Functional and regulatory diversification of circadian rhythm *period* genes during the evolution of 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**

Affiliations

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

[2] Departamento de Zoología, Instituto de Biología, Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico

Corresponding Author: marie.saitou@nmbu.no

Abstract:

The Period genes (*Per*) play essential roles in modulating the molecular circadian clock, which regulates the physiological and cellular timing in a broad range of species through the transcription-translation feedback loop in organisms. While the period gene paralogs are widely observed among vertebrates, the evolutionary history and the functional diversification of *Per* gene paralogs across vertebrates are not well known. In this study, we comprehensively investigated the evolution of *Per* genes, their lineage-specific transcriptome landscape and phenotypic effects, and the rapid turnover of transcription binding sites by comparative genomics and transcriptomics. We observed multiple lineage-specific gain and loss events of *Per* genes, though there was no simple association observed between ecological factors and *Per* gene numbers in each species. Most salmonid species lost the *per3* gene, while the species that retained the *per3* gene in salmonid did not show the signature of relaxed selective constraint, suggesting functional role of *per3* gene in retained species. We also determined the signature of adaptive diversification of the CRY-binding region in *Per1* and *Per3*, which modulates the circadian rhythm. Collectively, our findings revealed the evolution of *Per* genes and their fine-tuned contribution to the plastic regulation of circadian rhythms in various vertebrate taxa.

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

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 in different cells and organs [1–4]. The system is essential for an organism's rhythmic behavior and phisiological responses, including sleep/wake cycle and foraging, so that it can efficiently interact with the dynamic and temporal external resources and environments by adjusting its physiological processes [5,6]. 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 [7–9]. Mammalian homologs of the period gene were also characterized soon after that [10], and the genetics, molecular mechanism, and behavioral biology of period genes have been extensively studied in model organisms [11–13]. Such studies revealed that in at least fish and mammals, the molecular circadian clock is regulated by transcription-translation feedback loops, core clock genes with changes in external cues, in which Period genes (*Per*) play the key role [14–19]. 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 [20]. 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 [5,21].

Humans and mice have three copies of *PER* genes, *PER1*, *PER2*, and *PER3*, likely due to the two rounds of ancient genome duplications [22], which is one case of the “evolution by genome duplication” theory [23]. In fish species, some known gene absence/presence of *per* genes include a lack of *per3* in sticklebacks (Gasterosteidae) are observed, and multiple additional duplicates and lineage-specific evolution of *per* genes in several fish species [24–26]. *Per* paralogues show some sequence similarity and some 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 [27]. On the contrary, a genetic variant in *PER3* in humans is associated with sleep disorders [28]. In non-model species, for example, the evolutionary loss of circadian rhythms in cavefish (*Astyanax* sp.) is reported, suggesting the impact of ecological factors on circadian gene evolution [16,26]. However, the

comprehensive characterization of *Per* genes is yet to be conducted in most non-model vertebrates despite the recent availability of high-quality genome data, and thus their evolutionary history and diversification process are unclear.

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 [29,30]. Contrary to the historical notion, studies revealed that *Per* genes are also expressed in cell lines and non-brain tissues [31], sometimes with circadian rhythmic expression patterns, such as in mouse liver [32] and cell lines [33], and human cell lines [34]. 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 [5,6]. Interestingly, studies reported that circadian rhythm starts functioning even before birth in some species [35]. For example, differential regulation of *per2* and *per3* expression during zebrafish embryogenesis is reported [36]. 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 [37][38], 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 [35,39]. In addition, *Per* genes have non-circadian functions, such as DNA damage response [40] and embryonic neuron development [41]. Although there are numerous reports about the *Per* expression in various conditions in model organisms, *Per* gene paralogs are often not investigated simultaneously, and the comprehensive dynamics of how *PER* gene paralogs share roles and how conserved these roles are across species is not known. Besides, in earlier studies, before genome information was available, *Per* paralogs were sometimes not correctly annotated. Building on that, this paper aims to comprehensively investigate the evolutionary history and the functional diversification of *Per* genes in vertebrates.

Results

Lineage-specific gene gain and loss of period genes in vertebrates

We first curated the *Per* genes in vertebrates (**Fig 1A**) from Ensembl Compara (version 106) and constructed a species-based phylogenetic tree to visualize the evolutionary history of *Per* gene gain and loss. 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. The phylogenetic signal (Pagel's lambda λ) [42] 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* showed the strongest phylogenetic signal, and the remaining genes also showed high phylogenetic signals: *Per1* λ= 1.00, *Per2* λ= 0.85, and *Per3* λ= 0.85. 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, suggesting that *Per2* plays a major, universal role. Elephant sharks (*Callorhinchus milii*) have only one *per2* gene, and in most teleost fish species, two *per2* genes were observed, likely due to the teleost fish-specific whole genome duplication event (Ts3R) [43–45]. In addition, there was a Cyprinidae-specific whole genome duplication (Cs4R) [46] and a salmonid-specific whole genome duplication (Ss4R) [47], which led to the additional duplications of *per* gene in each lineage **(Fig 1C**). Conversely, *per3* gene loss is observed in multiple distant fish lineages, including pike (*Esox lucius*), most salmonid species [48], cod (*Gadus morhua*), stickleback (*Gasterosteus aculeatus*), and lumpfish (*Cyclopterus lumpus*) (**Fig 1C**). Although the loss of another circadian gene and retrogression of circadian rhythm is reported in cavefish (*Astyanax mexicanus*) [26], 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)*, brown trout *(Salmo trutta)*, and Danube salmon *(Hucho hucho)*, consistent with a previous study [48]. To test if *per3* gene is under relaxed selection in these three species, we used a codon-based statistical test to detect relaxed selection, RELAX [49], using the *per3* genes in these three species with retained *per3*, compared to the

Acanthomorphata group. As a result, we did not find evidence for relaxed or intensified selection in the *per3* genes in these species with retained *per3*, compared to the control group (k=1.22, p=0.072) (**Supplementary Figure S1**). This implies that the *per3* gene maintains indispensable roles in salmonid species which have this gene.

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) 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 factors on each species is complex, mutually influential, and variable [50–53]. We hypothesized that some environmental factors may have driven the evolution of *Per* gene gain and loss. To investigate the evolutionary association between each *Per* gene number and in each species, we conducted the Phylogenetic Generalized Least Squares test [54] between per gene copy number and environmental/ecological factors of 138 species 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/ecological variable

(Phylogenetic Generalized Least Squares test, *p*-value less than 0.05 was considered to be significant). We assume that ecological adaptation on circadian rhythm is based upon a more complex system, such as spatiotemporal regulation of the enzymatic activity, rather than the simple *Per* gene copy number.

Functional diversification of Period genes and their taxonomy-specificity

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 [27], and *Per3* is often excluded from later investigations and discussions probably for this reason [20,55]. 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 is 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 [24]. 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 functional difference 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 [56]. The gene-level p-value of association with traits was computed by MAGMA [57], 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 (**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. There is a deletion (54 bases) polymorphism reported in the GNOMAD paper [58], which is associated with sleep [59]. This variant is not reported in the 1000 Genomes phase 3 dataset [60] and the updated 1000 genome dataset [61] 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 [12,27,62]. Interestingly, *Per3* knockout mice have relatively moderate changes in circadian rhythm, observed light-dependent phenotypes, and reduced body size, growth, and metabolism [59,63]. In our meta-analysis, using data from Mouse Genome Database (MGD), 2021 [64], *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 [65], 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 Per 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 [66], mouse (*Mus musculus*) expression from the mouse Gene Expression Database (GXD) [67], and African clawed frog (*Xenopus laevis)* Xenbase [68] **(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 [69] 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 [70], 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 [71], 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 [70] and the Japanese medaka (*Oryzias latipes*) which has likely duplicated the *per2* gene because of the teleost genome duplication around 310 Mya [72]. 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 [68] (**Fig 4F**), 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 (**Fig 4E**). 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 [73,74]. 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/) [75,76] **(Fig S2)**. 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 [77]. Neither the suprachiasmatic nucleus nor the broader region, the hypothalamus, which is responsible for controlling circadian rhythms [29,30], 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. [78] 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. [78] (**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) [79] on Datamonkey version 2.0 [80] (**Table S5**). 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* and *Per3*, as shown in **Fig 5A and 5B**, **Fig S3 ,** contrary to the general notion that protein binding sites in a gene are evolutionarily conserved [81].

As summarized in the introduction section, the PER and Cryptochrome CRY proteins orchestrate the circadian feedback loop with the CLOCK–BMAL1 [21][5]. CRY-binding region of the *Per* gene was

characterized in rats (*Rattus norvegicus*) [82]. 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 at site 343 in the alignment (**Fig S4),** while we did not find functional annotation of this region in (corresponding with the 93rd "S" in NP**S**TSGCSSD in the human sequence, O15055 · PER2_HUMAN at UNIProt, https://www.uniprot.org/uniprotkb/O15055/entry, last accessed, 23rd. January 2023,[83]).

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. [84] 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. [85] 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 [85]. 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 [86,87] -based SalmoBase [88] (https://salmobase.org/, Last accessed, Jan 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. [84], 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 5D**). 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. [85] estimated is corresponded to Japanese medaka HdrR (ASM223467v1) 5:28606354-28606380 region, near the *per3* gene, while it to located near the *Per2* sequence in mammalian species [85]. In Ensembl 65 fish EPO-Extended, this region is widely observed in *Percomorpha* (**Fig 5D**) as represented in Paquet et al. [85]. 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 [89,90].

Discussion

One of the major questions in evolutionary genomics is how duplicated genes develop the diverse phenotype in various species by fine-tuning their original functions. In this study, we focused on the highly conserved *Per* gene family in vertebrates, which plays a key role in circadian rhythm. We revealed how the duplicated *Per* gene copies modified their function between species and paralogs. While *Per1*, *Per2*, and *Per3* paralogous genes were duplicated in the common ancestor of vertebrates, there have been multiple lineage-specific gain and loss events. Most birds lost the *Per1* gene, while most reptiles lost the *Per3* gene. The *Per2* gene was observed in most species. Salmonid species lost the *per3* gene except for *H. hucho, S. trutta*, and *S. salar* salmonids. For the three species with the *per3* gene, we detected no signature of relaxed selection compared to the Acanthomorphata group. This is contrary to the current view that *Per3* is the least important, based on earlier knock-out mouse studies, and thus, least studied among the paralogs [27]. By comparative analysis of public gene, phenotype, and transcriptome datasets, we determined that the phenotypic effect of *Per* gene paralogs and their expression patterns are highly variable between species and paralogues, which indicates the rapid evolution of sub-

functionalization. We also determined signatures of diversifying selection in the *Per1* and *Per3* genes' CRY-binding region and a rapid turnover of binding site sequences. However, one binding site sequence was widely maintained among taxa (**Fig 3E**), suggesting the lineage-specific conserved role of this region. 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.

We postulate that as more transcriptome/gene-phenotype datasets of a variety of species become available, not only genome datasets, the evolution of gene function and regulation will be more comprehensive in the context of how genome evolution contributes to cross-species phenotypic diversity. 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 a 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 the *Per* gene duplicates 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 genomics, especially evolutionary genomics, is to connect insights from genome information to their functional roles on molecular and macroscopic phenotypes and environmental conditions.

Methods

Gene Sequences

We have retrieved all available *Per1*, *Per2*, and *Per3* sequences from Ensembl Compara, version 106, 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 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, 15. Aug. 2022) and two species with projection-built genomes.

(https://www.ensembl.org/info/about/species.html). Lastly, we removed six species that are not available at TimeTree5 [91] and four non-vertebrate species, which additionally four do not have geographic range data on Map of Life/IUCN (last accessed, 15. Aug. 2022). A total of 160 vertebrate species were used for subsequent analyses, species lists and *Per* gene numbers are available in **Table S1**. Amino acid alignment was conducted by Mafft version 7 [92].

Evolutionary history of gene gain and loss of period genes in vertebrates

We used the phylogenetic hypothesis from TimeTree5 [91] and was pruned for the species of interest using ape v5.6 package [93] to reconstruct the evolution of each *Per* gene in vertebrates. We reconstruct ancestral *Per* states with the equal rates (ER) model, assuming that evolutionary transitions between different gene numbers have the same rate, using a maximum likelihood approach as implemented in the *ace* function in ape package and plotted into the tree using phytools v1.2.0 [94] in R v4.2.2 (https://www.R-project.org/). 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 for 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) [95], we performed this analysis in geiger v2.0.10 package [96].

Association between environmental and ecological factors and period gene numbers

Selection of environmental variables

We chose a set of terrestrial and marine world easter layers as likely candidates to predict the evolution of *Per* genes. For terrestrial environments, six climatic layers at 10 arc min (~340 km²) were obtained from Worldclim 2.1 [97]: elevation, 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 v.55 layers at 0.1 degrees downloaded from NASA for 2000-2016 [98] mean value was calculated for each year, and finally, the mean value for the whole period. For marine environments, we selected five available layers analogous to the terrestrial, these were: bathymetry and annual temperature range downloaded from MARSPEC at 10 arc min [99], annual photosynthetic available radiation and annual primary productivity downloaded from Bio-ORACLE 2.0 at 5 arc min [100], and annual wind speed downloaded at 5 arc min from Global Marine Environment Datasets 2.0 [101].

All raster layers were scaled to 10 arc min. A total of 138 available geographic range species maps were obtained from the International Union for Conservation of Nature (IUCN. 2022. The IUCN Red List of Threatened Species. Version 2022-2. https://www.iucnredlist.org. Accessed on October 2022) and Map of Life [102] for extracting raster values and mean values calculated for each species using raster v3.6 and exactextractr v0.9.1 [103] in R.

Selection of ecological variables.

Ecological traits available for main taxonomic groups were obtained from public databases. For fishes, migration, body mass, vision, and swimming mode were downloaded from [104] through rfishbase v4.0.0 R package [105]. For amphibians, body mass, habitat, and migration were downloaded from AmphiBio

[106]. For birds, body mass, habitat, and migration were downloaded from AVONET [107]. For mammals, activity cycle, body mass, and habitat breadth were downloaded from PanTHERIA [108].

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; [54]. 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 [109]. 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 [110] before running all models. Variables with no normality and a VIF major of 5 were discarded from the final model.

The phylogenetic hypothesis was obtained from TimeTree5 and pruned for species with environmental/ecological available data using ape v5.6 package. PGLS models were computed for each *Per* gene with the final predictors and using a phylogenetic signal calculated using the Pagel’s lambda (λ) estimated by maximum likelihood [42] using caper package v1.0.1 (<https://CRAN.R-project.org/package=caper>). Diagnostic plots of residuals (e.g., outlier analysis, Q-Q plots) were checked to make sure that assumptions were not violated. Species and *Per*/environmental/ecological data analyzed are in **Table S3 and S3**.

Gene-phenotype association datasets

Reported phenotypes associated with *PER1*, *PER2,* and *PER3* in humans at GWAS Atlas [56] 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 Mouse Genome Database (MGD), 2021 [64]. Traits in the *per* deficient zebrafish were retrieved from the Zebrafish Information Network (zfin.org), 2022 [65].

Transcriptome datasets

Human adult transcriptome data set was retrieved from GTEx portal version 8 (last accessed: Jan. 18th, 2023) [66]. Mouse transcriptome dataset was retrieved from The mouse Gene Expression Database (GXD) [67] 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 [68]. 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 cerreblum). We used Spearman's rank correlation coefficient test to investigate the correlation between gene expression patterns between paralogs.

Molecular evolution analysis

Intensification or relaxation of selection was estimated using [49] implemented in [111]. 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, Danube salmon) *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) [79] on Datamonkey version 2.0 [80]. We conducted the analysis for 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 [86,87] -based SalmoBase [88] (https://salmobase.org/, Last accessed, Jan 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. [84], 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 [85]:

Medaka cgtTCACGTgga-tccccatGTTACA

Zebratish cggTCACCTgtt-tctccacATGCTG

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

Acknowledgments

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

Funding:

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

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

Figures and Tables

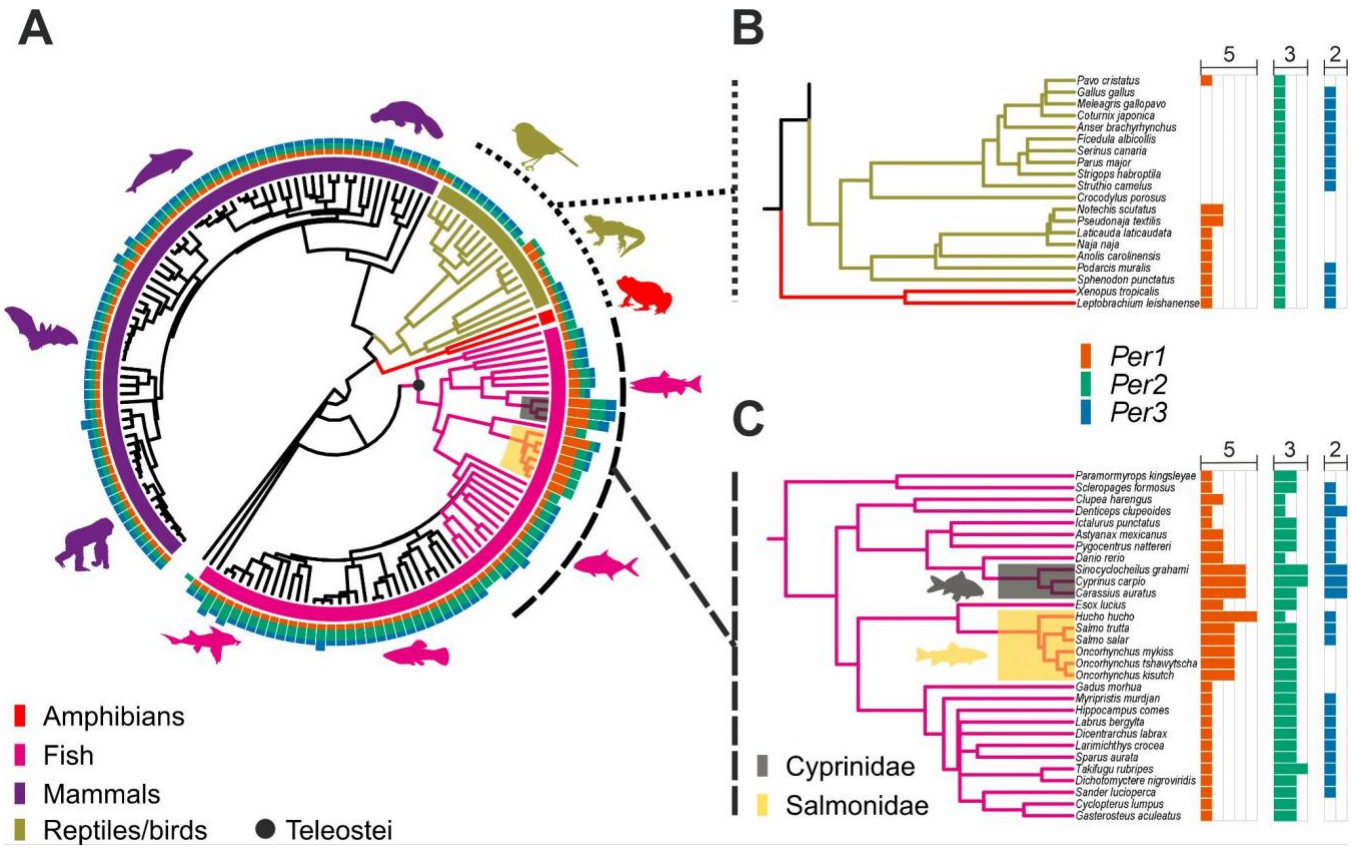

Fig 1. The evolutionary history of the period gene (*Per*) gain and loss in vertebrates. **A**. Phylogenetic tree and numbers of *Per* paralogs in vertebrates. Bar colors around the tree denote the main vertebrate taxonomic clades; amphibians in red, fishes in deep pink, mammals in purple, and reptiles/birds 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.

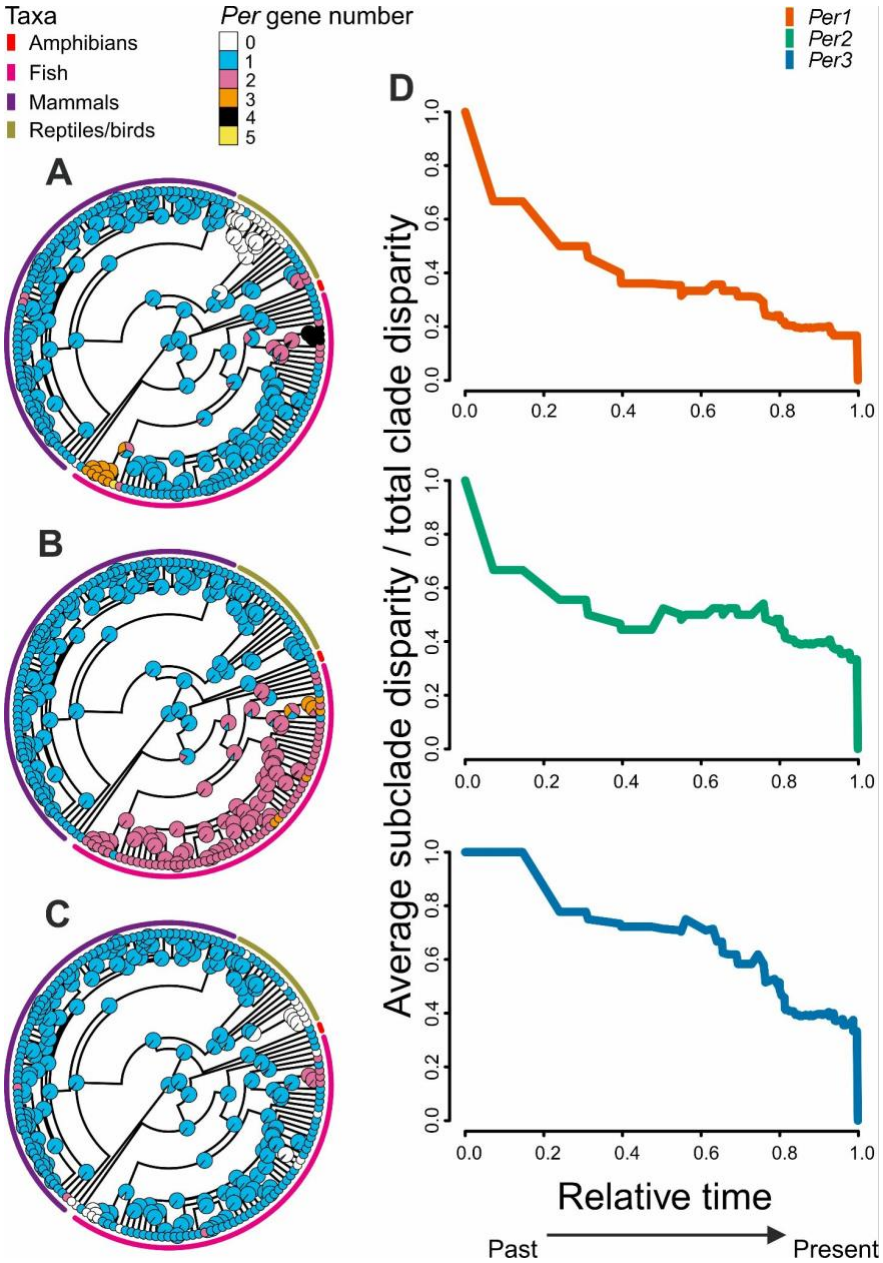

Figure 2. Ancestral state reconstruction using equal rates model for Period genes (*Per*) gain/loss numbers in vertebrates, **A**. *Per1*, **B**. *Per2*, and **C** *Per3*. **D**. Disparity Through Time (DTT) plots for each *Per* gene. The horizontal axis represents relative time values, 0.0 represents 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.

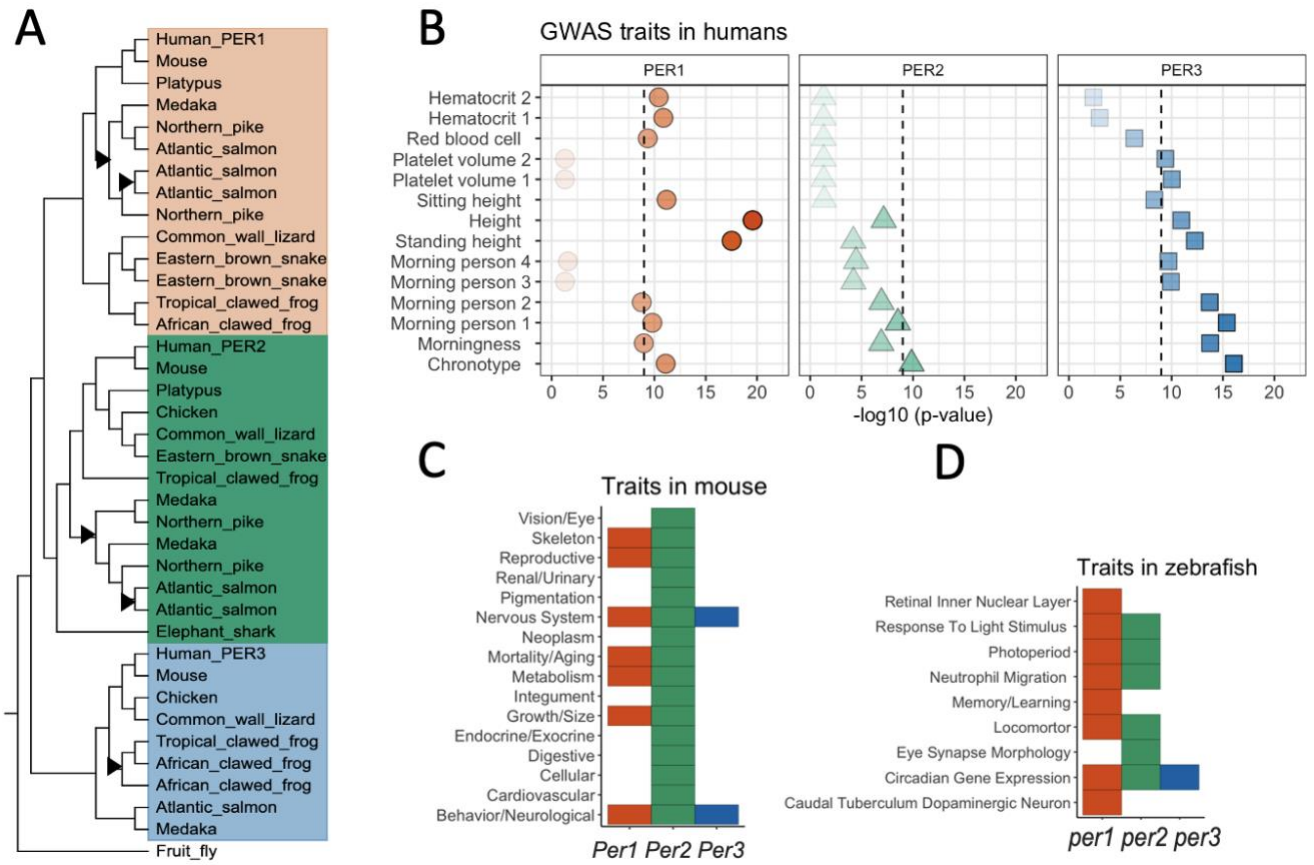

Fig 3. Phenotypic effect differentiation of period genes (*Per*) in various vertebrate species. **A**. Amino-acid-based phylogenetic tree of vertebrates with the outlier fruit fly (*Drosophila melanogaster*) *per*. Black triangles indicate whole genome duplication events. The background color shows the gene classification (Red: *Per1*, Green: *Per2*, Blue: *Per3*). **B.** Reported phenotypes associated with *PER1*, *PER2*, and *PER3* in humans at GWAS Atlas [56]. 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 [64]. **D.** Reported traits in *per* deficient zebrafish (*Danio rerio*), the Zebrafish Information Network (zfin.org), 2022 [65].

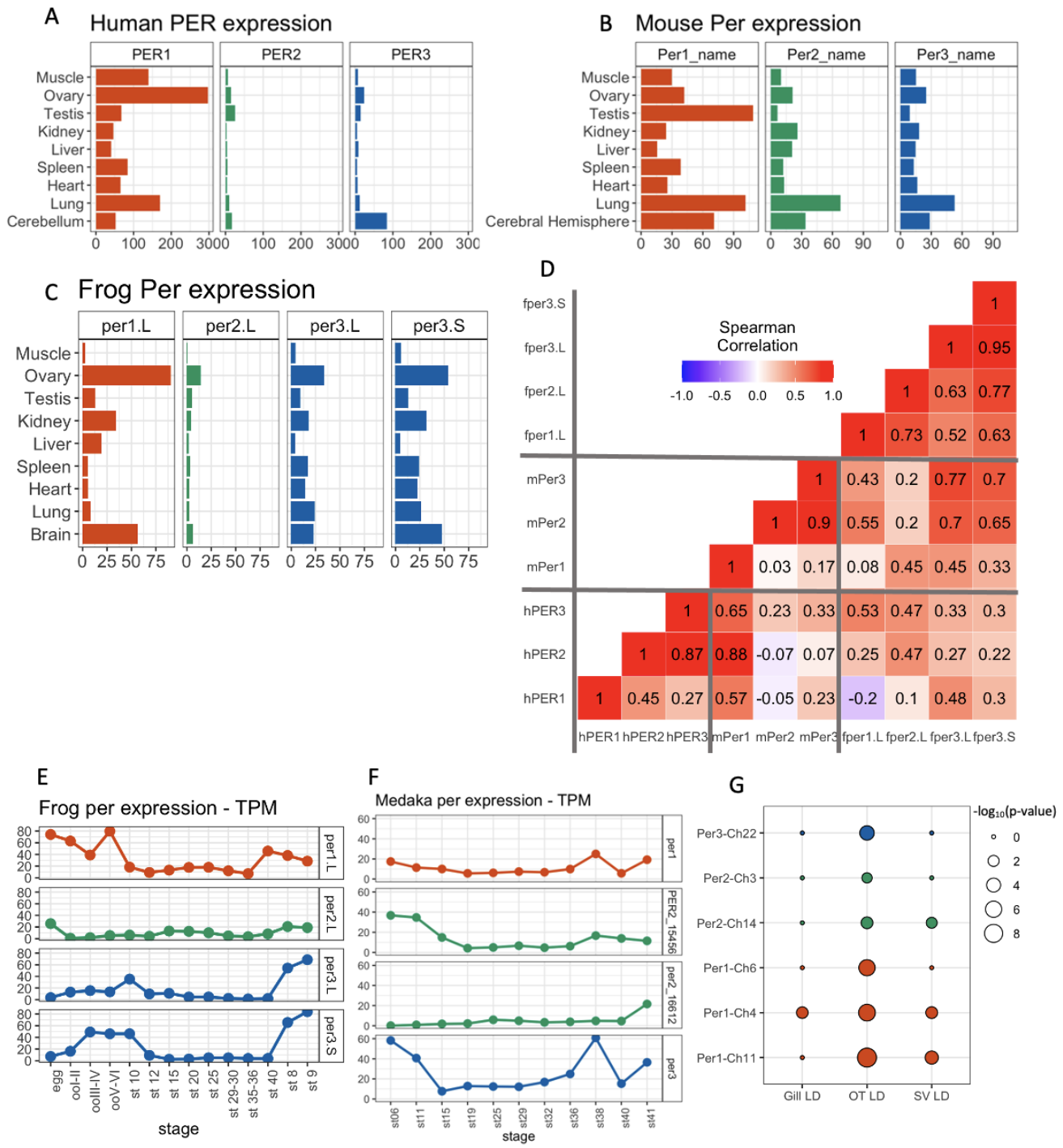

Fig 4. Transcriptomic differentiation of period genes (*Per*) in **A**. Human, **B**, Mouse (*Mus musculus*), and **C** Frog (*Xenopus laevis*). The expression data was obtained from GTEx, Mouse Genome Database, and the Zebrafish Information Network. **D** Correlation of gene expression patterns between per genes (hPER: human, mPer: mouse, fper: frog). **E**. Medaka development. **F**. Frog development, **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 [78], one gene with no expression is not displayed.

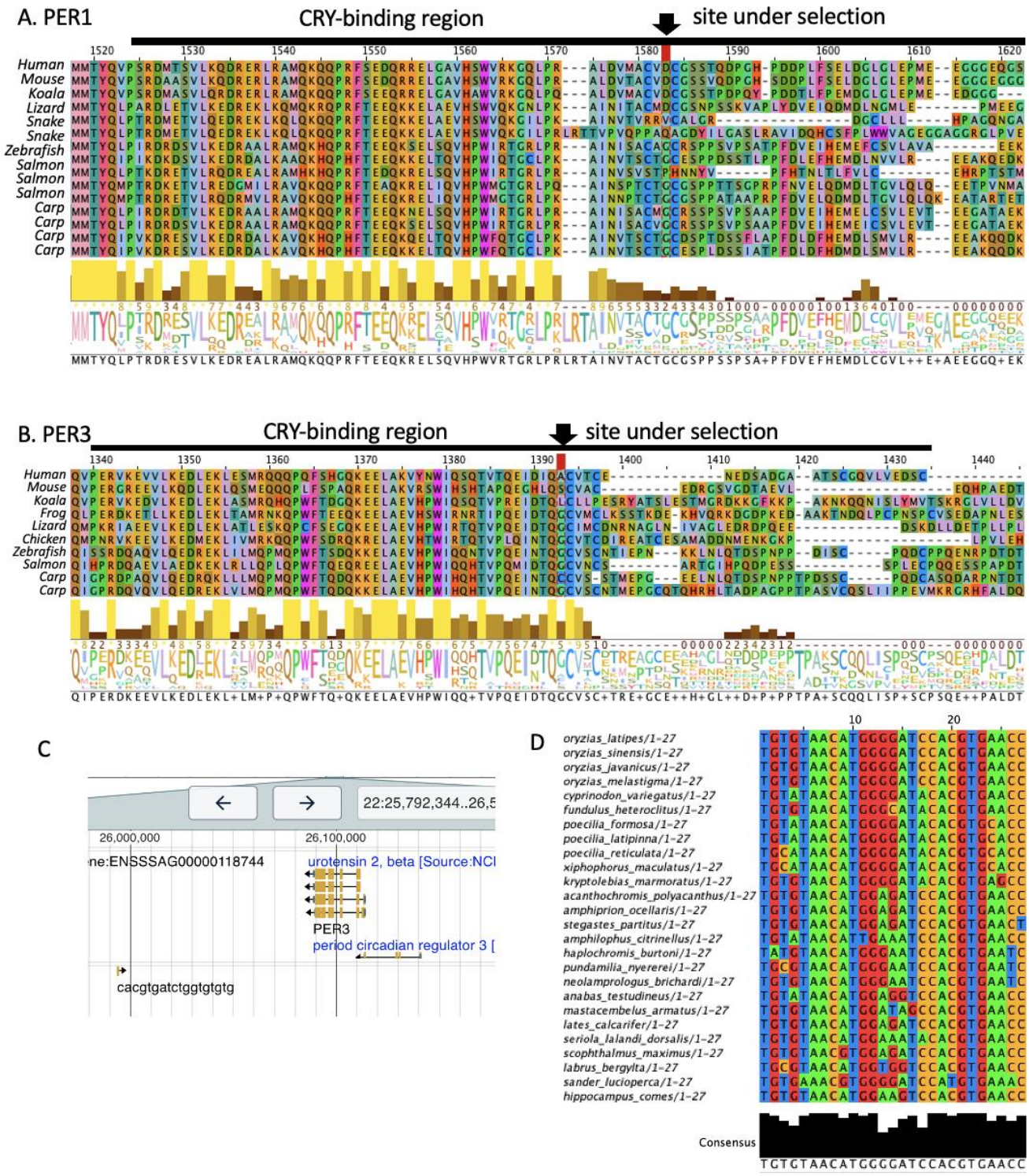

Fig 5. AB. The site with diversifying selection detected (arrow) by MEME (Mixed Effects Model of Evolution) [79] on Datamonkey version 2.0 [80], which are both located in the CRY_binding sites of *Per1* (**A**) and *Per3* (**B**). **C**. A sequence, which is similar to the mammalian *PER1*-regulatory 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 a broad fish species, in *Percomorpha.*

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. MEME results for *Per* genes.

Figure S1 RELAX result of salmonid *per3* gene.

Figure S2 Brain *PER* gene expression in humans.

Figure S3 Diversifying selection signature in Period genes in vertebrates.

File S1, S2, S3. Period gene (*Per*) alignment from high-quality species (from Wasabi App).

File S4, R script for sequence curation and visualization.

File S5, A timetree-based species tree.

References

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

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

3. Roenneberg T, Merrow M. 2016 The Circadian Clock and Human Health. *Curr. Biol.* **26**, R432–43.

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

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

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

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

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

9. Zehring WA, Wheeler DA, Reddy P, Konopka RJ, Kyriacou CP, Rosbash M, Hall JC. 1984 P-element transformation with period locus DNA restores rhythmicity to mutant, arrhythmic Drosophila melanogaster. *Cell* **39**, 369–376.

10. Tei H, Okamura H, Shigeyoshi Y, Fukuhara C, Ozawa R, Hirose M, Sakaki Y. 1997 Circadian oscillation of a mammalian homologue of the Drosophila period gene. *Nature* **389**, 512–516.

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

12. 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.

13. Pendergast JS, Friday RC. 2009 Endogenous rhythms in Period1 mutant suprachiasmatic nuclei in vitro do not represent circadian behavior. *Journal of Neuroscience*

14. 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.

15. Relógio A, Westermark PO, Wallach T, Schellenberg K, Kramer A, Herzel H. 2011 Tuning the mammalian circadian clock: robust synergy of two loops. *PLoS Comput. Biol.* **7**, e1002309.

16. Idda ML, Bertolucci C, Vallone D, Gothilf Y, Sánchez-Vázquez FJ, Foulkes NS. 2012 Circadian clocks: lessons from fish. *Prog. Brain Res.* **199**, 41–57.

17. 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.

18. 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.

19. Ono D, Honma S, Nakajima Y, Kuroda S, Enoki R, Honma K-I. 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.

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

21. 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.

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

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

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

25. 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.

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

27. Bae K, Jin X, Maywood ES, Hastings MH, Reppert SM, Weaver DR. 2001 Differential functions of mPer1, mPer2, and mPer3 in the SCN circadian clock. *Neuron* **30**, 525–536.

28. 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.

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

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

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

32. Siepka SM, Yoo S-H, Park J, Song W, Kumar V, Hu Y, Lee C, Takahashi JS. 2007 Circadian mutant Overtime reveals F-box protein FBXL3 regulation of cryptochrome and period gene expression. *Cell* **129**, 1011–1023.

33. Ramanathan C, Xu H, Khan SK, Shen Y, Gitis PJ, Welsh DK, Hogenesch JB, Liu AC. 2014 Cell type-specific functions of period genes revealed by novel adipocyte and hepatocyte circadian clock models. *PLoS Genet.* **10**, e1004244.

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

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

36. 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.

37. Shimomura H, Moriya T, Sudo M, Wakamatsu H, Akiyama M, Miyake Y, Shibata S. 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.

38. 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.

39. 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.

40. 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.

41. 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)

42. Pagel M. 1997 Inferring evolutionary processes from phylogenies. *Zool. Scr.* **26**, 331–348.

43. Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Smith WL. 2012 Resolution of ray-finned fish phylogeny and timing of diversification. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 13698–13703.

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

45. 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.

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

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

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

49. 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.

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

51. 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.

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

53. 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)

54. 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 (ed LZ Garamszegi), pp. 105–130. Berlin, Heidelberg: Springer Berlin Heidelberg.

55. Chiou Y-Y, Yang Y, Rashid N, Ye R, Selby CP, Sancar A. 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.

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

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

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

59. 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.

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

61. 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.

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

63. Pereira DS, van der Veen DR, Gonçalves BSB, Tufik S, von Schantz M, Archer SN, Pedrazzoli M. 2014 The effect of different photoperiods in circadian rhythms of per3 knockout mice. *Biomed Res. Int.* **2014**, 170795.

64. Blake JA, Baldarelli R, Kadin JA, Richardson JE, Smith CL, Bult CJ. 2020 Mouse Genome Database (MGD): Knowledgebase for mouse–human comparative biology. *Nucleic Acids Res.* **49**, D981–D987.

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

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

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

68. 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.

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

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

71. Fukushima K, Pollock DD. 2020 Amalgamated cross-species transcriptomes reveal organ-specific

propensity in gene expression evolution. *Nat. Commun.* **11**, 4459.

72. Davesne D, Friedman M, Schmitt AD, Fernandez V, Carnevale G, Ahlberg PE, Sanchez S, Benson RBJ. 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)

73. 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.

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

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

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

77. 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.

78. West AC, Iversen M, Jørgensen EH, Sandve SR, Hazlerigg DG, Wood SH. 2020 Diversified regulation of circadian clock gene expression following whole genome duplication. *PLoS Genet.* **16**, e1009097.

79. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012 Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8**, e1002764.

80. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. 2018 Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Mol. Biol. Evol.* **35**, 773–777.

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

82. 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.

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

84. Nakahata Y, Yoshida M, Takano A, Soma H, Yamamoto T, Yasuda A, Nakatsu T, Takumi T. 2008 A direct repeat of E-box-like elements is required for cell-autonomous circadian rhythm of clock genes. *BMC Mol. Biol.* **9**, 1.

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

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

87. 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)

88. Samy JKA, Mulugeta TD, Nome T, Sandve SR, Grammes F, Kent MP, Lien S, Våge DI. 2017 SalmoBase: an integrated molecular data resource for Salmonid species. *BMC Genomics* **18**, 482.

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

90. 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.

91. Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. 2022 TimeTree 5: An Expanded Resource for Species Divergence Times. *Mol. Biol. Evol.* **39**. (doi:10.1093/molbev/msac174)

92. 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.

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

94. 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)

95. 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)

96. Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, FitzJohn RG, Alfaro ME, Harmon LJ. 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)

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

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

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

100. Assis J, Tyberghein L, Bosch S, Verbruggen H, Serrão EA, De Clerck O, Tittensor D. 2018 Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Glob. Ecol. Biogeogr.* **27**, 277–284.

101. 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.

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

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

104. Froese R, Pauly D, Others. 2010 FishBase.

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

106. 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.

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

108. 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.

109. 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* (ed LZ Garamszegi), pp. 131–153. Berlin, Heidelberg: Springer Berlin Heidelberg.

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

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