



# Systematic Analysis of Mouse Genome Reveals Distinct Evolutionary and Functional Properties Among Circadian and Ultradian Genes

Stefano Castellana<sup>1†</sup>, Tommaso Mazza<sup>1†</sup>, Daniele Capocefalo<sup>1</sup>, Nikolai Genov<sup>2,3</sup>, Tommaso Biagini<sup>1</sup>, Caterina Fusilli<sup>1</sup>, Felix Scholkmann<sup>4,5</sup>, Angela Relógio<sup>2,3</sup>, John B. Hogenesch<sup>6</sup> and Gianluigi Mazzoccoli<sup>7\*</sup>

## OPEN ACCESS

### Edited by:

Giovanni Li Volti,  
Università degli Studi di Catania, Italy

### Reviewed by:

Luciana A. Campos,  
Universidad Camilo José Cela, Spain  
Giovanni Messina,  
University of Foggia, Italy

### \*Correspondence:

Gianluigi Mazzoccoli  
g.mazzoccoli@operapadrepio.it

†These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Integrative Physiology,  
a section of the journal  
Frontiers in Physiology

**Received:** 21 May 2018

**Accepted:** 06 August 2018

**Published:** 23 August 2018

### Citation:

Castellana S, Mazza T, Capocefalo D,  
Genov N, Biagini T, Fusilli C,  
Scholkmann F, Relógio A,  
Hogenesch JB and Mazzoccoli G  
(2018) Systematic Analysis of Mouse  
Genome Reveals Distinct Evolutionary  
and Functional Properties Among  
Circadian and Ultradian Genes.  
*Front. Physiol.* 9:1178.  
doi: 10.3389/fphys.2018.01178

<sup>1</sup> Bioinformatics Unit, IRCCS “Casa Sollievo della Sofferenza”, San Giovanni Rotondo, Italy, <sup>2</sup> Institute for Theoretical Biology (ITB), Charité – Universitätsmedizin Berlin and Humboldt University of Berlin, Berlin, Germany, <sup>3</sup> Molekulares Krebsforschungszentrum (MKFZ), Charité – Universitätsmedizin Berlin, Berlin, Germany, <sup>4</sup> Research Office for Complex Physical and Biological Systems (ROCoS), Zürich, Switzerland, <sup>5</sup> Department of Neonatology, University Hospital Zürich, University of Zürich, Zürich, Switzerland, <sup>6</sup> Divisions of Human Genetics and Immunobiology, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH, United States, <sup>7</sup> Division of Internal Medicine and Chronobiology Unit, IRCCS “Casa Sollievo della Sofferenza”, San Giovanni Rotondo, Italy

In living organisms, biological clocks regulate 24 h (circadian) molecular, physiological, and behavioral rhythms to maintain homeostasis and synchrony with predictable environmental changes, in particular with those induced by Earth’s rotation on its axis. Harmonics of these circadian rhythms having periods of 8 and 12 h (ultradian) have been documented in several species. In mouse liver, harmonics of the 24-h period of gene transcription hallmarked genes oscillating with a frequency two or three times faster than circadian periodicity. Many of these harmonic transcripts enriched pathways regulating responses to environmental stress and coinciding preferentially with subjective dawn and dusk. At this time, the evolutionary history of genes with rhythmic expression is still poorly known and the role of length-of-day changes due to Earth’s rotation speed decrease over the last four billion years is totally ignored. We hypothesized that ultradian and stress anticipatory genes would be more evolutionarily conserved than circadian genes and background non-oscillating genes. To investigate this issue, we performed broad computational analyses of genes/proteins oscillating at different frequency ranges across several species and showed that ultradian genes/proteins, especially those oscillating with a 12-h periodicity, are more likely to be of ancient origin and essential in mice. In summary, our results show that genes with ultradian transcriptional patterns are more likely to be phylogenetically conserved and associated with the primeval and inevitable dawn/dusk transitions.

**Keywords:** clock, gene, evolution, rhythmicity, circadian, ultradian

## INTRODUCTION

Recurring changes in environmental variables (i.e., temperature, atmospheric pressure, magnetism, ultra-violet radiation, humidity, food/water availability, etc.) greatly influence living organisms, impinging on processes and activities essential for individual and species survival. These include rest/wake and feeding/fasting cycles, hunting, courtship and mating behavior, *inter alia*.

Different molecular oscillatory mechanisms, consisting of endogenous non-transcriptional circuits as well as transcriptional/translational feedback loops, appeared within the Tree of Life as a result of convergent evolution, to cope with energetic cycles driven by the solar light (Pittendrigh, 1993; Dunlap, 1999; Takahashi, 2017). These biological clocks are common across species and phyla, allowing appropriate physiological/behavioral adaptation and rhythmicity to anticipate predictable environmental changes and providing survival advantage (Pittendrigh, 1993; Dunlap, 1999; Takahashi, 2017).

The most common and investigated biological rhythms have a roughly 24-h periodicity and resonate with the daily switch from the bright light of day to the darkness of night. The mammalian biological clock drives the oscillatory expression of ~43% of all protein coding genes in an organ-specific manner (Bozek et al., 2009; Zhang et al., 2014; Lehmann et al., 2015). Oscillating genes comprise core clock genes encoding biological clock components, which operate feedback loops with time delay through the contribution of post-translational modifications, and output genes, which drive time-related tissue-specific functions through downstream signaling pathways enriched in central and peripheral oscillators (Chiou et al., 2016; Kriebs et al., 2017; Trott and Menet, 2018). In addition to the 24-h periodicity, two clusters of genes cycling at the second (12-h) and third (8-h) harmonics of the circadian rhythm were identified through high temporal resolution analyses of the mouse transcriptome. These harmonics were found in intact mice and were lost *ex vivo* or under restricted feeding conditions (Hughes et al., 2009). The repertoire of rhythms observed in gene transcription could depend on binding strengths of the elements involved (genes and/or proteins) as well as transcription/translation dynamics involving degradation speed and temporal delays (Fuhr et al., 2015). As suggested, two-peak circadian oscillations in output gene expression may result from the interaction between transcription factors with definite circadian phase relationships and non-competitive binding to the promoters of regulated genes (Westermarck and Herzel, 2013) or through an oscillatory incoherent feed-forward loop (Martins et al., 2016). Despite the pervasiveness of biological clocks among species, evolutionary and functional properties of oscillating genes remain largely unexplored.

Among higher organisms, the mouse is the most commonly used animal model for studying biological rhythms from a genomic and molecular point of view. Oscillating genes have been observed to cluster together along the mouse genome, to be longer and to have more alternative transcripts with respect to non-oscillating genes (Zhang et al., 2014). Regulation of

oscillating genes expression has been also extensively studied in mouse: three main *cis*-regulatory elements are located in the proximity of gene promoters (E-box, D-box, Ror/Rev-Erb response elements) and allow their time-regulated expression (Takahashi, 2017).

Furthermore, insights into the physiological role of ultradian oscillating genes have been published, suggesting the importance of the ultradian expression of genes involved in metabolism regulation, feeding behavior and response to dawn/dusk transition-related stress situations (Hughes et al., 2009).

Another intriguing aspect of oscillating gene regulation regards codon usage and translation: recent experimental evidences report an unexpected non-optimized codon usage for key circadian genes within simple organisms (i.e., *Neurospora crassa*, *Synechococcus elongatus*, *Drosophila melanogaster*) (Xu et al., 2013; Zhou et al., 2013, 2015; Fu et al., 2016).

Thus, in this study we asked ourselves how and when genes with the three different oscillatory patterns originated along the Tree of Life. Furthermore, inspired by the abovementioned data regarding codon usage impact on circadian gene expression, we explored Synonymous Codon Usage for the whole set of mouse oscillating genes, by partitioning the gene set in the three period classes and comparing them to the global genetic background. Moreover, we inspected upstream genomic sequences within the mouse genome to extract possible Transcription Factor Binding Sites (TFBS), other than the known *cis*-regulatory elements that could provide additional insights into the regulation of oscillating genes.

We also carried out a comparative analysis of functional annotations for the 24, 12, and 8 h oscillating genes and evaluated the impact of these genes on organism viability (gene essentiality). Function and evolutionary conservation of ultradian mouse genes would suggest that such genes have played a crucial role during the evolutionary history of life on Earth, possibly determining a proper response to environmental stress conditions. We hypothesized that at various levels of biological organization their regulation was adapted in order to finely manage stress conditions dictated by recurrent transitions along the day, like the dawn/dusk transitions.

## RESULTS

We investigated different aspects of mouse genomics in order to better understand the relationships between evolutionary origin, regulation and function of the different oscillating gene subsets. Our reference mouse gene lists were taken from a single publicly available data set (Hughes et al., 2009). Different valuable although not fully comparable methodological approaches have been recently used in order to detect circadian and ultradian signals within transcriptomes (e.g., van der Veen and Gerkema, 2017; Zhu et al., 2017). The results of our well-established bioinformatics approach partly overlap with ultradian gene sets identified by these studies, which use different criteria to identify harmonics of circadian rhythmicity in gene expression.

## Phylogenetic Origin of Cycling Genes

We operated a systematic inspection of oscillating gene trees retrieved from the PhylomeDB, NCBI HomoloGene, Phyletic age resources (Huerta-Cepas et al., 2014; Chen et al., 2017; Ncbi Resource Coordinators, 2017). We considered the oldest phylogenetic node for each gene and those genes that originated early during evolution (Opisthokonts) were labeled as “ancient.”

Genes that presumably arose after Chordate radiation were labeled as “recent” (cf., **Table 1**). Raw data obtained from PhylomeDB, NCBI HomoloGene, and Phyletic age are reported in **Supplementary Table S1**.

From PhylomeDB, we were able to retrieve phylogenetic information for 15443 genes out of the global mouse transcriptome and for 1787 out of 2312 oscillating mouse

**TABLE 1** | Evolutionary origin of 8, 12, and 24 h oscillating genes sorted out of the global mouse transcriptome, according to PhylomeDB, NCBI HomoloGene, and Phyletic age.

	8 h	12 h	24 h	All genes
Taxonomic groups	(n = 43/56) % (n)	(n = 164/202) % (n)	(n = 1580/2054) % (n)	(n = 15443) % (n)
<b>PhylomeDB</b>				
<i>Prokaryota</i>	9.3 (4)	9.8 (16)	4.8 (76)	5.7 (879)
<i>Eukaryota</i>	27.9 (12)	52.4 (86)	30.4 (481)	29.2 (4516)
<i>Opisthokonts</i>	37.2 (16)	20.1 (33)	20 (316)	21.7 (3357)
“Ancient” genes	74.4 (32)	82.3 (135)	55.2 (873)	56.6 (8752)
<i>Chordate</i>	–	–	–	–
<i>Teleostomi</i>	20.9 (9)	12.2 (20)	20.2 (319)	22.8 (3525)
<i>Tetrapoda</i>	–	–	–	–
<i>Amniota</i>	0 (0)	2.4 (4)	11.3 (179)	8.8 (1346)
<i>Mammalia</i>	–	–	–	–
<i>Eutheria</i>	4.7 (2)	3 (5)	13.3 (209)	11.8 (1820)
“Recent” genes	25.6 (11)	17.7 (29)	44.8 (707)	43.4 (6691)
	(n = 49/56) % (n)	(n = 183/202) % (n)	(n = 1474/2054) % (n)	(n = 19032) % (n)
<b>NCBI HomoloGene</b>				
<i>Eukaryota</i>	20.4 (10)	35.5 (65)	15.9 (234)	13.1 (2500)
<i>Opisthokonts</i>	10.2 (5)	2.7 (5)	3.7 (54)	3.1 (587)
“Ancient” genes	30.6 (15)	38.2 (70)	19.6 (288)	16.2 (3087)
<i>Metazoa</i>	–	–	–	–
<i>Bilateria</i>	22.4 (11)	25.7 (47)	17.6 (259)	15.1 (2883)
<i>Vertebrata</i>	–	–	–	–
<i>Euteleostomi</i>	30.6 (15)	25.1 (46)	40.2 (592)	34.6 (6579)
<i>Tetrapoda</i>	4.2 (2)	7.2 (13)	8.6 (127)	7.7 (1459)
<i>Amniota</i>	–	1.6 (3)	3 (45)	3.5 (657)
<i>Mammalia</i>	–	–	–	–
<i>Boroetheria</i>	12.2 (6)	2.2 (4)	10 (148)	14.8 (2825)
<i>Euarchontoglires</i>	–	–	0.9 (14)	1.9 (357)
<i>Murinae</i>	–	–	0.1 (1)	5.2 (996)
exclusive to <i>Mus musculus</i>	–	–	–	1 (189)
“Recent” genes	69.4 (34)	61.8 (113)	80.4 (1186)	83.7 (15945)
	(n = 52/56) % (n)	(n = 189/202) % (n)	(n = 1882/2054) % (n)	(n = 22544) % (n)
<b>Phyletic age</b>				
<i>Eukaryota</i>	36.5 (19)	41.3 (78)	25.1 (472)	23.2 (5262)
<i>Opisthokonts</i>	1.9 (1)	8.5 (16)	6.6 (124)	54.2 (1227)
“Ancient” genes	38.5 (20)	49.7 (94)	31.7 (596)	28.8 (6489)
<i>Metazoa</i>	32.7 (17)	28 (53)	25.9 (488)	2.2 (5351)
<i>Chordate</i>	21.1 (11)	13.2 (25)	24.2 (455)	23.7 (4753)
<i>Mammalia</i>	3.8 (2)	3.7 (7)	9 (169)	21.1 (2011)
Exclusive to <i>Mus musculus</i>	3.8 (2)	5.3 (10)	9.2 (174)	17.5 (3940)
“Recent” genes	61.5 (32)	50.3 (95)	68.3 (1286)	71.2 (16055)

All genes refers to the global mouse genetic background. For each taxon, the number of genes with putative origin in that taxon are shown.

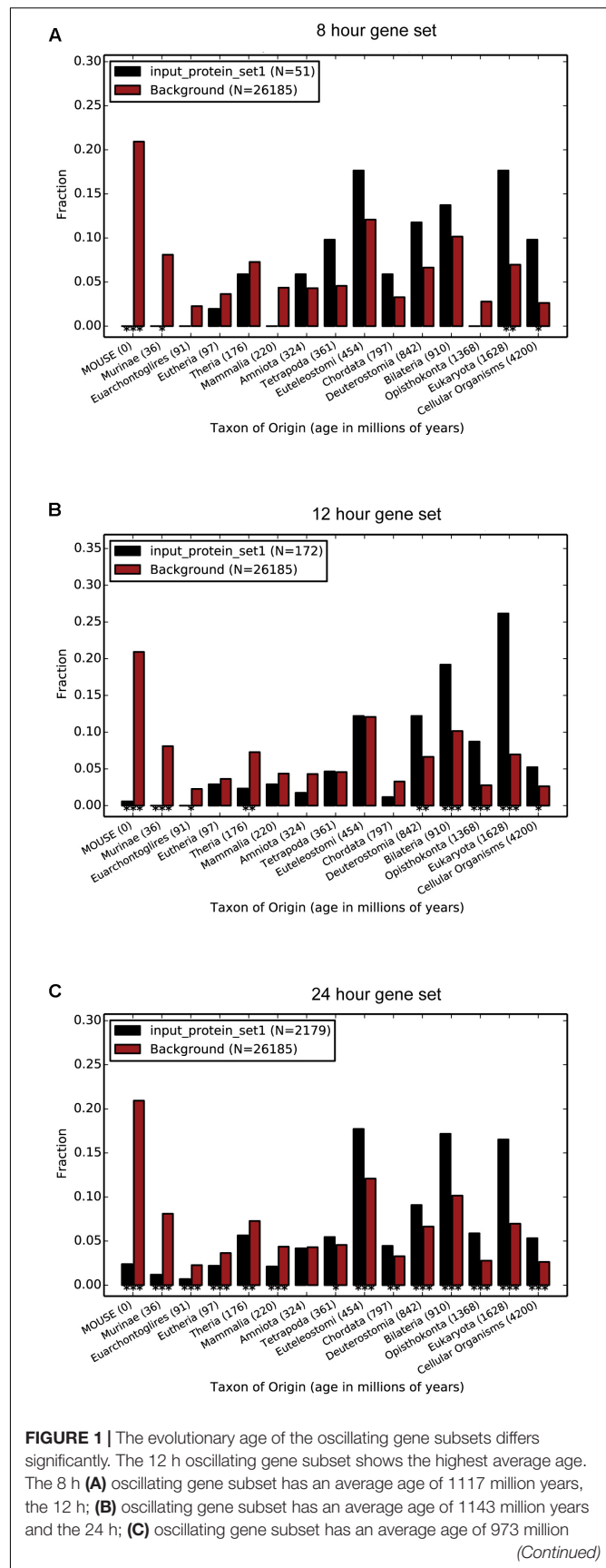
genes. Genes oscillating with an 8- and 12-h periodicity were significantly enriched in “ancient” genes with respect to the global mouse transcriptome (Fisher’s Exact Test for Count Data; respectively,  $p = 0.01$  and  $p = 3.58e^{-12}$ ; 95% CI 1.21-Inf, 2.51-Inf; Odds Ratios 2.22, 3.55). No enrichment was found for mouse genes oscillating with a 24-h period. Besides, 8 and 12 h oscillating genes were enriched in “ancient” genes with respect to the 24 h oscillating ones (respectively,  $p = 0.008$  and  $p = 2.37e^{-12}$ ; 95% CI: 1.26-Inf and 2.63-Inf; Odds Ratios: 2.35 and 3.76).

We then analyzed homologous genes in NCBI Homologene, with particular focus on *Mus musculus* species group. We retrieved homologous accession numbers, gene symbols, and conservation data from the summary output, mapped oscillatory genes, and calculated the relative proportions. We further retrieved 1706 oscillating genes out of the global mouse transcriptome (19032 genes), split them according to the “ancient” and “recent” classes, and calculated asymmetries among counts. The 12 h oscillating gene subset was strongly enriched in “ancient” genes ( $p = 9.72e^{-13}$ ; 95% CI 2.45-Inf; Odds Ratio 3.19). In this analysis, circadian genes and 8 h oscillating genes were slightly enriched in “ancient” genes with respect to the global mouse transcriptome ( $p = 0.01$ ; 95% CI: 1.11-Inf; Odds Ratio 1.25;  $p = 0.0091$ ; 95% CI: 1.28-Inf; Odds Ratio 2.27). Statistically significant differences emerged when comparing the sets of 12 and 24 h oscillating genes ( $p = 3.57e^{-8}$ ; 95% CI: 1.91-Inf; Odds Ratio 2.55) and, less significantly, for the comparisons of 8 and 24 h oscillating gene sets ( $p = 0.046$ , 95% CI: 1.01-Inf; Odds Ratio 1.81).

Finally, we analyzed oscillatory genes with the Phyletic age dataset. Genes oscillating with a 12-h periodicity were still enriched in “ancient” genes with respect to the global mouse transcriptome (Fisher’s Exact Test  $p = 1.28e^{-9}$ ; 95% CI: 1.9-Inf; Odds Ratio 2.44) and compared to the 24 h oscillating gene subset ( $p = 7.15e^{-7}$ ; 95% CI: 1.63-Inf; Odds Ratio 2.13). The 8 h oscillating gene subset did not differ from the other subsets and the global mouse transcriptome. Circadian genes were also significantly enriched in “ancient” genes when compared to the global mouse transcriptome ( $p = 0.004$ ; 95% CI: 1.05-Inf; Odds Ratio 1.14).

## Protein Age Estimation and Enrichment Analysis

Proteins found in different living species arose at specific evolutionary times and their evolutionary origins provide information about their possible function and interactions. To further explore the evolutionary conservation of the three oscillating gene subsets at the protein level, we determined their age using the ProteinHistorian tool (Capra et al., 2012). Each comparison was performed with respect to the average age of the genomic background (553 million years). Our data showed that the 8 h and the 12 h oscillating gene subsets have a similar average computed age of 1117 million years and of 1143 million years, respectively, while the 24 h oscillating gene subset was found to be less ancient with an average age of 973 million years (Figures 1A–C). Interestingly, the three oscillating gene subsets were more ancient than the genomic background,



**FIGURE 1** | The evolutionary age of the oscillating gene subsets differs significantly. The 12 h oscillating gene subset shows the highest average age. The 8 h (A) oscillating gene subset has an average age of 1117 million years, the 12 h; (B) oscillating gene subset has an average age of 1143 million years and the 24 h; (C) oscillating gene subset has an average age of 973 million

(Continued)

**FIGURE 1 | Continued**

years. All the three subsets are significantly older than the genomic background, having an average age of 553 million years. Since the gene subsets are derived from experiments performed with modern mice, the age of the mouse has been set to 0 million years on the x-axis. The remaining ages refer to the emergence of the corresponding branches of the Tree of Life. Data for the computed evolutionary age of the oscillating gene subsets were gathered by the Protein Historian tool.

and can be traced back to very ancient branches of the Tree of Life. In the case of the 8 h oscillating gene subset, the bins containing the highest number of genes are located at the Euteleostomi and Eukaryota branches (**Figure 1A**). As for the 12 h oscillating gene subset, the bins containing the highest number of genes are located at the Eukaryota branch (**Figure 1B**). As for the 24 h oscillating gene subset, we found three large bins at Euteleostomi, Eukaryota and Deuterostomia branches (**Figure 1C**). Nevertheless, given the larger size of the 24 h oscillating gene subset, as compared to the genes oscillating with 12- and 8-h periodicity, it is conceivable that smaller sets of more ancient genes exist, but that these are masked by the large size of the background set of genes. The fact that all the three oscillating gene subsets contain a large bin at the Eukaryota level, hints that most of the oscillating genes developed early during the evolution of life on Earth and have been conserved over a long period of evolutionary time, suggesting that they may play an essential function in many organisms.

## Essentiality of Oscillating Genes

Genes that are necessary for an organism's survival are defined as "essential." Gene essentiality refers to the fact that such genes play a fundamental role so that their dysfunction cannot be compensated by other genes or molecular mechanisms. Essential genes are then, theoretically, more likely to be of ancient origin because they constitute the "minimal" necessary genetic repertoire in order for a living being to survive in its environment.

By consulting the OGEEv2 databank<sup>1</sup> (Chen et al., 2017), we retrieved information for mouse genome essentiality, by focusing on oscillating gene subsets (**Table 2**).

We extracted and focused our analysis on 9041 mouse genes with some experimental indication of essentiality/non-essentiality. The 68.5% of genes oscillating with 12-h periodicity were categorized as essential with respect to the set of all genes (Fisher's Exact Test for Count Data;  $p = 6.305 \times 10^{-5}$ , CI: 1.59-Inf, Odds Ratio: 2.35). No significant disproportion was found for the 8 and 24 h oscillating gene subsets. These data corroborate the hypothesis that the 12 h oscillating genes are critical components of the eukaryotic cell, in order that they tend to be evolutionary conserved and essential.

## Synonymous Codon Usage in Oscillating Genes

The different patterns of synonymous codon usage (SCU) within and among genomes have been generally related to efficient

translation or accurate protein folding. Although intriguing findings emerged for important circadian genes within basal species (*N. crassa*, *S. elongatus*, *D. melanogaster*) (Vicario et al., 2007), no systematic inspection of codon usage has been carried out in complex species, such as *M. Musculus*.

First, we calculated the Effective Number of Codons (ENC) and the Codon Adaptation Index (CAI) for the total pool of mouse protein-coding genes (26780 common sequences between the CodonW and Jcat outputs) and then evaluated potential differences among oscillatory gene subsets and the genetic background.

Given that preferentiality of SCU is a tradeoff between gene-specific mutational bias and purifying selection, we performed a further analysis by removing genes with extreme compositional properties. We considered the GC content of the third codon positions ('GC3s') as an estimate of the mutational pressure within genes: thus, we focused only on genes within the lowest and highest quartile ( $0.48 < \text{GC3s} < 0.66$ ). **Table 3** resumes comparative tests for the two indexes and the 3 oscillatory gene classes ('n' indicates the number of the resulting sequences after GC correction). Genes that oscillate with 12-h periodicity generally have non-optimized SCU (higher ENC and smaller CAI with respect to the 13323 background sequences), while the SCU of the other gene subsets was found to be similar to the general scenario. Raw outputs of CodonW and JCat tools are presented in **Supplementary Table S3** and original output of CodonW and JCat in, respectively, **Supplementary Files S1, S2**.

**Figure 2** shows box plots rendering a descriptive view of the ENC and CAI value distribution among the oscillating gene subsets with different oscillation periodicity.

## Functional Differences Among Oscillating Gene Subsets

We surveyed the functional impact of mouse genes oscillating with 8-h, 12-h, and 24-h periods by performing functional (pathway, gene ontology) enrichment analyses through the use

**TABLE 2 |** Number and proportion of essential/non-essential oscillating mouse genes according to the OGEEv2 database.

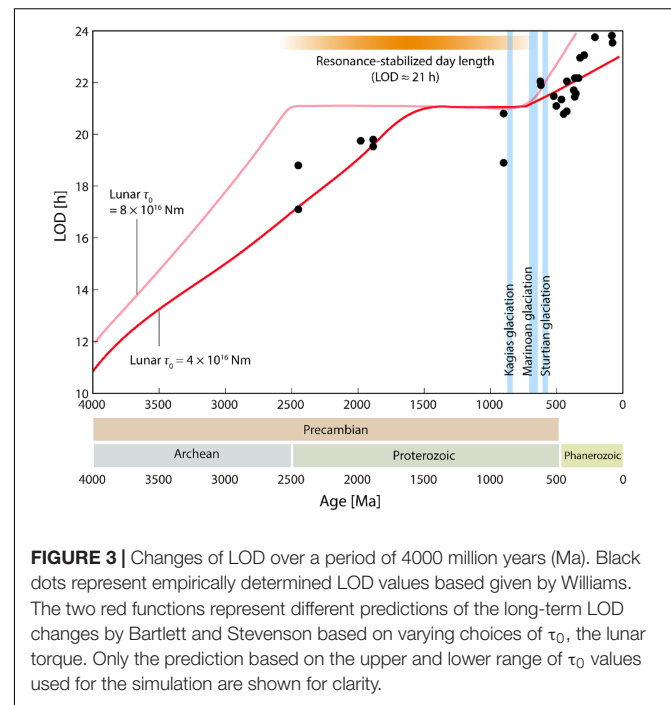
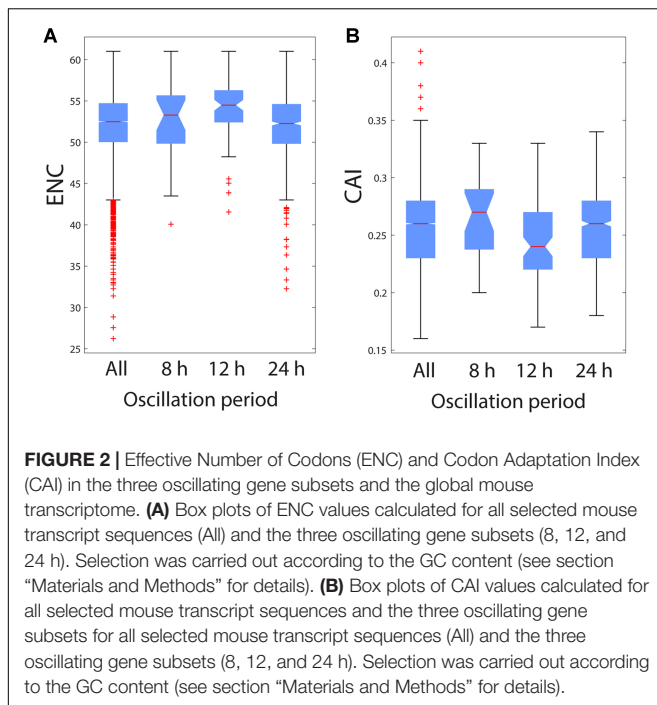
Essentiality	8 h	12 h	24 h	All genes
Essential	15 (43%)	63 (68.5%)	392 (47%)	4341 (48%)
Non-essential	20 (57%)	29 (31.5%)	446 (53%)	4700 (52%)
Total genes retrieved	35 (100%)	92 (100%)	838 (100%)	9041 (100%)

All genes refers to the global mouse genetic background.

**TABLE 3 |** Results of Wilcoxon Tests for ENC/CAI distributions of 24, 12, and 8 h oscillating gene subsets against the whole mouse genome; "n" represent s the size of each distribution.

Comparison	ENC		CAI	
	(n = 13309/1031/88/25)		(n = 13309/1031/88/25)	
All genes versus	24 h	Greater ( $p = 0.04$ )	Not different ( $p = 0.27$ )	
	12 h	Smaller ( $p = 1.64 \times 10^{-6}$ )	Greater ( $p = 8.26 \times 10^{-5}$ )	
	8 h	Not different ( $p = 0.48$ )	Not different ( $p = 0.31$ )	

<sup>1</sup><http://ogee.medgenius.info/downloads/>



of Ingenuity Pathway Analysis (IPA) (QIAGEN Inc., Ingenuity Pathway Analysis) (**Supplementary Table S5**).

Furthermore, we performed a comparative analysis, evidencing differentially enriched pathways among the three oscillating gene subsets, although only the top-ranking ones are reported (**Supplementary Table S6**).

The most overrepresented pathways for genes oscillating with 8-h periodicity are related to sterol and fatty acids metabolism as well as oxidative stress, while a significant portion of 12 h oscillating genes is involved in protein folding/ubiquitination and hormone signaling. Regarding 24 h oscillating genes, many functions/pathways resulted to be significantly enriched and among these the category ‘circadian rhythm’ was ranking among the most statistically significant pathways in our gene expression analysis in mouse liver (see **Supplementary Table S4**).

## Enrichment Analysis of TFBS Within Oscillating Genes

We explored the genomic context of ultradian genes in search of over-represented binding sites. The possible presence of additional *cis*-regulatory elements that are specific for 8 and 12 h oscillating gene subsets beyond well-established *cis*-acting elements associated with 24 h oscillating gene expression could influence their different oscillatory pattern with respect to the circadian genes.

The implementation of AME tool on oscillating vs. non-oscillating genes upstream sequences revealed very interesting results, particularly with regard to the subset of 202 genes oscillating with a 12-h period. We were able to retrieve a set of enriched motifs (adjusted *p*-value < 0.05) for the comparisons with five distinct random datasets (12, 13, 10, 14, 9; see section “Materials and Methods”). These pools

greatly overlap within comparative analyses, with five common inferred TFBS (Zfp1\_primary, Zfp1\_secondary, XBP1.C, MBD2.B, GABPA.A) frequently resulting as over-represented. A significant enrichment for these five elements was also observed by considering upstream sequences of genes oscillating with a 12-h periodicity and upstream sequences of 24 h oscillating genes as control dataset.

No significantly enriched motif was detected when comparing upstream sequences of 8 h oscillating genes with the random datasets and for 4 out of 5 comparisons of upstream sequences of genes oscillating with a period of 8 h and upstream sequences of 24 h oscillating genes as control dataset.

Two motifs (ERR3.B, E4F1.D) were enriched within upstream regions of genes oscillating with a 24-h periodicity with respect to a single random sample of 1500 upstream sequences.

**Supplementary Table S7** lists the top ranking motifs with significant enrichment in almost all comparative analyses; reported *p*-values are specific for the first comparison, although they result very low for all the performed comparative analyses (command lines and raw output in **Supplementary File S3**).

## DISCUSSION

The basic role played by biological timing systems is reflected by the fact that molecular oscillators tick in the three domains of life and oscillating genes have been discovered in phylogenetically distant species, ranging from cyanobacteria to humans. The cycling period of biological rhythms spans from minutes to years (Mazzocchi et al., 2016), but perhaps the largest class is attuned with a 24-h rhythmicity dictated by Earth’s rotation around its axis. Different astronomical and geological factors affected

Earth's rotation during its history. While estimates of the rotation period are imprecise, when early life appeared, day length was significantly shorter than today. Interestingly, in relation to the highly metamorphosed nature of the oldest sedimentary rocks, there are no confirmed microfossils older than 3.5 billion years on Earth, but proofs support biological activity in submarine-hydrothermal environments and oxidized biomass more than 3.8 billion years ago (Dodd et al., 2017). Day length at the time of origin of very primitive photosynthesis is estimated to be around 12 h, while day length at the dawn of eukaryotes is estimated to be less than 18 h (Zahnle and Walker, 1987; Sonett et al., 1996; Williams, 2000; Bartlett and Stevenson, 2016) (**Figure 3** and **Supplementary File S4**).

Interestingly, our systematic investigation of mouse gene phylogenies revealed that genes exhibiting ultradian oscillatory patterns of expression are enriched of ancient genes, which arose at the early stages of life evolution. Many eukaryote species retain one (or more) copies of such genes; consistently, they have been classified as “essential” for mouse survival. Ultradian period oscillating genes comprised in mouse gene subsets are particularly enriched for the Unfolded Protein Response, (*Ddit3*, *Insg1*, *Dnajb9*, *Edem3*, *Hspa1b*, *Hspa5*, *Hspa13*) as well as the Protein Ubiquitination pathway (*Usp28*, *Usp46*, *Usp11*, *Ube1dc1*, *Ube1l2*, *Ube2g2*, *Anapc10*, *Dnajb9*, *Hspa5*, *Psmc5*, *Anapc4*, *Dnajb1*, *Vhl*).

Ultradian rhythms are known to play a key role in physiology, as evidenced by the 12-h activation rhythm of the IRE1 $\alpha$ -XBP1 pathway in the endoplasmic reticulum driving the two-peak activation of the Unfolded Protein Response (UPR)-regulated genes hallmarking hepatic lipid metabolism (Cretenet et al., 2010). UPR is an adaptive response conserved across kingdoms, phyla and species and committed to manage the accumulation of unfolded proteins in the endoplasmic reticulum to increase expression of genes entailed in endoplasmic reticulum function to re-establish/augment the capability of folding and modify proteins (Iwata and Koizumi, 2012; Hollien, 2013; Zhang et al., 2016). Also, the Ubiquitin/Proteasome interactive pathway is evolutionary conserved and essential in organisms ranging from yeast to mammals for the coordinated and temporal targeted degradation of short-lived proteins or proteins that do not fold accurately within the endoplasmic reticulum. Monoubiquitination is involved in endocytosis, subcellular protein localization/trafficking variations and DNA damage response, whilst multiple ubiquitination rounds play a role in protein targeting for proteasomal degradation (Budhidarmo et al., 2012; Amm et al., 2014).

The preferential usage among synonymous codons is a common property of many genomes from Prokaryotes to Eukaryotes. This bias, CUB, is thought to reflect a balance between mutational pressure and selective forces. The differential SCU among genes could act as a regulator of the translational process, by shifting from a rapid peptide synthesis to a slower one in which the peptide chain has an appropriate time to fold correctly (Yu et al., 2015). In simple model organisms, efforts were focused on the study of the relationship between SCU and translational efficiency/accuracy. For example, in the cyanobacterium *S. elongatus*, two strains with optimized and

non-optimized (wild type) SCU for the KaiABC core clock system were tested for rhythmicity preservation. The wild type strain with a lower CUB was found to be more efficient in maintaining rhythmicity in different environmental conditions (Xu et al., 2013). A non-optimal SCU was also tested in the Frequency (FRQ) circadian protein in *N. crassa* through construction of an optimized version of the FRQ coding sequence: circadian rhythm obliteration was observed *in vivo* (Zhou et al., 2013, 2015). The usage of “unpreferred” synonymous codons was shown to slow-down FRQ mRNA translation, ensuring appropriate protein folding, especially for the Intrinsically Disordered Regions (‘IDP’) within it (Zhou et al., 2013, 2015). Studies on the *D. melanogaster* circadian gene *Period* and CUB confirmed that artificially induced codon optimization would cause protein instability and alteration of circadian rhythmicity regulation (Fu et al., 2016). Our analyses performed on mouse oscillating genes showed that ultradian genes generally have more relaxed codon usage. This phenomenon could be linked to the need for appropriate protein folding, although it is very hard to demonstrate this hypothesis for such a large gene set using computational modeling and functional assays.

The results of our study are globally in agreement with the general notion that ancient genes have a propensity for non-biased codon usage, as detected by a recent analysis on the human genome showing that gene age has a significant inverse correlation with SCU bias (with older genes having Codon Deviation Coefficient values toward 0) (Yin et al., 2016). This issue is further complicated when gene expression is taken into account. However, while housekeeping (ubiquitous, highly expressed) genes show optimized SCU (Ma et al., 2014), selection for translational efficiency could be less relevant in case of tissue-specific genes or genes with highly variable expression, such as oscillating genes. However, we cannot exclude that, for some specific oscillatory gene, CUB could be determined by mechanisms resembling those found for circadian genes in *N. Crassa* or *S. elongatus*.

Dawn/dusk transitions impact rhythmic gene expression as evidenced by experiments performed in *N. crassa* and showing that most of the oscillating transcripts managing metabolic functions are produced by dawn and dusk specific transcription (Sancar et al., 2015).

Besides, typical circadian genes show peak expression levels all over the 24 h with modest trend in their phase distribution, whereas the largest fraction of genes oscillating with 12-h periodicity gather into a unique group with comparable phases and most of them peak at dusk and dawn, hinting that these genes may anticipate light/darkness transition-related stress on a daily basis (Hughes et al., 2009).

In addition to diurnal patterns of solar illumination, Earth's rotation on its axis dictates environmental temperature fluctuations. Even though life forms are accustomed to temperatures from the freezing to over the boiling point of water, heat stress caused by temperatures barely exceeding the relative most favorable growth temperature severely limits survival (Richter et al., 2010). Interestingly, some of the oscillating genes comprised in the 12-h periodicity subset (*Hspa1b*, *Hspa5*, *Hspa13*, *Dnajb1*, *Dnaja4*, *Dnajb9*) encode heat shock proteins (HSP).

HSP work as molecular chaperones to avert the formation of anomalous protein aggregates and support proteins in the acquisition of their native structures, enriching an ancient signaling pathway activated by abrupt temperature increase (heat stress) harming vital cellular structures and obstructing crucial functions (Carra et al., 2017; Conway de Macario et al., 2017). Moreover, UPR activity fluctuates with ultradian pattern in mouse liver cells and especially in the endoplasmic reticulum. Interestingly, the expression of XBP1, driving one of the three UPR effector pathways, oscillates with 12-h periodicity (Cretenet et al., 2010) and regulates the downstream expression of genes implicated in protein folding, turnover and proteo-toxicity control (Chaix et al., 2016). The presence of XBP1 binding sites on upstream sequences of 12 h oscillating genes would support the hypothesis that these ultradian genes played a key role during the evolution of all life forms on planet Earth.

In addition to the genes oscillating with a 12-h period, the 8 h oscillating gene subset included ancient genes and enriched pathways related to sterol synthesis and oxidative stress. Fascinatingly, sterol biosynthesis is an O<sub>2</sub>-intensive process, and relevant molecular fossils contain steroidal hydrocarbons (steranes), which are the diagenetic remains of sterol lipids and hint at Paleoproterozoic sterol biosynthesis (Gold et al., 2017). More elementary sterols can be synthesized by some bacteria, and complex sterols with modified side chains are synthesized by eukaryotes. Complex steranes found in ancient rocks suggest aerobic metabolic processes rather than the presence of eukaryotes, dating the divergence time of bacterial and eukaryal sterol biosynthesis genes around 2.31 billion years ago. This is consistent with the most recent geochemical evidence for the great oxidation event (Gold et al., 2017). O<sub>2</sub> production completely transformed Earth's surface and atmosphere during the Proterozoic Eon, leading to substantial atmospheric O<sub>2</sub> levels rise about 3 billion years ago with fading of the reducing conditions prevalent during the Archean. The great oxygenation event was hallmarked by considerable and irreversible O<sub>2</sub> accumulation in the atmosphere starting in the early Paleoproterozoic and boosting during the Mesoproterozoic (Bekker et al., 2004; Cavalier-Smith, 2006; Parnell et al., 2010; Crowe et al., 2013; Hamilton et al., 2016; Luo et al., 2016). During this period, life increased exponentially, however, oxygen is harmful for nucleic acids and avoiding exposure of these biomolecules to the reductive phases of the respiratory cycle could prevent oxidative damage. Possibly, the oldest unicellular autotrophic organisms, for instance the cyanobacteria, evolved ultradian rhythmicity to coordinate DNA replication and transcription with mitochondrial redox activity, differently from the more recent life forms, such as unicellular algae, hallmarked by circadian rhythmicity. Intriguingly, aerobic metabolism was more proficient, but O<sub>2</sub> reactivity and toxicity threatened living organisms, which developed biochemical tools to defend against oxidative damage. These comprise enzymatic and non-enzymatic antioxidants blocking dangerous reactive oxygen species and among the genes oscillating with 8-h periodicity we found several genes, such as *Pon1*, *Cat*, *Apoa1*, *Apoa2*, *Prdx1*, *Hmox2*, encoding proteins involved in redox reactions and other molecular interactions directly associated to antioxidant activity

(Oda et al., 2001; Neumann et al., 2003; Barter et al., 2004; Lacher et al., 2015). Among these genes, particularly interesting is the *Cat* gene, encoding catalase, a highly conserved heme enzyme, capable to remove hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) and present in the peroxisomes of nearly all aerobic cells; accordingly, the catalase evolutionary process started in the Archaean and boosted in the Proterozoic, in line with the increase in atmospheric O<sub>2</sub> (Zámocký et al., 2012).

## CONCLUSION

The expression of oscillating genes shows rhythmicity in more than one time domain and the harmonics of transcription periodicity could depend, further than on transcriptional dynamics, also on phylogenetic components related to geochemical changes and evolutionary mechanisms. Ultradian rhythms are crucial for organismal physiology and characterize essential biological functions and adaptative processes conserved across kingdoms, phyla and species, such as metabolic cycles and the response to cell stress. In particular, 12 h oscillating genes are more likely to be ancient and essential, suggesting that these genes may have evolved to deal with the twice-daily cycles of transition between light/heat and dark/cold. Our comprehensive computational analyses show that the genes oscillating at the second and third harmonic of circadian rhythmicity are phylogenetically more conserved across all kingdoms of life and may represent an evolutionary genomic footprint.

## MATERIALS AND METHODS

### Primary Dataset

We exploited a dataset compiled as previously described (Hughes et al., 2009) and available from GEO (GSE11923). Rhythmic genes were identified in mouse hepatic tissue using both COSOPT and Fisher's *G*-test at a false-discovery rate of <0.05. Clusters of rhythmic genes with period lengths of approximately 24 h (>20 and <30 h), 12 h (>10 and <14 h) and 8 h (>7 and <9 h) were observed. We selected lists of array probe IDs/nucleotide sequences for which oscillating expression levels were determined and performed our analyses on three oscillating gene subsets: 56 (8 h), 202 (12 h), and 2396 (24 h) Ensembl Transcript IDs were retrieved from the primary dataset, by using BioDBnet<sup>2</sup> db conversion tool, respectively.

### Analysis of Gene Genealogies

Gene phylogenies can be resolved by numerous computational techniques. Given the elevate number of investigated genes (more than 2000 oscillating genes), we opted to consider pre-computed phylogenetic gene trees from three popular bioinformatic resources, PhylomeDb<sup>3</sup>, NCBI HomoloGene<sup>4</sup>, and the OGEEv2 Portal<sup>5</sup>.

<sup>2</sup><https://biodbnet-abcc.ncifcrf.gov/db/db2db.php>

<sup>3</sup><http://phylomedb.org/>

<sup>4</sup><http://www.ncbi.nlm.nih.gov/homologene/>

<sup>5</sup><http://ogee.medgenius.info/>



PhylomeDB was programmatically accessed through the web resource, focusing on the Reference Model Species Metaphylome (id: 504) of *M. musculus* and extracting from it data consisting of: the submitted mouse gene Ensembl ID (Ensembl Genes 83 release, January 2016; mouse genome version GRCm38.p4), official gene symbol and taxon name for the largest taxonomic group that comprises the investigated gene.

Secondarily, we extracted gene phylogenetic information from NCBI HomoloGene (build 146, accessed January 2016) using “*M. musculus* [ORGN]” as query term. In particular, we manipulated the NCBI “summary output” consisting of HomoloGene Accession Number, gene-specific degree of conservation (resumed with the formula “Conserved in taxon XX”) and relative gene symbol.

From the OGEEv2 databank (accessed January 2017), we downloaded the “gene essentiality” dataset, a collection of tested essential and non-essential genes in multiple species, and “gene age,” an additional dataset for gene evolutionary properties (Supplementary Tables S1, S2).

## Synonymous Codon Usage Analysis in *Mus musculus* Genome

Genomic investigations in a large spectrum of organisms have shown that the usage of synonymous codons is variable across species and genes. Several researchers demonstrated that asymmetry in the usage of codon within a certain codon family derives from the counterbalance of mutational pressure, selection and random drift. The quantification of the bias in synonymous codon usage, SCU (‘codon usage bias’ or ‘CUB’) can be performed by implementing numerous methods (Behura and Severson, 2013).

We primarily opted to calculate the “effective number of codons” (“ENC” or “ $N_c$ ”) as a measure of the bias that, similarly to the “effective number of alleles,” that is based on “ $F_k$ ” (average homozygosity of one degenerate codon family with “ $k$ ” synonymous codons), while integer numbers are relative to the twenty codon family (2 is associated to methionine and tryptophan, that are encoded by one single codon) (Wright, 1990).

$$N_c = 2 + \frac{9}{F_2} + \frac{1}{F_3} + \frac{5}{F_4} + \frac{3}{F_6}, \quad 20 < N_c < 61$$

With some adjustments, ENC ranges from 20 (maximum bias with a single synonymous codon used for each amino acid) to 61 (minimum bias, with all possible codons used). Importantly, this index is not based on any knowledge about organism-specific highly expressed genes and this is quite useful especially for species in which gene expression levels are poorly known.

Effective Number of Codons was calculated by using the popular CodonW Package<sup>6</sup> for 26814 protein-coding FASTA sequences as retrieved by Ensembl Biomart (Ensembl v82, mouse genome: GRCm38.p4).

Secondarily, we considered another CUB index, the “Codon Adaptation Index” (CAI) (Sharp and Li, 1987), defined as

“the geometric mean of the relative adaptiveness values ( $\omega$ ) of individual codons”:

$$CAI = \left( \prod_{i=1}^L (\omega_i) \right)^{1/L}$$

where “ $L$ ” represents the total number of codons, while  $\omega_i$  is the ratio of the observed codon frequency over the frequency of the most used codon for a certain amino acid. CAI ranges from 0 (usage of non-optimized codons) to 1 (usage of optimized codons). The “most used” or “most abundant” codons are calculated on codon usage of highly expressed genes. Thus, this measure necessitates of the knowledge of constitutive/highly expressed genes for a certain organism. Focusing on the mouse genome, we relied on the JCat webtool<sup>7</sup> that calculates CAI for a series of prokaryotic genomes and a limited number of Eukaryotes, among which *M. musculus* (run with default parameters, accessed at January 2016). We used the same 26814 FASTA sequences as input.

Distributions of ENC/CAI values for the entire mouse genome and the oscillatory gene subsets were compared by Wilcoxon rank sum tests. Furthermore, in order to “minimize” the impact of mutational pressure in determining CUB, we considered only the sequences with a comparable base composition. In detail, we calculated the GC content at third codon positions (‘GC<sub>3s</sub>’) for each mouse sequences and discarded those with GC content within the first or the third quartile. We then compared ENC or CAI distributions for the retained sequences.

## Relative Enrichment Analysis for Known TFBS in Oscillating Genes

We performed a differential enrichment analysis for known transcription binding, by focusing on the 5000 bp at the 5’ end of CDS initiation site for the classes of oscillating genes and, as controls, five sample datasets of 1500 randomly chosen upstream regions from mouse genes and additional five samples of 500 sequences. Sequences were retrieved from Ensembl build 84, by querying the GRCm38p2 mouse genome version through the Biomart tool. Background sequences were randomly picked up from the entire dataset (23111 regions), with the exclusion of those related to oscillating genes.

AME good usage practices<sup>8</sup> suggest that sequences should have comparable lengths, so we opted to work with a constant region size (5000 bp) for all the considered genes. We carried out 18 distinct runs of the AME program (MEME suite, 4.11.4 version) (Bailey et al., 2009) as follows: upstream sequences of 8 h oscillating genes versus the five control datasets; the same approach for 12 and 24 h oscillating genes sequences; 8 h oscillating genes versus 12 h oscillating genes (using the second dataset as control); 12 h oscillating genes versus 24 h oscillating genes; 8 h oscillating genes versus 24 h oscillating genes.

The procedure was repeated also for the 500 sequences control datasets. Three datasets of “meme-formatted” mouse regulatory motifs were screened, for a total of 825 elements.

<sup>7</sup><http://www.jcat.de/>

<sup>8</sup>[http://meme-suite.org/doc/ame.html?man\\_type=web](http://meme-suite.org/doc/ame.html?man_type=web)

<sup>6</sup><http://codonw.sourceforge.net/>

The AME output is produced in the form of textual and html files in which relevant enriched motifs are sorted according Fisher's Exact test  $p$ -values and adjusted  $p$ -values. Command lines and raw output are present in **Supplementary File S3**.

## Eukaryotic Protein Age Analyses

Protein age was evaluated by ProteinHistorian, a software based on three inputs, a species tree, a protein family database, and an ancestral family reconstruction algorithm with two parsimony algorithms: Dollo parsimony and Wagner parsimony. Dollo parsimony is based on the assumption that gaining a complex structure is much rarer than losing one. Thus, it assumes that there was a single gain event for each family, potentially followed by many losses in specific lineages. In other words, under Dollo parsimony, a family's origin is the most recent common ancestor (MRCA) of all species in which it is observed. Wagner parsimony allows multiple gain and loss events in an ancestral family reconstruction, as well as the ability to set weights on the relative likelihood of these events. By default, ProteinHistorian uses a relative gain penalty of 1. Since we focused on eukaryotic species in which horizontal gene transfer is rare, this largely serves to prevent false positives in the protein family databases from biasing age distributions (Capra et al., 2012).

## AUTHOR SUMMARY

Many crucial biological processes show regular changes, mainly with 24 h periodicity and resonate with the light/darkness alternation dictated by Earth's rotation on its own axis. Biological rhythms are driven by molecular clocks working through mechanisms conserved across phyla and species and maneuvered by proteins encoded by genes with circadian expression changes. A number of genes are hallmarked by fluctuation with less than 24 h periodicity, precisely at the second (12 h) and third (8 h) harmonic of circadian rhythmicity. At present, evolution of oscillating genes remains largely ignored and in this research article we present evidence that oscillatory patterns in gene transcription are partly related to evolutionary mechanisms. Considering that Earth's rotation slowed while the length of the day increased across centuries (from a few hours to approximately 24 h of the current solar day), we assumed that oscillatory harmonics in gene expression are likely to be ancient and well conserved across species. To investigate this hypothesis we carried out a systematic survey of evolutionary, regulatory, and functional properties of the entire mouse transcriptome in relation to frequency harmonics of gene expression and exploited sequence analysis of circadian and ultradian genes and proteins, across several species. Our results show that genes and proteins hallmarked by harmonics of circadian expression are conserved and are more likely to be essential and regulate stress responsive pathways. Ultradian patterns in the rhythmicity of gene expression could represent a characteristic genomic signature of evolutionistic eras and might correspond to evolutionary dynamics related to the biosphere response to environmental changes and to the geophysical context in which these rhythmic patterns originally developed.

## AUTHOR CONTRIBUTIONS

GM conceived the study. TM, AR, NG, JH, CF, DC, TB, and SC performed the bioinformatics analysis. FS computed length of day across centuries. GM, NG, and AR provided funding for the study. GM, SC, AR, NG, and JH wrote the paper. GM, TM, AR, NG, CF, FS, DC, TB, SC, and JH approved the final version of the manuscript.

## FUNDING

This study was supported by the "5×1000" voluntary contribution and by a grant to GM through Division of Internal Medicine and Chronobiology Unit (RC1603ME43, RC1703ME43, and RC1803ME40), IRCCS Scientific Institute and Regional General Hospital "Casa Sollievo della Sofferenza", Opera di Padre Pio da Pietrelcina, San Giovanni Rotondo (FG), Italy. AR and NG were funded by the German Federal Ministry of Education and Research (BMBF)—eBio-CIRSPLICE - FKZ031A316. NG was additionally funded by the Berlin School of Integrative Oncology (BSIO) of the Charité – Universitätsmedizin Berlin.

## ACKNOWLEDGMENTS

We acknowledge that this manuscript has previously been released as a pre-print at <https://www.biorxiv.org/content/early/2018/03/30/197111>.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphys.2018.01178/full#supplementary-material>

**TABLE S1** | Raw data obtained from PhylomeDB, NCBI HomoloGene, and Phyletic age.

**TABLE S2** | Dataset for gene evolutionary properties.

**TABLE S3** | Raw outputs of CodonW and JCat tools.

**TABLE S4** | Most statistically significant pathways in gene expression analysis in mouse liver.

**TABLE S5** | Top five enriched pathways for the three oscillating gene subsets.

**TABLE S6** | Comparative analysis based on  $-\log(p\text{-value})$  of enriched pathways among the three oscillating gene subsets.

**TABLE S7** | Summary of the top ranking over-represented Transcription Factor Binding Sites, as estimated by AME (MEME suite) comparative analyses.

**FILE S1** | Original output of CodonW.

**FILE S2** | Original output of JCat.

**FILE S3** | Command lines and raw output.

**FILE S4** | Day length estimation through ages.

## REFERENCES

- Amm, I., Sommer, T., and Wolf, D. H. (2014). Protein quality control and elimination of protein waste: the role of the ubiquitin-proteasome system. *Biochim. Biophys. Acta* 1843, 182–196. doi: 10.1016/j.bbamcr.2013.06.031
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335
- Barter, P. J., Nicholls, S., Rye, K. A., Anantharamaiah, G. M., Navab, M., and Fogelman, A. M. (2004). Antiinflammatory properties of HDL. *Circ. Res.* 95, 764–772. doi: 10.1161/01.RES.0000146094.59640.13
- Bartlett, B. C., and Stevenson, D. J. (2016). Analysis of a Precambrian resonance-stabilized day length. *Geophys. Res. Lett.* 43, 5716–5724. doi: 10.1002/2016GL068912
- Behura, S. K., and Severson, D. W. (2013). Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. *Biol. Rev. Camb. Philos. Soc.* 88, 49–61. doi: 10.1111/j.1469-185X.2012.00242.x
- Bekker, A., Holland, H. D., Wang, P. L., Rumble, D. III, Stein, H. J., Hannah, J. L., et al. (2004). Dating the rise of atmospheric oxygen. *Nature* 427, 117–120. doi: 10.1038/nature02260
- Bozek, K., Relógio, A., Kielbasa, S. M., Heine, M., Dame, C., Kramer, A., et al. (2009). Regulation of clock-controlled genes in mammals. *PLoS One* 4:e4882. doi: 10.1371/journal.pone.0004882
- Budhidarmo, R., Nakatani, Y., and Day, C. L. (2012). RINGs hold the key to ubiquitin transfer. *Trends Biochem. Sci.* 37, 58–65. doi: 10.1016/j.tibs.2011.11.001
- Capra, J. A., Williams, A. G., and Pollard, K. S. (2012). ProteinHistorian: tools for the comparative analysis of eukaryote protein origin. *PLoS Comput. Biol.* 8:e1002567. doi: 10.1371/journal.pcbi.1002567
- Carra, S., Alberti, S., Arrigo, P. A., Benesch, J. L., Benjamin, I. J., and Boelens, W. (2017). The growing world of small heat shock proteins: from structure to functions. *Cell Stress Chaperones* 22, 601–611. doi: 10.1007/s12192-017-0787-8
- Cavalier-Smith, T. (2006). Cell evolution and Earth history: stasis and revolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 969–1006. doi: 10.1098/rstb.2006.1842
- Chaix, A., Zarrinpar, A., and Panda, S. (2016). The circadian coordination of cell biology. *J. Cell Biol.* 215, 15–25. doi: 10.1083/jcb.201603076
- Chen, W. H., Lu, G., Chen, X., Zhao, X. M., and Bork, P. (2017). OGEE v2: an update of the online gene essentiality database with special focus on differentially essential genes in human cancer cell lines. *Nucleic Acids Res.* 45, D940–D944. doi: 10.1093/nar/gkw1013
- Chiou, Y. Y., Yang, Y., Rashid, N., Ye, R., Selby, C. P., and Sancar, A. (2016). Mammalian Period represses and de-represses transcription by displacing CLOCK-BMAL1 from promoters in a Cryptochrome-dependent manner. *Proc. Natl. Acad. Sci. U.S.A.* 113, E6072–E6079. doi: 10.1073/pnas.1612917113
- Conway de Macario, E., Robb, F. T., and Macario, A. J. (2017). Prokaryotic chaperonins as experimental models for elucidating structure-function abnormalities of human pathogenic mutant counterparts. *Front. Mol. Biosci.* 3:84. doi: 10.3389/fmolb.2016.00084
- Cretenet, G., Le Clech, M., and Gachon, F. (2010). Circadian clock-coordinated 12 Hr period rhythmic activation of the IRE1alpha pathway controls lipid metabolism in mouse liver. *Cell Metab.* 11, 47–57. doi: 10.1016/j.cmet.2009.11.002
- Crowe, S. A., Dössing, L. N., Beukes, N. J., Bau, M., Kruger, S. J., Frei, R., et al. (2013). Atmospheric oxygenation three billion years ago. *Nature* 501, 535–538. doi: 10.1038/nature12426
- Dodd, M. S., Papineau, D., Grenne, T., Slack, J. F., Rittner, M., Pirajno, F., et al. (2017). Evidence for early life in Earth's oldest hydrothermal vent precipitates. *Nature* 543, 60–64. doi: 10.1038/nature21377
- Dunlap, J. C. (1999). Molecular bases for circadian clocks. *Cell* 96, 271–290. doi: 10.1016/S0092-8674(00)80566-8
- Fu, J., Murphy, K. A., Zhou, M., Li, Y. H., Lam, V. H., Tabuloc, C. A., et al. (2016). Codon usage affects the structure and function of the *Drosophila* circadian clock protein PERIOD. *Genes Dev.* 30, 1761–1775. doi: 10.1101/gad.281030.116
- Fuhr, L., Abreu, M., Pett, P., and Relógio, A. (2015). Circadian systems biology: when time matters. *Comput. Struct. Biotechnol. J.* 13, 417–426. doi: 10.1016/j.csbj.2015.07.001
- Gold, D. A., Caron, A., Fournier, G. P., and Summons, R. E. (2017). Paleoproterozoic sterol biosynthesis and the rise of oxygen. *Nature* 543, 420–423. doi: 10.1038/nature21412
- Hamilton, T. L., Bryant, D. A., and Macalady, J. L. (2016). The role of biology in planetary evolution: cyanobacterial primary production in low-oxygen Proterozoic oceans. *Environ. Microbiol.* 18, 325–340. doi: 10.1111/1462-2920.13118
- Hollien, J. (2013). Evolution of the unfolded protein response. *Biochim. Biophys. Acta* 1833, 2458–2463. doi: 10.1016/j.bbamcr.2013.01.016
- Huerta-Cepas, J., Capella-Gutiérrez, S., Pryszcz, L. P., Marcet-Houben, M., and Gabaldón, T. (2014). PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome. *Nucleic Acids Res.* 42, D897–D902. doi: 10.1093/nar/gkt1177
- Hughes, M. E., Di Tacchio, L., Hayes, K. R., Vollmers, C., Pulivarthy, S., Baggs, J. E., et al. (2009). Harmonics of circadian gene transcription in mammals. *PLoS Genet.* 5:e1000442. doi: 10.1371/journal.pgen.1000442
- Iwata, Y., and Koizumi, N. (2012). Plant transducers of the endoplasmic reticulum unfolded protein response. *Trends Plant Sci.* 17, 720–727. doi: 10.1016/j.tplants.2012.06.014
- Kriebs, A., Jordan, S. D., Soto, E., Henriksson, E., Sandate, C. R., Vaughan, M. E., et al. (2017). Circadian repressors CRY1 and CRY2 broadly interact with nuclear receptors and modulate transcriptional activity. *Proc. Natl. Acad. Sci. U.S.A.* 114, 8776–8781. doi: 10.1073/pnas.1704955114
- Lacher, S. E., Lee, J. S., Wang, X., Campbell, M. R., Bell, D. A., and Slattery, M. (2015). Beyond antioxidant genes in the ancient Nrf2 regulatory network. *Free Radic. Biol. Med.* 88, 452–465. doi: 10.1016/j.freeradbiomed.2015.06.044
- Lehmann, R., Childs, L., Thomas, P., Abreu, M., Fuhr, L., Herzel, H., et al. (2015). Assembly of a comprehensive regulatory network for the mammalian circadian clock: a bioinformatics approach. *PLoS One* 10:e0126283. doi: 10.1371/journal.pone.0126283
- Luo, G., Ono, S., Beukes, N. J., Wang, D. T., Xie, S., and Summons, R. E. (2016). Rapid oxygenation of Earth's atmosphere 2.33 billion years ago. *Sci. Adv.* 2:e1600134. doi: 10.1126/sciadv.1600134
- Ma, L., Cui, P., Zhu, J., Zhang, Z., and Zhang, Z. (2014). Translational selection in human: more pronounced in housekeeping genes. *Biol. Direct* 9:17. doi: 10.1186/1745-6150-9-17
- Martins, B. M., Das, A. K., Antunes, L., and Locke, J. C. (2016). Frequency doubling in the cyanobacterial circadian clock. *Mol. Syst. Biol.* 12:896. doi: 10.15252/msb.20167087
- Mazzoccoli, G., Miscio, G., Fontana, A., Copetti, M., Francavilla, M., Bosi, A., et al. (2016). Time related variations in stem cell harvesting of umbilical cord blood. *Sci. Rep.* 6:21404. doi: 10.1038/srep21404
- Ncbi Resource Coordinators. (2017). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 45, D12–D17. doi: 10.1093/nar/gkw1071
- Neumann, C. A., Krause, D. S., Carman, C. V., Das, S., Dubey, D. P., Abraham, J. L., et al. (2003). Essential role for the peroxiredoxin Prdx1 in erythrocyte antioxidant defence and tumour suppression. *Nature* 424, 561–565. doi: 10.1038/nature01819
- Oda, M. N., Bielicki, J. K., Berger, T., and Forte, T. M. (2001). Cysteine substitutions in apolipoprotein A-I primary structure modulate paraoxonase activity. *Biochemistry* 40, 1710–1718. doi: 10.1021/bi001922h
- Parnell, J., Boyce, A. J., Mark, D., Bowden, S., and Spinks, S. (2010). Early oxygenation of the terrestrial environment during the Mesoproterozoic. *Nature* 468, 290–293. doi: 10.1038/nature09538
- Pittendrigh, C. S. (1993). Temporal organization: reflections of a Darwinian clock-watcher. *Annu. Rev. Physiol.* 55, 16–54. doi: 10.1146/annurev.ph.55.030193.000313
- Richter, K., Haslbeck, M., and Buchner, J. (2010). The heat shock response: life on the verge of death. *Mol. Cell* 40, 253–266. doi: 10.1016/j.molcel.2010.10.006
- Sancar, C., Sancar, G., Ha, N., Cesbron, F., and Brunner, M. (2015). Dawn- and dusk-phased circadian transcription rhythms coordinate anabolic and catabolic functions in *Neurospora*. *BMC Biol.* 13:17. doi: 10.1186/s12915-015-0126-4
- Sharp, P. M., and Li, W. H. (1987). The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15, 1281–1295. doi: 10.1093/nar/15.3.1281
- Sonett, C. P., Kvale, E. P., Zakharian, A., Chan, M. A., and Demko, T. M. (1996). Late proterozoic and paleozoic tides, retreat of the moon, and

- rotation of the Earth. *Science* 273, 100–104. doi: 10.1126/science.273.5271.100
- Takahashi, J. S. (2017). Transcriptional architecture of the mammalian circadian clock. *Nat. Rev. Genet.* 18, 164–179. doi: 10.1038/nrg.2016.150
- Trott, A. J., and Menet, J. S. (2018). Regulation of circadian clock transcriptional output by clock:BMAL1. *PLoS Genet.* 14:e1007156. doi: 10.1371/journal.pgen.1007156
- van der Veen, D. R., and Gerkema, M. P. (2017). Unmasking ultradian rhythms in gene expression. *FASEB J.* 31, 743–750. doi: 10.1096/fj.201600872R
- Vicario, S., Moriyama, E. N., and Powell, J. R. (2007). Codon usage in twelve species of *Drosophila*. *BMC Evol. Biol.* 7:226. doi: 10.1186/1471-2148-7-226
- Westermarck, P. O., and Herzog, H. (2013). Mechanism for 12 hr rhythm generation by the circadian clock. *Cell Rep.* 3, 1228–1238. doi: 10.1016/j.celrep.2013.03.013
- Williams, G. E. (2000). Geological constraints on the Precambrian history of Earth's rotation and the Moon's orbit. *Rev. Geophys.* 38, 37–59. doi: 10.1029/1999RG900016
- Wright, F. (1990). The 'effective number of codons' used in a gene. *Gene* 87, 23–29. doi: 10.1016/0378-1119(90)90491-9
- Xu, Y., Ma, P., Shah, P., Rokas, A., Liu, Y., and Johnson, C. H. (2013). Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. *Nature* 495, 116–120. doi: 10.1038/nature11942
- Yin, H., Wang, G., Ma, L., Yi, S. V., and Zhang, Z. (2016). What signatures dominantly associate with gene age? *Genome Biol. Evol.* 8, 3083–3089.
- Yu, C. H., Dang, Y., Zhou, Z., Wu, C., Zhao, F., Sachs, M. S., et al. (2015). Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Mol. Cell* 59, 744–754. doi: 10.1016/j.molcel.2015.07.018
- Zahnle, K., and Walker, J. C. (1987). A constant day length during the precambrian era? *Precambrian Res.* 37, 95–105. doi: 10.1016/0301-9268(87)90073-8
- Zámocký, M., Gasselhuber, B., Furtmüller, P. G., and Obinger, C. (2012). Molecular evolution of hydrogen peroxide degrading enzymes. *Arch. Biochem. Biophys.* 525, 131–144. doi: 10.1016/j.abb.2012.01.017
- Zhang, L., Zhang, C., and Wang, A. (2016). Divergence and conservation of the major UPR Branch IRE1-bZIP signaling pathway across Eukaryotes. *Sci. Rep.* 6:27362. doi: 10.1038/srep27362
- Zhang, R., Lahens, N. F., Ballance, H. I., Hughes, M. E., and Hogenesch, J. B. (2014). A circadian gene expression atlas in mammals: implications for biology and medicine. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16219–16224. doi: 10.1073/pnas.1408886111
- Zhou, M., Guo, J., Cha, J., Chae, M., Chen, S., Barral, J. M., et al. (2013). Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature* 495, 111–115. doi: 10.1038/nature11833
- Zhou, M., Wang, T., Fu, J., Xiao, G., and Liu, Y. (2015). Nonoptimal codon usage influences protein structure in intrinsically disordered regions. *Mol. Microbiol.* 97, 974–987. doi: 10.1111/mmi.13079
- Zhu, B., Zhang, Q., Pan, Y., Mace, E. M., York, B., Antoulas, A. C., et al. (2017). A cell-autonomous mammalian 12 hr clock coordinates metabolic and stress rhythms. *Cell Metab.* 25, 1305.e9–1319.e9. doi: 10.1016/j.cmet.2017.05.004

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Castellana, Mazza, Capocefalo, Genov, Biagini, Fusilli, Scholkmann, Relógio, Hogenesch and Mazzoccoli. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.