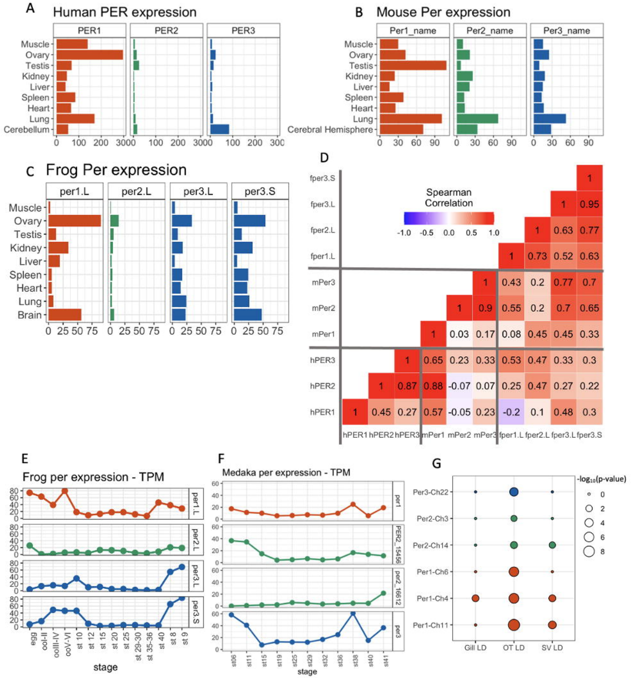
Zheng B et al. 2001. Nonredundant roles of the mPer1 and mPer2 genes in the mammalian circadian clock. Cell. 105:683–694.

Zheng X, Zhang K, Zhao Y, Fent K. 2021. Environmental chemicals affect circadian rhythms: An underexplored effect influencing health and fitness in animals and humans. Environ. Int. 149:106159.

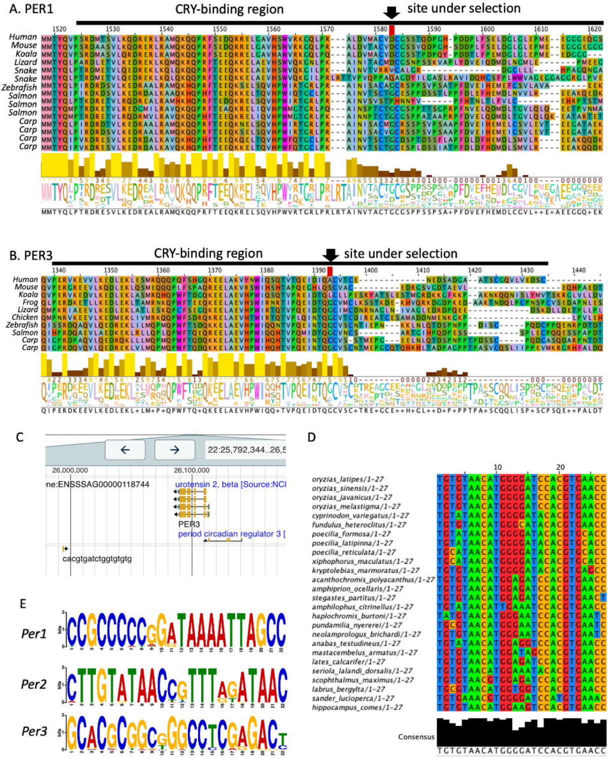

A
B
C
4
3
2
Pavo cristatus
Gallus gallus
Meleagris gallopavo
Coturnix japonica
Anser brachyrhynchus
Anas platyrhynchos
Ficedula albicollis
Serinus canaria
Parus major
Strigops habroptila
Aquila chrysaetos
Struthio camelus
Crocodylus porosus
Naja naja
Laticauda laticaudata
Anolis carolinensis
Podarcis muralis
Sphenodon punctatus
Xenopus tropicalis
Leptobrachium leishanense
Per1
Per2
Per3
Paramormyrops kingsleyae
Scleropages formosus
Clupea harengus
Denticeps clupeoides
Ictalurus punctatus
Astyanax mexicanus
Pygocentrus nattereri
Danio rerio
Cyprinus carpio
Carassius auratus
Esox lucius
Oncorhynchus kisutch
Salmo trutta
Salmo salar
Gadus morhua
Myripristis murdjan
Labrus bergylta
Dicentrarchus labrax
Larimichthys crocea
Sparus aurata
Takifugu rubripes
Cyprinidae
Salmonidae
Amphibians
Fish
Mammals
Reptiles/birds
Teleostei


Taxa
Amphibians
Fish
Mammals
Reptiles/birds
*Per* gene number
0
1
2
3
4
*Per1*
*Per2*
*Per3*
A
B
C
D
Average subclade disparity / total clade disparity
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Relative time
Past
Present


A
Human_PER1
Mouse
Platypus
Medaka
Northern_pike
Atlantic_salmon
Atlantic_salmon
Atlantic_salmon
Northern_pike
Common_wall_lizard
Eastern_brown_snake
Eastern_brown_snake
Tropical_clawed_frog
African_clawed_frog
Human_PER2
Mouse
Platypus
Chicken
Common_wall_lizard
Eastern_brown_snake
Tropical_clawed_frog
Medaka
Northern_pike
Medaka
Northern_pike
Atlantic_salmon
Atlantic_salmon
Elephant_shark
Human_PER3
Mouse
Chicken
Common_wall_lizard
Tropical_clawed_frog
African_clawed_frog
African_clawed_frog
Atlantic_salmon
Medaka
Fruit_fly
100
67
74
100
61
100
100
100
100
62
100
77
100
92
62
89
100
69
48
100
99
98
96
100
90
28
100
94
89
96
97
100
100
100
100
B
GWAS traits in humans
PER1
PER2
PER3
Hematocrit 2
Hematocrit 1
Red blood cell
Platelet volume 2
Platelet volume 1
Sitting height
Height
Standing height
Morning person 4
Morning person 3
Morning person 2
Morning person 1
Morningness
Chronotype
0
5
10
15
20
-log10 (p-value)
C
Traits in mouse
Vision/Eye
Skeleton
Reproductive
Renal/Urinary
Pigmentation
Nervous System
Neoplasm
Mortality/Aging
Metabolism
Integument
Growth/Size
Endocrine/Exocrine
Digestive
Cellular
Cardiovascular
Behavior/Neurological
Per1 Per2 Per3
D
Traits in zebrafish
Retinal Inner Nuclear Layer
Response To Light Stimulus
Photoperiod
Neutrophil Migration
Memory/Learning
Locomortor
Eye Synapse Morphology
Circadian Gene Expression
Caudal Tuberculum Dopaminergic Neuron
per1 per2 per3



| | hPER1 | hPER2 | hPER3 | mPer1 | mPer2 | mPer3 | fper1.L | fper2.L | fper3.L | fper3.S |
|---|---|---|---|---|---|---|---|---|---|---|
| fper3.S | | | | | | | | | | 1 |
| fper3.L | | | | | | | | | 1 | 0.95 |
| fper2.L | | | | | | | | 1 | 0.63 | 0.77 |
| fper1.L | | | | | | | 1 | 0.73 | 0.52 | 0.63 |
| mPer3 | | | | | | 1 | 0.43 | 0.2 | 0.77 | 0.7 |
| mPer2 | | | | | 1 | 0.9 | 0.55 | 0.2 | 0.7 | 0.65 |
| mPer1 | | | | 1 | 0.03 | 0.17 | 0.08 | 0.45 | 0.45 | 0.33 |
| hPER3 | | | 1 | 0.65 | 0.23 | 0.33 | 0.53 | 0.47 | 0.33 | 0.3 |
| hPER2 | | 1 | 0.87 | 0.88 | -0.07 | 0.07 | 0.25 | 0.47 | 0.27 | 0.22 |
| hPER1 | 1 | 0.45 | 0.27 | 0.57 | -0.05 | 0.23 | -0.2 | 0.1 | 0.48 | 0.3 |


A. PER1
CRY-binding region
site under selection
Human
Mouse
Koala
Lizard
Snake
Snake
Zebrafish
Salmon
Salmon
Salmon
Carp
Carp
Carp
Carp
B. PER3
CRY-binding region
site under selection
Human
Mouse
Koala
Frog
Lizard
Chicken
Zebrafish
Salmon
Carp
Carp
C
22:25,792,344..26,5
26,000,000
26,100,000
ne:ENSSSAG00000118744
urotensin 2, beta [Source:NCI
PER3
period circadian regulator 3 [
cacgtgatctggtgtgtg
D
oryzias_latipes/1-27
oryzias_sinensis/1-27
oryzias_javanicus/1-27
oryzias_melastigma/1-27
cyprinodon_variegatus/1-27
fundulus_heteroclitus/1-27
poecilia_formosa/1-27
poecilia_latipinna/1-27
poecilia_reticulata/1-27
xiphophorus_maculatus/1-27
kryptolebias_marmoratus/1-27
acanthochromis_polyacanthus/1-27
amphiprion_ocellaris/1-27
stegastes_partitus/1-27
amphilophus_citrinellus/1-27
haplochromis_burtoni/1-27
pundamilia_nyererei/1-27
neolamprologus_brichardi/1-27
anabas_testudineus/1-27
mastacembelus_armatus/1-27
lates_calcarifer/1-27
seriola_lalandi_dorsalis/1-27
scophthalmus_maximus/1-27
labrus_bergylta/1-27
sander_lucioperca/1-27
hippocampus_comes/1-27
Consensus
TGTGTAACATGGGGATCCACGTGAACC
E
Per1
Per2
Per3
bits