

## RESEARCH ARTICLE

# Rhythmicity is linked to expression cost at the protein level but to expression precision at the mRNA level

David Laloum<sup>1,2</sup>, Marc Robinson-Rechavi<sup>1,2\*</sup>

**1** Department of Ecology and Evolution, Batiment Biophore, Quartier UNIL-Sorge, Université de Lausanne, Lausanne, Switzerland, **2** Swiss Institute of Bioinformatics, Batiment Génopode, Quartier UNIL-Sorge, Université de Lausanne, Lausanne, Switzerland

\* [marc.robinson-rechavi@unil.ch](mailto:marc.robinson-rechavi@unil.ch)**OPEN ACCESS**

**Citation:** Laloum D, Robinson-Rechavi M (2022) Rhythmicity is linked to expression cost at the protein level but to expression precision at the mRNA level. *PLoS Comput Biol* 18(9): e1010399. <https://doi.org/10.1371/journal.pcbi.1010399>

**Editor:** Guang-Zhong Wang, Partner Institute for Computational Biology Chinese Academy of Sciences and Max Planck Society, CHINA

**Received:** December 8, 2021

**Accepted:** July 17, 2022

**Published:** September 12, 2022

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1010399>

**Copyright:** © 2022 Laloum, Robinson-Rechavi. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data and scripts are available at: [https://github.com/laloumdav/cost\\_noise\\_conservation\\_rhythmicity](https://github.com/laloumdav/cost_noise_conservation_rhythmicity).

## Abstract

Many genes have nycthemeral rhythms of expression, *i.e.* a 24-hours periodic variation, at either mRNA or protein level or both, and most rhythmic genes are tissue-specific. Here, we investigate and discuss the evolutionary origins of rhythms in gene expression. Our results suggest that rhythmicity of protein expression could have been favored by selection to minimize costs. Trends are consistent in bacteria, plants and animals, and are also supported by tissue-specific patterns in mouse. Unlike for protein level, cost cannot explain rhythm at the RNA level. We suggest that instead it allows to periodically reduce expression noise. Noise control had the strongest support in mouse, with limited evidence in other species. We have also found that genes under stronger purifying selection are rhythmically expressed at the mRNA level, and we propose that this is because they are noise sensitive genes. Finally, the adaptive role of rhythmic expression is supported by rhythmic genes being highly expressed yet tissue-specific. This provides a good evolutionary explanation for the observation that nycthemeral rhythms are often tissue-specific.

## Author summary

For many genes, their expression, *i.e.* the production of RNA and proteins, is rhythmic with a 24-hour period. Here, we study and discuss the evolutionary origins of these rhythms. Our analyses of data from different species suggest that the rhythmicity of protein level may have been favored by selection for cost minimization. Furthermore, we have shown that cost cannot explain the rhythmic variations in RNA levels. Instead, we suggest that it periodically reduces the stochasticity of gene expression. We also found that genes under stronger purifying selection are rhythmically expressed at the mRNA level, and propose that this is because they are noise-sensitive genes. Finally, rhythmic expression involves genes that are often highly expressed and tissue-specific. This provides a good evolutionary explanation for the tissue-specificity of these rhythms.

**Funding:** MRR received grant 31003A\_173048 of Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (SNF) <https://www.snf.ch/en>. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Living organisms have to adapt to complex and changing environments. Physiological systems able to accommodate themselves to changing but predictable circumstances are expected to have a higher stability of survival and reproduction [1]. The strongest predictable change for most organisms is the light/dark cycle and the associated nycthemeral temperature variations. In an organism, “circadian rhythms” denote entities characterized by an endogenous and entrainable oscillator clock which is able to persist in constant conditions (such as in constant darkness) and whose phases can be altered (reset or entrained). However, many physiological systems display non-autonomous nycthemeral rhythms, directly or indirectly controlled by the local clock or mainly by the environment itself, or by both [2–6]. Such rhythms are found at all levels: molecular, cellular, organs, and behavioural, and several regulatory networks appear to play roles in the synchronization of these levels [7].

Here, we studied some of the evolutionary costs and benefits that shape the rhythmic nature of gene expression at the RNA and protein levels. For this, we analysed several characteristics that we expect to determine the rhythmic nature of gene expression in a trade-off between its advantages (economy of energetic costs over 24h, ribosomal non-occupancy) and disadvantages (costs of complexity due to precise temporal regulation). We have not studied here the potential negative effect of proteins when they are not needed, such as mis-interaction of proteins, or unwanted enzymatic activity. We call “rhythmic genes” all genes displaying a 24-hours periodic variation of their mRNA or protein level, or of both, constituting the nycthemeral transcriptome or proteome. Their rhythmic expression can be entrained directly by the internal clock but also directly or indirectly by external inputs, such as the light-dark cycle or food-intake [2–6]. Hence we use the term “nycthemeral” to avoid confusion with the specific features of “circadian” rhythms, although these are included in the nycthemeral rhythms, which includes downstream regulated genes. Because the alternation of light and dark can be considered as a permanent signal for most life on earth [8], we consider that the entirety of nycthemeral biological rhythms are relevant as a phenotype under selection.

## Rhythmic gene expression: An adaptation in cycling environments

The endogenous nature of circadian rhythms is a strategy of anticipation. The ability to anticipate improves the adaptation of organisms to their fluctuating environment. Most rhythmic genes are tissue-specific [9–11], *i.e.* a given gene can be rhythmic in some tissues, and constantly or not expressed in others, which means that their rhythmic regulation is not a general property of the gene and is therefore expected to be advantageous only in those tissues in which they are found rhythmic. This argues that rhythmic regulation has costs, since it is not general. These costs are probably related to the complexity of regulation to maintain precise temporal organisation, and could also be related to toxicity. Thus, cyclic biological systems are expected to have adaptive origins. Gene expression is costly for the cell in terms of energy and cellular materials usage. Wang et al. [12] have shown that in the liver of mice, abundant proteins that are required at one time are down-regulated at other times, apparently to economize on overall production. Wang et al. also reported that at each time-point, the total metabolic cost was 4 fold higher for the set of cycling genes compared to the non-cycling genes set at both transcriptional and translational levels [12]—although the proteomic data used from mouse fibroblasts appear to have been underestimated and have since been corrected [13]. Furthermore, we have shown in our previous work that rhythmic genes are largely enriched in highly expressed genes and that the differences in rhythm detection obtained between highly and lowly expressed genes either reflect true biology or a lower signal to noise ratio in lowly expressed genes [14]. Here, we present results supporting the hypothesis that cyclic expression

of highly expressed proteins were selected as part of a “low-cost” strategy to minimize the overall use of cellular energy. Thus, a first evolutionary advantage given by rhythmic biological processes would be a reduction of the overall cost (over a 24-hour period), compared to the costs generated over the same period by a sufficient constant level of proteins. Furthermore, we provide a first explanation for the tissue-specificity of rhythms in gene expression by showing that genes are more likely to be rhythmic in tissues where they are specifically highly expressed.

### Noise and cost reduction

Expression costs at the protein level are at least 160 times higher than at the RNA level (section 3 in [S1 File](#)). Thus, it requires considerably more metabolic activity to produce a significant change in protein than in transcript levels. Moreover, small increases in protein expression levels have been shown to incur energy costs large enough to be opposed by natural selection, at least in bacteria [15, 16]. Consequently, protein synthesis costs are more constraining than RNA synthesis costs, especially in eukaryotes for which translation is the major contribution to making expression costs visible to selection [16].

Many transcripts show nycthemeral fluctuations without rhythmicity of their protein abundances, even when measured in the same study. This is the case for instance in mouse liver [17] or in plants [18]. Yet relative to protein synthesis costs, costs at RNA level are probably too small to provide a satisfactory evolutionary explanation of rhythms at mRNA level.

Thattai et al. [19] and Hausser et al. [20] propose that living systems have made tradeoffs between energy efficiency and noise reduction. Indeed, the control of noise (stability against fluctuations) plays a key role in the function of biological systems (*i.e.* in the robustness of gene expression) as supported by observations of a noise reduction during key periods, *e.g.* in development [19, 21]. These considerations lead to predictions which we study here: i) a strategy to periodically decrease stochasticity for genes with rhythmically accumulated mRNAs; ii) a cost-saving strategy for genes whose protein expression is rhythmic; and iii) a combined strategy for genes rhythmic at both levels.

## Results

### Cyclicity of costliest genes

Expression costs at the protein level are at least 160 times higher than at the RNA level (section 3 in [S1 File](#)). They are dominated by the cost of translation [15, 16]. This is mainly due to the higher abundance of proteins relative to transcripts, by a factor of 1000 (section 3 in [S1 File](#)). Thus, based on the calculation developed by Lynch and Marinov [16], we estimated the expression cost ( $C_p$ ) of each gene by the formula (1) which takes into account the averaged amino-acid (AA) synthesis cost ( $\bar{c}_{AA}$ ) of the protein ([S1 Table](#)), protein length ( $L_p$ ), and protein abundance ( $N_p$ ).

$$C_p = N_p L_p \bar{c}_{AA} \quad (1)$$

Other costs such as AA polymerization or protein decay are not expected to change our results, since longer or higher expressed proteins need more chain elongation activity and degradation (section 4 in [S1 File](#)). We show in [S1 File](#) (section 4) that the costs of maintaining proteins is higher for highly expressed than for lowly expressed proteins, even taking into account that highly expressed proteins have longer half-lives (also see [Discussion](#)). Thus, the cost provided by formula (1) is representative of gene expression costs and can be used for comparison between genes. To compare expression costs between rhythmic and non-rhythmic proteins, we calculated the average and the maximum protein expression level over time-points (see

Methods, Eqs (2) and (3) (Fig 1d). The AA biosynthesis costs estimated in *E. coli* (S1 Table) were used as representative for all species since biosynthetic pathways are nearly universally conserved [15].

If the biological function of a protein is periodic, we expect the rhythmic regulation of its expression to be determined by a trade-off between the benefits of not producing proteins when they are not needed (energy costs saved, as we did not include the potential toxicity of unneeded proteins) and the potential costs involved in making it rhythmic and temporally coordinated (costs of complexity) (Fig 1b).

This leads to expect costlier genes to be more frequently rhythmic. First, we confirm that cycling genes are indeed enriched in highly expressed genes, as previously reported [12, 14]. The higher cost of cycling genes (Fig 2a and S2 Fig) was especially correlated with their higher expression levels, observed both at the protein level (Fig 2d and 2e) and at the RNA level (S1 Fig). We find rhythmic proteins to be longer only in mouse and in cyanobacteria (Fig 2c). However, conserved genes are known to be longer [22, 23] and are often more expressed [24, 25], therefore more expensive to produce, which could explain the rhythmic expression we observe for longer proteins. On the other hand, the higher cost for rhythmic proteins in *Synechococcus elongatus* appears only due to their length. Finally, in mouse liver, cycling proteins seem to contain more expensive amino acids than non-rhythmic proteins (Fig 2b).

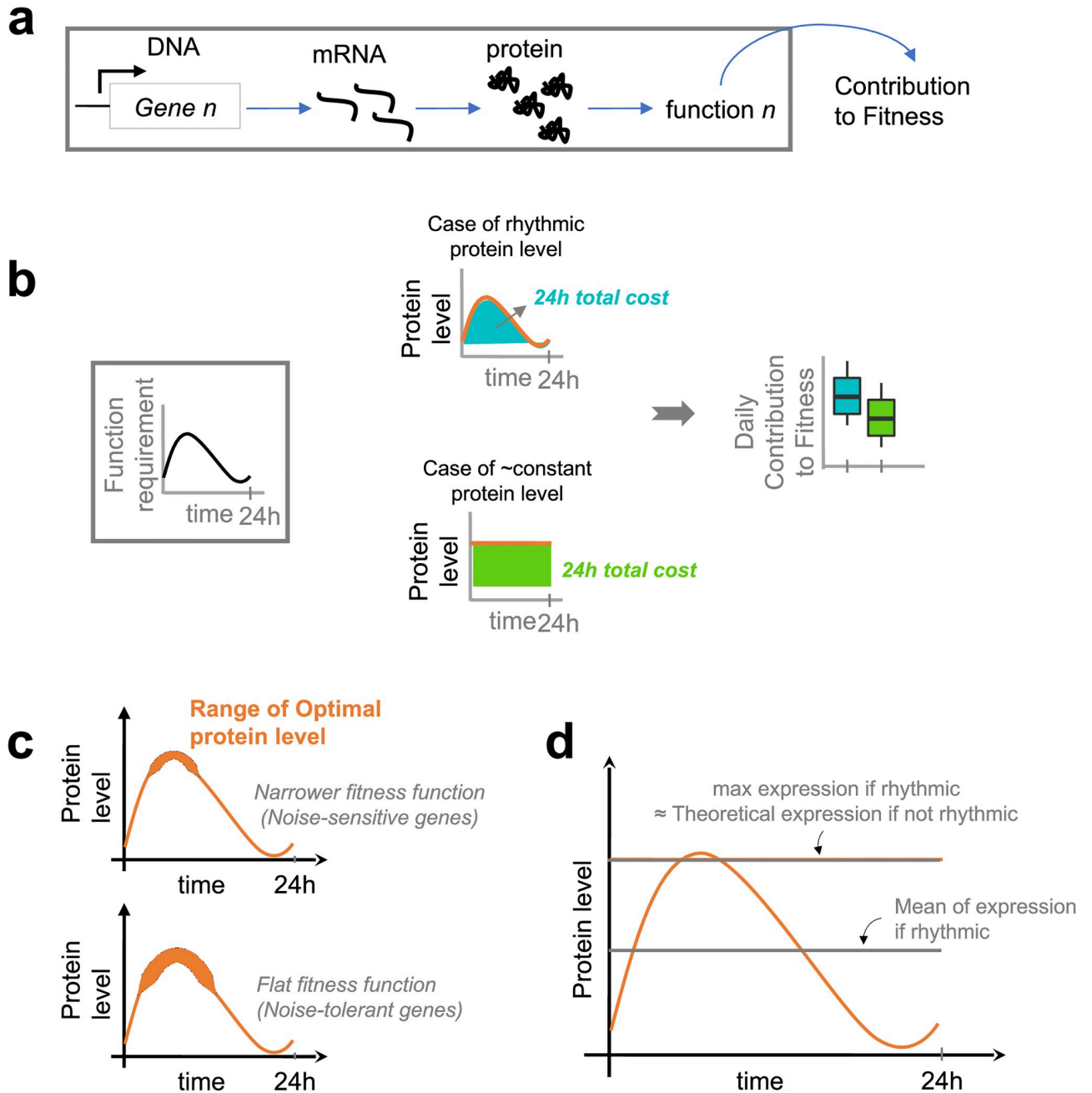
Overall, results are consistent with the expectation that costlier genes are preferentially under rhythmic regulation. Furthermore, our results show that while rhythmicity at protein level lowers costs, it does not drive the costs of rhythmic genes down to the level of constant proteins. A more specific prediction from this hypothesis is that each gene should be found rhythmic specifically in tissues where a high expression of that gene is needed, *i.e.* where its function is costlier.

### Genes tend to be rhythmic in the tissues in which they are highly expressed

To test this hypothesis, we used a mouse circadian dataset with 11 tissues (transcriptomics from Zhang et al. [9], S2 Table). For each gene we separated the tissues into two groups, those for which the gene was rhythmic (Rhythmic detection method, GeneCycle:  $p$ -value  $\leq 0.01$ ) and those for which it was not (Rhythmic detection method, GeneCycle:  $p$ -value  $> 0.5$ ). Because of the difficulty of setting reliable thresholds for rhythmicity [14], we ignored intermediate  $p$ -values. For each gene, we estimated the difference  $\delta$  of expression levels between these two groups of tissues (see Formula 7 in Methods). As expected, genes tend to be more highly expressed in tissues where they are rhythmic (Student's test, testing the hypothesis that the  $\delta$  distribution mean is equal to 0:  $\delta_{mean} = 0.0146$ ,  $t = 28.29$ ,  $df = 11683$ ,  $p$ -value  $< 2.2e-16$ , 95% CI [0.0136, 0.0156]). We also provide results obtained from other datasets in S3 Table, although they must be taken with caution since only 2 to 4 tissues were available, and sometimes data were coming from different experiments. Of note, for proteomic data, the distributions of  $\delta$  are bimodal (S3 Fig), separating rhythmic proteins into two groups, with low or high protein levels in the tissues in which they are rhythmic. A hypothesis is that for some tissue-specific proteins the rhythmic regulation is not tissue-specific, making them rhythmic also in tissues where they are lowly expressed. But the very small sample size does not allow us to test it, and we caution against any over-interpretation of this pattern before it can be confirmed.

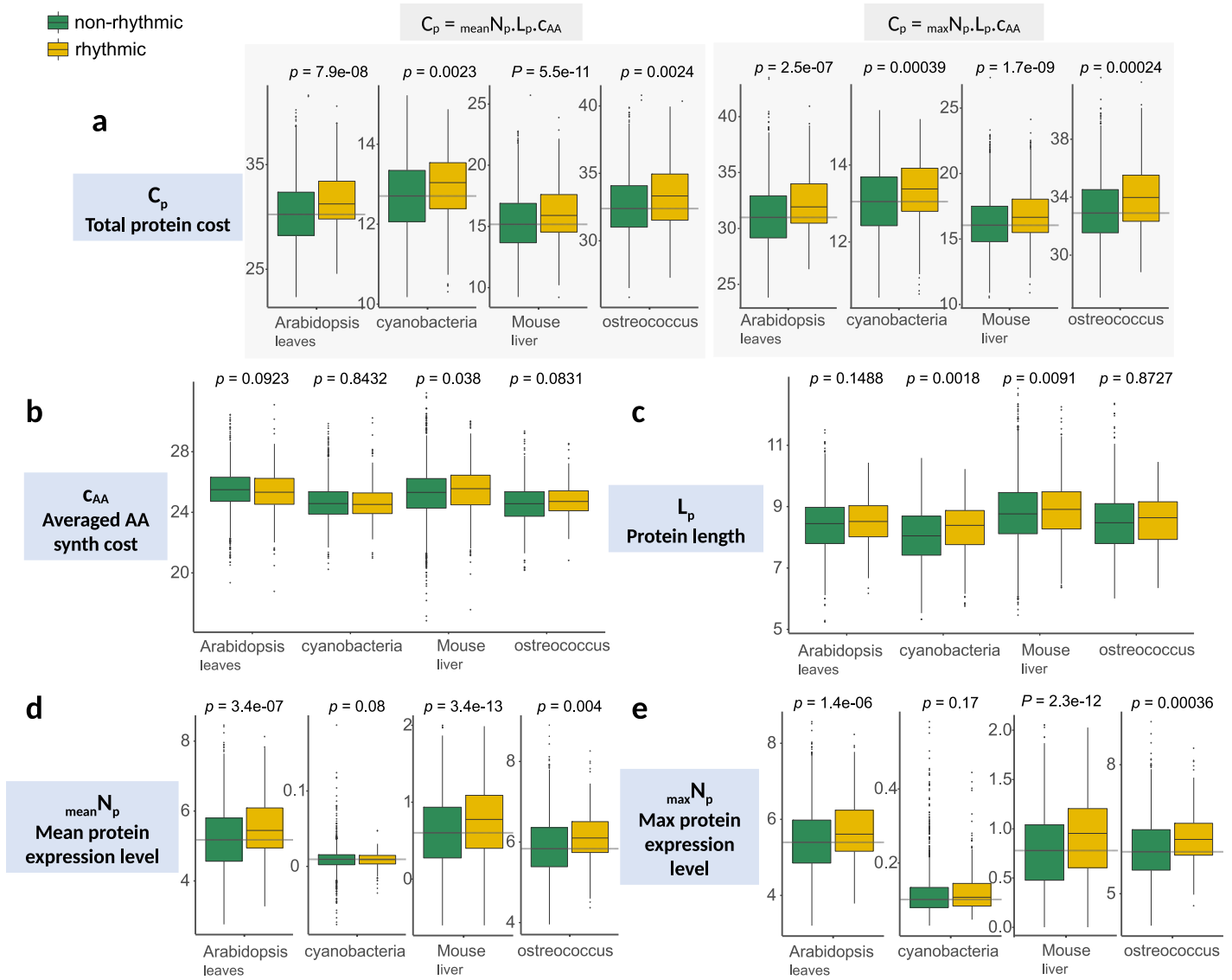
### Rhythmic genes are tissue-specific

To clarify whether rhythmic genes tend to be tissue-specific highly expressed genes, we also analysed the relation between the number of tissues in which a gene is rhythmic and its tissue-specificity of expression  $\tau$  [26, 27]. For the mouse circadian dataset (transcriptomic) with 11



**Fig 1. a)** Gene expression contributes to organismal fitness. **b)** Rhythmic protein regulation presents a trade-off between the costs generated by integration into the rhythmic system (costs of complexity) and the advantages provided plus the costs saved over 24 hours. The middle exemplifies two extreme behaviors, while the right shows the distribution expected from populations of genes which follow these behaviors. **c)** The range of high fitness protein levels depends on the sensitivity of the function to deviations from an optimal level. We use the term “narrower” following Hausser et al. [20]. Noise sensitive genes have narrower fitness function, *i.e.* a small deviation from the optimum rapidly decreases the contribution to fitness. Precision is less important for genes with flat fitness functions. **d)** Mean or maximum expression level calculated from time-series datasets (see [Methods](#)). We assume that, in the absence of rhythmic regulation, the constant optimal level is included between the mean and the maximum expression level observed in rhythmic expression.

<https://doi.org/10.1371/journal.pcbi.1010399.g001>



**Fig 2. Among the factors of expression costs, expression level is the main factor explaining the higher cost observed in rhythmic proteins.** a) The total cost of rhythmic proteins is higher than those of other proteins. b) With the exception of mouse liver, rhythmic proteins do not contain more expensive amino-acids than other proteins. c) Rhythmic proteins can be longer in some species. d-e) Mean or maximum expression level calculated from time-series datasets: rhythmic proteins are highly expressed proteins. Boxplots are log scaled except for the averaged AA synthesis cost. The first 15% of proteins from *p*-values ranking (from the most rhythmic to the most un-rhythmic genes) obtained from the rhythm detection algorithms were classified as rhythmic.

<https://doi.org/10.1371/journal.pcbi.1010399.g002>

tissues, partial correlations show that genes whose rhythmicity is tissue-specific have tissue-specific expression ( $\tau$ , Formula 8, versus number of rhythmic tissues: Pearson’s correlation = -0.37,  $t = -5.3e+01$ ,  $p < 2.2e^{-16}$ ; see S4 Table for Spearman’s correlation test). Results obtained with other datasets are in S4 Table; results for the baboon dataset should be taken with caution for reasons discussed in our previous work [14]. The partial correlation seems to be stronger at the transcriptional level, although data was available in much less tissues at the protein level (only mouse forebrain, cartilage, liver, and tendon). Moreover, for tissue-specific genes ( $\tau > 0.5$ ), the signal of rhythmicity correlates with expression level over tissues (S4(a) Fig). While Spearman’s correlation is clearly skewed towards negative correlations, *i.e.* lower *p*-values thus

stronger signal of rhythmicity in the tissue where genes are more expressed, Pearson's correlation also has a smaller peak of positive correlations (S4(a) Fig), suggesting a subset of genes which are less rhythmic in the tissues where they are most expressed. We show that tissue-specific genes which are mostly rhythmic in tissues where they are highly expressed are under stronger selective constraint than those which are rhythmic in tissues where they are lowly expressed (S4(b) Fig). Thus, rhythmic expression of this second set of genes might be under weaker constraints. The dominant signal overall is that in a given tissue, rhythmically expressed genes tend to be those which are tissue-specifically expressed in this tissue.

### Lower cell-to-cell variability of genes with rhythmic transcripts

Increasing transcription for a fixed amount of protein can decrease the noise in final protein levels [20], thus, genes with rhythmic mRNAs can have lower noise at their mRNA level peak. We predict that genes with rhythmic mRNAs are noise-sensitive genes (Fig 1c). To test this, we compared the noise distribution between rhythmic and non-rhythmic genes. We used  $F^*$  of Barroso et al. [28] to control for the correlation between expression mean ( $\mu$ ) and variance ( $\sigma^2$ ). It was the most efficient compared with other methods (section 6.4 in S1 File). To evaluate expression noise, we used single-cell RNA data from *Mus musculus* liver, lung, limb muscle, heart, and aorta [29] and *Arabidopsis thaliana* roots [30] (S2 Table). These data are from a single time-point each time. In the absence of other data, we have to assume that this noise calculated at the transcriptional level is representative of noise at the protein level, or in any case functionally relevant. If anything, we expect this to provide an under-estimation of any relation between noise and rhythmicity, since noise at other time points is missed. Then, we assessed rhythmicity based on time-series datasets: RNA and proteins from Arabidopsis leaves and mouse liver from the data used above; and RNA only in mouse lung, kidney, muscle, heart, and aorta (S2 Table). Results are given in a simplified way in Table 1 for the mouse; full results, which include Arabidopsis, are provided in S5 Table. In most cases, we found lower noise for genes with rhythmic mRNA (row a of Table 1 and S5 Table).

We obtained the same results using less stringent cut-offs when calling genes rhythmic ( $p < 0.05$ , results not shown). Genes with both rhythmic proteins and mRNA had lower noise (row c of Table 1), although differences weren't significant. Our results show that noise estimated at a single time-point is globally reduced for genes with rhythmic regulation at the transcriptional level. Since rhythmic genes are not all in the same phase (Fig 2a and S1 File), we expect this result obtained for a given time-point (average noise estimated at a single time-point from scRNA) to be general to all time-points (section 6.3 in S1 File). Assuming that genes with low noise are noise-sensitive (and thus noise is tightly controlled), these results suggest that rhythmic genes have, individually, their noise periodically and drastically reduced through periodic high accumulation of their mRNAs. In Arabidopsis, the single-cell data used are from the root, while transcriptomic time-series data used to detect rhythmicity are from the leaves, which limits the interpretation. Despite this limitation, we found no evidence of lower noise for genes that are rhythmic at the protein level (rows b and e of Table 1, and S5 Table), and trends towards lower noise in almost all cases for genes with rhythmic mRNAs (rows a, c, and d of Table 1).

### Genes with rhythmic transcripts are under stronger selection

Finally, we compared protein evolutionary conservation (estimated by the dN/dS ratio) between rhythmic and non-rhythmic genes. Results are given in a simplified way in Table 2; full results are provided in S6 Table. In all cases, in plants, vertebrates, and insects, we found that genes with rhythmic mRNA levels were significantly more conserved (triangles in S8

**Table 1. Simplified table showing the results of the Welch two sample t-test testing the hypothesis that the noise is equal between rhythmic versus non-rhythmic transcripts (a), or proteins (b), or between rhythmic versus non-rhythmic transcripts among rhythmic proteins (c) or among non-rhythmic proteins (d), and between rhythmic versus non-rhythmic proteins among genes with constant transcripts (e).  $F^*$  is an estimation of the noise based on Barroso et al. method [28]. Complete results are provided in S5 Table.**

		<i>Mus musculus</i>					
		Liver	Lung	Muscle	Heart	Kidney	Aorta
a	All genes: rRNA VS nrRNA						
b	All genes: rProtein VS nrProtein						
c	rProteins ( $p \leq 0.05$ ): rRNA VS nrRNA						
d	nrProteins ( $p > 0.2$ ): rRNA VS nrRNA						
e	nrRNA ( $p > 0.2$ ): rProtein VS nrProtein						

Noise r < Noise nr ;  $p < 0.05$   
 Noise r < Noise nr ;  $p \geq 0.05$   
 Noise r > Noise nr ;  $p < 0.05$   
 Noise r > Noise nr ;  $p \geq 0.05$

Welch two sample t-test

r: rhythmic  
nr: non-rhythmic

<https://doi.org/10.1371/journal.pcbi.1010399.t001>

Table). We obtained similar results after controlling for gene expression bias (Table 2 and background boxes in S8 Table), since higher expressed genes are known to be more conserved genes. Interestingly, we didn't find a clear tendency for higher conservation in rhythmic proteins (Table 2). In Arabidopsis, genes rhythmic at both levels were less conserved than genes rhythmic at only one level (Table 2), and results were unclear for mouse. These results suggest that rhythmic expression at the transcriptional level plays an important role for genes under strong purifying selection, consistent with the results for tissue-specific rhythmic genes.

### Discussion


In this work, we investigate the evolutionary origins of rhythms in gene expression that respond to factors oscillating on a 24-hour time scale. This scale is longer than those of many biological oscillators at the molecular level, which are ultradian; for instance ~3h for the p53-Mdm2 system, or on the order of minutes to seconds for calcium oscillations. Hence, natural selection was probably instrumental in reaching the longer nycthemeral period of gene expression for many genes.


The endogenous generation of circadian rhythms is an anticipation strategy, which is optimized if the internal clock resonates with the external cycle [31, 32]. The autonomous nature of such mechanisms provides a clear advantage to the organism able to anticipate its environmental changes before they take place, allowing it to be "ready" before organisms who would not be endowed with such capacity. In terms of adaptability in a given environment, if we consider nycthemeral rhythms without considering their endogenous or exogenous nature, one can ask the question of the evolutionary origin of maintaining large cyclic biological systems, with gene and tissue specific regulation. Indeed, they are very widely found among living organisms, at every level of biological systems, highlighting their necessity for adapted phenotypes.





**Table 2. Simplified table showing the results of the Welch two sample t-test testing the hypothesis that dN/dS ratio is equal between rhythmic versus non-rhythmic transcripts (a), or proteins (b), or between rhythmic versus non-rhythmic transcripts among rhythmic proteins (c), and between rhythmic versus non-rhythmic proteins among genes with rhythmic transcripts (d). In S8 Table, triangles give the result of the Welch two sample t-test without controlling for the effect of gene expression level. Complete results are provided in S6 Table.**

			All genes: rRNA vs. nrRNA	All genes: rProtein vs. nrProtein	rProteins ( $p \leq 0.05$ ): rRNA vs. nrRNA	rRNA ( $p \leq 0.01$ ): rProtein vs. nrProtein
PLANT	Arabidopsis	Leaves				
VERTEBRATES	Mouse	Liver				
		Lung				
		Muscle				
		Heart				
		Kidney				
		Aorta				
		Tendon				
		Forebrain				
		Cartilage				
	Rat	Lung				
INSECTS	Anopheles	Head				
		Body				
	Aedes	Head				

 dN/dS of rhythmic genes < dN/dS of non-rhythmic genes;  $p < 0.05$

 dN/dS of rhythmic genes < dN/dS of non-rhythmic genes;  $p \geq 0.05$

 dN/dS of rhythmic genes > dN/dS of non-rhythmic genes;  $p < 0.05$

 dN/dS of rhythmic genes > dN/dS of non-rhythmic genes;  $p \geq 0.05$

<https://doi.org/10.1371/journal.pcbi.1010399.t002>

An important question which we have addressed is why some genes are rhythmic at the RNA but not the protein level, and vice versa. Apart from mechanistic causes that explain how, it was not clear why in some cases the rhythmicity is initiated early then lost over the processes (from transcription initiation to protein decay), while in others it is initiated later. In some cases, these lost or gained rhythms might be by-products of the evolution of regulatory processes, but it seems unlikely that all nycthemeral genes be due only to genetic drift. Our previous results showing that rhythmicity of gene expression tends to be a conserved property [14] support this perspective. Indeed, we have found a stronger conservation signal of rhythmic expression for evolutionarily closer species [14]. Moreover, a reasonable assumption is that rhythmic expression can have costs, e.g., of regulation, and thus it is not always obvious that it should be adaptive. For instance, regulatory dynamics can cause substantial changes in noise levels, e.g. the noise strength immediately following gene induction is almost twice the final steady-state value [19].

### Rhythmic proteins: Reduction of overall costs

Our calculation of energetic costs is based on the costs that the cell needs to produce a given steady state protein level. The cell also consumes energy to maintain the steady state protein level. We argue that these costs should correlate with our estimation of expression costs (see section 3 in S1 File). Indeed, the costs of maintaining a steady state protein level are higher for highly expressed than for lowly expressed genes, even if higher expressed proteins have longer half-lives (see section 4 in S1 File). Thus, even taking into account a more complex cost calculation would not change our observation that rhythmic genes are costlier.

Our results suggest that rhythmicity of protein expression has been favored by selection for cost control of gene expression, while keeping sufficient expression levels. In the case of rhythmic genes, what would this sufficient constant level be? We can propose two hypotheses. The first is that it would be the mean expression over the period, since this maintains the same overall amount of protein. The second is that it would be the maximum over the rhythm period, since that is the level needed at least at some point. The second hypothesis explains better the existence of this maximum level during the cycle. Of note, it also strengthens the case for selection on expression cost. Our results also suggest that rhythmicity at the protein level only minimizes costs within the constraints of the rhythmic proteins, without making these costs similar or smaller than those of constant proteins. They also suggest that rhythmic regulation of transcription of highly expressed genes does impact the energy expenditure, consistent with Cheng et al. [33], but as a by-product of the selection acting at protein level. Finally, rhythmic regulation of highly expressed genes might be driven by other costs than energy, *i.e.* high expression levels could be detrimental, *e.g.* because of toxicity. For instance, by altering protein-protein interactions, generating undesirable enzyme activities, or by antagonistic pleiotropy of genes. Thus, for rhythmic genes, the constant level should at least correspond to the mean expression level (Fig 1d). We provide results obtained using both the maximum and the mean of expression in Fig 2a.

Another hypothesis is that rhythmicity at protein level could have been selected for noise adjustment. If that were the case, we would expect that for a given mRNA level, the peak level of proteins (*i.e.*, the highest translation rate) would be associated with lower noise to increase the precision when the protein level is highest. However, both theoretical and empirical arguments contradict this expectation. For a given mRNA level, theoretical models predict that protein level will be negatively correlated with noise [20, 34] (section 6.2 in S1 File). Moreover, measurements of cell-to-cell variability of protein level show that for a given mRNA level, changes in protein level have little to no impact on gene expression precision [20] (section 6.2 in S1 File). Thus, translational efficiency does not seem to be the main driver of expression noise, and selection on noise cannot explain rhythmicity at protein level.

### Rhythmic mRNAs: Noise reduction

Transcription rather than translation seems to be the main source of overall expression noise (section 6.1 in S1 File). However, we can only conclude regarding the transcriptional noise. Small fluctuations of expression away from the optimal wild-type expression have been shown to impact organismal fitness in yeast, where noise is nearly as detrimental as sustained (mean) deviation [35]. Furthermore, noise-increasing mutations in endogenous promoters have been found to be under purifying selection [36], and noise has been shown to be under selection in fly development [21]. In this study, we found lower noise at least at a single time-point among rhythmic transcripts, compared to constant genes. Rhythmic expression of RNAs might be a way to periodically reduce expression noise of highly expressed genes (Fig 2 and S1 Fig), which are under stronger selection. Indeed, we found that genes with rhythmic transcripts are under stronger selection, even controlling for expression level effect. As proposed by Horvath et al. (2019) and supported by results in mouse by Barroso et al. (2018), genes under strong selection could be less tolerant to expression noise. Thus, periodic accumulation of mRNAs might be a way to periodically reduce transcriptional noise of noise-sensitive genes (Fig 1c), *i.e.* genes under stronger selection. However, our results are limited by the fact that noise estimation is based on a single time-point measurement since no scRNA time-series data are currently available for these species. Since the peak time of rhythmic transcripts is distributed across all times (Fig 2a in S1 File), the mean noise estimated at a given time-point includes the noise of the

genes that are peaking at that time (lowest noise) and all the others that have a higher noise than those at their own peak time-point (Fig 2b in [S1 File](#)). Thus, genes peaking at the time-point of the scRNA measurement had sufficiently low noise to produce an averaged noise in rhythmic genes much lower than constant genes. We suggest that, over a gene-specific threshold of mRNA level, the likelihood that ribosomes interact with mRNA molecules increases much faster than under this threshold, allowing to maximize expression accuracy at the peaks. Future experiments would allow to test this suggestion.

We are aware that this theory is difficult to test experimentally. An interesting direction would be the measurement the cell-to-cell protein fluorescence at different time-point for different tissues and species, providing real-time information on the noise fluctuations at various levels. Moreover, the fitness effect of noise could be estimated from growth rates in unicellular organisms measured at different time-point, under conditions which favor rhythmicity or which don't.

Stochasticity of gene expression could also allow to maintain an oscillator system that would be damped otherwise, according to deterministic simulations [37]. Furthermore, expression noise could generate phenotypic diversity, and improve fitness in fluctuating environments [38–43]. Thus, rhythmicity of transcripts might lead to periodic stochasticity that maintains the amplitude of oscillations as well as lead to alternative molecular phenotypes.

Higher precision allows to increase the robustness of gene expression when the function is most needed, while higher stochasticity allows to present diverse molecular phenotypes. Thus alternating between these two states might improve the fitness of organisms in fluctuating environments. We also found that rhythmic expression involves highly expressed, tissue-specific genes, consistent with the observation of tissue-specific rhythms. Selection for rhythmic regulation of highly expressed genes might be due to an efficient reduction of periodic protein-protein misinteractions. Finally, since genes under strong selection could also become less tolerant to high expression noise levels, our results suggest that rhythmicity at the mRNA level might have been under strong selection for these noise-sensitive genes.

## Materials and methods

### Datasets

Details of the datasets are available in [S2 Table](#).

**Mus musculus.** Mouse liver transcriptomic and proteomic time-series datasets come from Mauvoisin et al. [17]. Original protein counts dataset was downloaded from ProteomeX-change (PXD001211) file Combined\_WT, and cleaned from data with multiple Uniprot.IDs affectations between raw peptides data. Time-series transcriptomic data was downloaded from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) accession (GSE33726) [44] Multi-tissues time-serie transcriptomic data used for tissue-specificity expression comparisons are microarray data from Zhang et al. [9] downloaded from the NCBI GEO accession GSE54652 (see Materials of Laloum and Robinson-Rechavi [14] for more details). Tissues analysed are: adrenal gland, aorta, brain stem, brown adipose, cerebellum, heart, kidney, liver, lung, muscle, and white adipose. We have voluntarily excluded the hypothalamus since it contains the Supra-Chiasmatic Nucleus known to have core clock genes that synchronise the body. Therefore, their rhythmicity is probably dominated by other causes than a regulation of cost.

Multi-tissues time-serie proteomic data come from: Mauvoisin et al. [17] for the liver, Noya et al. [45] for the forebrain, Chang et al. [46] for the tendon, and Dudek et al. [47] for the cartilage. Finally, single-cell data of thousand of cells in organs for which we had time-series datasets, *i.e.* liver, lung, kidney, muscle, aorta, and heart, were downloaded from figshare using R

objects from FACS single-cell datasets [29]. We kept cellular sub-types which were found in at least 100 cells and that we considered to be characteristic of the tissue. This leads to keep: endothelial cell of hepatic sinusoid and hepatocyte for the liver; epithelial cell of lung, lung endothelial cell, and stromal cell for the lung; endothelial cell, mesenchymal stem cell, and skeletal muscle satellite cell for the muscle; cardiac muscle cell, endocardial cell, and endothelial cell for the heart; endothelial cell for the aorta; endothelial cell, epithelial cell of proximal tubule, and kidney collecting duct epithelial cell for the kidney.

**Arabidopsis thaliana.** Leaves time-series proteomic data are the Dataset I from Krahmer et al. [48], cleaned from data with multiple protein identifications. Leaves time-series transcriptomic data were downloaded from the NCBI GEO accession (GSE3416) [49]. Single-cell data of twenty root cells were downloaded from the NCBI GEO accession (GSE46226) [30].

**Ostreococcus tauri.** Unicellular alga proteomics time-series dataset (normalized abundances) come from Noordally et al. [50]. Transcriptomic time-series dataset come from Monnier et al. [51], was downloaded from the NCBI GEO accession (GSE16422) and was cleaned for genes with too much missing values (more than seven).

**Synechococcus elongatus (PCC 7942).** Unicellular cyanobacterium proteomics time-series dataset come from Guerreiro et al. [52]. Transcriptomic time-series dataset come from Ito et al. [53] and was recovered from Guerreiro et al. [52].

**Drosophila melanogaster.** Transcriptomic time-series datasets for the body, the head, and the heart, come from Gill et al. [54] and were downloaded from the NCBI GEO accession (GSE64108) (see Materials of Laloum and Robinson-Rechavi [14] for more details).

**Papio anubis [Olive baboon].** Multi-tissues time-series transcriptomic data used for tissue-specificity expression comparisons are RNA-seq data from Mure et al. [55] (see Materials of Laloum and Robinson-Rechavi [14] for more details).

## Pre-processing

For each time-series dataset, only protein coding genes were kept. ProbIDs assigned to several proteins were removed. Probesets were cross-referenced to best-matching gene symbols by using either Ensembl BioMart software [56], or UniProt [57].

## Rhythm detection

To increase power of rhythm detection [14], we considered biological replicates as new cycles when it was possible. We used `GeneCycle` R package (version 1.1.4) [58] available from CRAN and used the `robust.spectrum` function developed by [59]—with parameters *periodicity.time* = 24 and *algorithm* = “*regression*”—that computes a robust rank-based estimate of the periodogram/correlogram and that we improved with the `try-catch` function to avoid error of dimension with MM-estimation method. When the *p*-values distribution obtained did not correspond to the expected distribution—skewed towards low *p*-values because of the presence of rhythmic genes—we used the results obtained by the rhythm detection method used by the original paper from where the data came after checking they presented a classic skewed *p*-values distribution. Finally, for each gene or protein having several data (ProbIDs or transcripts), we combined *p*-values by Brown’s method using the `EmpiricalBrownsMethod` R package (S1 File). Thus, for each dataset, we obtained a unique rhythm *p*-value per gene or per protein. Due either to low-amplitude or less-accurate measurements, it was more challenging to identify rhythms in proteomics data. That is why, in general, we used lower stringency for proteins.

## Consistent gene expression levels

Tissue-specific mRNA or protein abundances were the average or the maximum level of the  $n$  time-points such as for gene  $i$ :

$$\max N_i = \frac{\max_1 N_{i,j} + \max_2 N_{i,j}}{2} \quad \text{with} \quad \max_1 N_{i,j} \neq \max_2 N_{i,j} \quad 0 \leq j \leq n \quad (2)$$

$$\text{mean} N_i = \frac{\sum_{j=1}^n N_{i,j}}{n} \quad \text{with} \quad n: \text{the number of time - points} \quad (3)$$

$N_{pj}$ : for the abundance of the protein  $i$

$N_{RNAi}$ : for the abundance of the transcript  $i$

In mouse, we have used peptide level data to compare protein abundances in the calculation of expression costs.

## Expression costs

Energetic costs of each AA (unit: high-energy phosphate bonds per molecule) come from Akaishi and Gojobori [60], or Wagner [15] which are linearly correlated (S5 Fig). The averaged AA synthesis cost of one protein of length  $L_p$  is:

$$\bar{c}_{AA} = \frac{\sum_{j=1}^{L_p} c_{AAj}}{L_p} \quad (4)$$

Protein sequences come from FASTA files downloaded from EnsemblPlants [56] or UniProt [57] for: Proteome UP000002717 for *Synechococcus elongatus* PCC 7942, Uniprot Reviewed [Swiss-Prot] for *Mus musculus*, Proteome UP000009170 for *Ostreococcus tauri*.

Main results use Wagner [15] AA costs data. We provide supplementary results using Akaishi and Gojobori [60] AA costs data (S6 Fig).

## Multi-tissues analysis

To obtain comparable expressions levels between different tissues or datasets, we normalized expression values by Z-score transformation such as in the dataset of  $n$  genes, the mean expression of the gene  $i$  becomes:

$$\text{mean} Z_i = \frac{\text{mean} N_i - \bar{N}}{\max Z \cdot \sigma} \quad \text{with} \quad \text{mean} N_i: \text{the average expression level of gene } i \text{ (formula (3))} \quad (5)$$

$$\bar{N} = \frac{\sum_{i=1}^n N_i}{n}$$

$\sigma$ : the standard deviation of  $\text{mean} N_i$

$\max Z$ : the maximal value of the  $\text{mean} Z$ -scores

$n$ : the number of genes

and the maximal expression of the gene  $i$  becomes:

$$\max Z_i = \frac{\max N_i - \bar{N}}{\max Z \cdot \sigma} \quad \text{with} \quad \begin{array}{l} \max N_i: \text{ the maximal expression level of gene } i \text{ (formula (2))} \\ \max Z: \text{ the maximal value of the } \max Z\text{-scores} \\ \sigma: \text{ the standard deviation of } \max N_i \end{array} \quad (6)$$

$$Z_i \in [0, 1]$$

To compare the expression levels between the set of  $n_r$  rhythmic tissues and the set of  $n_{\bar{r}}$  non-rhythmic tissues, we estimated the difference ( $\delta$ ) of expression levels between these two groups for each gene  $i$  such as:

$$\delta_i = \frac{\sum_{j_r=1}^{n_r} (\text{mean } Z_{i,j_r})}{n_r} - \frac{\sum_{j_{\bar{r}}=1}^{n_{\bar{r}}} (\text{mean } Z_{i,j_{\bar{r}}})}{n_{\bar{r}}} \quad (7)$$

with  $n_r$ : the number of tissues in which gene  $i$  is rhythmic ( $p \leq 0.01$  or  $0.05$ )  
 $n_{\bar{r}}$ : the number of tissues in which gene  $i$  is not rhythmic ( $p > 0.1$  or  $0.5$ )  
 $\text{mean } Z_{i,j_r}$ : the mean expression level  $Z$  – score normalized of gene  $i$  in the tissue  $j_r$  in which it is rhythmic  
 $\text{mean } Z_{i,j_{\bar{r}}}$ : idem for non – rhythmic tissue  
 $\text{mean } Z_i \in [0, 1]$ , defined as formula (5)

Finally, we analysed the distribution of  $\delta_i$  and generated a Student’s t-test to compare its mean distribution with an expected theoretical mean of 0.

### Tissue-specificity of gene expression

To calculate a tissue-specificity  $\tau$  for each gene  $i$ , we log-transformed the averaged gene expression and followed Kryuchkova-Mostacci and Robinson-Rechavi [27] to make expression values manageable. Thus, among the  $n$  tissues, the tissue-specificity of gene  $i$  is:

$$\tau_i = \frac{\sum_{j=1}^n (1 - \hat{N}_{i,j})}{n - 1} \quad \text{with} \quad \begin{array}{l} n: \text{ the number of tissues} \\ \hat{N}_{i,j} = \frac{\log(\text{mean } N_{i,j})}{\max_{1 \leq j \leq n} (\log(\text{mean } N_{i,j}))} \\ \max_{1 \leq j \leq n} (\log(\text{mean } N_{i,j})) \text{ is the maximal expression level of gene } i \\ \text{among the } n \text{ tissues} \\ \hat{N}_{i,j} \in [0, 1] \end{array} \quad (8)$$

This tissue-specificity formula was described by Yanai et al. [26]. Finally, we performed linear regressions to analyse the respective and the interaction influences of gene expression level and tissue-specificity into the rhythmicity. For mouse [9], we used RNA-seq data to estimate the mean expression used for the calculation of  $\tau$ , and used rhythm  $p$ -values obtained from

microarray dataset since there was more time-points in the microarray time-series (as discussed in benchmark paper). [S7 Fig](#) shows the distributions of  $\tau$  obtained.

### Gene expression noise quantification

We estimated a unique expression noise per gene by taking advantages of *Arabidopsis thaliana* roots [30] and *Mus musculus* liver, lung, kidney, muscle, aorta, and heart [29] (from figshare using R objects of FACS single-cell datasets) single-cell RNAseq data (see [Materials](#)). For *Arabidopsis*, we obtained gene expression for twenty root quiescent centre (QC) cells for which a total of 14,084 genes had non-zero expression in at least one of the 20 single cells. For mouse, we obtained gene expression for a number of cells ranging from 180 to 3772 cells, for which a total of genes with non-zero expression was around 19,000 genes. As in Barroso et al. [28], we kept only genes whose expression level satisfied  $\log[\text{FPKM} + 1] > 1.5$  in at least one single cell. We calculated the stochastic gene expression ( $F^*$ ) defined by Barroso et al. [28] as a measure for gene expression noise (cell-to-cell variability) designed to control the biases associated with the correlation between the expression mean ( $\mu$ ) and the variance ( $\sigma^2$ ) by using the lowest degree polynomial regression which decorrelate them. We tested different degrees of the polynomial regression to estimate  $\log(\sigma^2)$  in the calculation of  $F^*$  and measured the correlation based on Kendall's rank and linear regression slope (see section 6.4 in [S1 File](#)). The  $F^*$  polynomial degree of Barroso et al. (2018) was the best method because it removed the correlation between mean expression and  $F^*$  (smallest linear regression slope and Kendall's) and keep a large range of noise values for every mean expression (smallest R2) (section 6.4 in [S1 File](#)). Finally, since the peak times of rhythmic genes is largely distributed ([Fig 2a in S1 File](#)), we expect the mean noise of a given time-point to be general to all time-points ([Fig 2b in S1 File](#)), therefore giving a good indicator to estimate its biologically relevant.

### dN/dS analysis

dN/dS data have been downloaded from Ensembl BioMart Archive 99 [56]. The homologous species used are: *Mus musculus*—*Rattus norvegicus*; *Arabidopsis thaliana*—*Arabidopsis lyrata*; and *Anopheles gambiae*—*Aedes aegypti*. We first tested the hypothesis that rhythmic ( $p\text{-value} \leq \text{cutoff.1}$ ) versus non-rhythmic genes ( $p\text{-value} > \text{cutoff.2}$ ) have equal ratio of non-synonymous to synonymous substitutions. Then, since rhythmic genes are now known to be enriched in highly expressed genes and because highly expressed genes are under purifying selection, we controlled for the effect of gene expression on dN/dS ratio by testing the hypothesis that rhythmic ( $p\text{-value} \leq \text{cutoff.1}$ ) versus non-rhythmic ( $p\text{-value} > \text{cutoff.2}$ ) genes among residuals of the linear regression fitting gene expression level and dN/dS ratio. Complete results are given in [S6 Table](#) and in a simplified way in [Table 2](#).

In R code it gives:

```
> lmTest = lm(log(dNdS) ~ log(gene.expr.level), data)
> rhythmic.residuals = lmTest$residuals[data$rhythm.
pvalue<=cutoff.1]
> nonrhythmic.residuals = lmTest$residuals[data$rhythm.
pvalue>cutoff.2]
```

```
> t.test(rhythmic.residuals, nonrhythmic.residuals)
```

Plots have been generated using ggplot2 package (version 3.3.2) run in R version 4.0.2.

## Supporting information

**S1 Fig. Mean or maximum mRNA expression level calculated from time-series datasets.** Rhythmic transcripts are highly expressed transcripts.  
(TIFF)

**S2 Fig. Linear regression analysis between the order of the rhythmicity signal (from the most rhythmic to the most un-rhythmic genes) and the total expression cost ( $C_p$ , calculated from the mean expression level).**  
(TIFF)

**S3 Fig. Distribution of the difference of expression levels ( $\delta$ ) between rhythmic tissues group and non-rhythmic tissues group obtained for every gene ( $\delta$  calculated from mean expression levels; see [Methods](#)).  $\delta$  shows a bimodal distribution at protein level.**  
(TIFF)

**S4 Fig. a)** Histograms of Pearson's and Spearman's coefficients testing the correlation between the expression level and the rhythmicity signal (rhythm p-value) measured across the body for every tissue-specific genes ( $\tau > 0.5$ ). **b)** Student's t-test comparing the mean of dN/dS between gene sets A and B: Tissue-specific genes which are mostly rhythmic in tissues where they are highly expressed are under stronger selective constraint than those which are rhythmic in tissues where they are lowly expressed.  
(TIFF)

**S5 Fig. Linear relationships of amino acid (AA) biosynthesis costs estimated by Akashi and Gojobori [60], and Wagner [15].**  
(EPS)

**S6 Fig. Comparison of the averaged AA synthesis costs calculated in both groups: Rhythmic VS random genes group, using Akashi and Gojobori versus Wagner AA costs data.**  
(EPS)

**S7 Fig. Histograms of tissue-specificity  $\tau$ .**  
(TIFF)

**S1 Table. Energetic costs estimated for each amino acid.**  
(XLSX)

**S2 Table. Protein and RNA time-series datasets & single-cell RNA datasets.**  
(XLSX)

**S3 Table. Comparison of expression level of genes in tissues in which they are rhythmic or not.**  
(XLSX)

**S4 Table. Pearson and Spearman correlation test:  $\tau$  vs. the number of tissues in which the gene is rhythmic.**  
(XLSX)

**S5 Table. Comparison of expression noise level ( $F^*$ ) between rhythmic versus non-rhythmic genes: Welch two sample t-test.**  
(XLSX)

**S6 Table. Student's test testing the hypothesis that the dN/dS distributions are equal among rhythmic versus non-rhythmic genes.**  
(XLSX)



**S7 Table. Linear regression slope, R2, and Kendall's correlation test of the relationships between gene expression level and variance, standard deviation (sd), or different noise estimation methods.**

(XLSX)

**S8 Table. Welch two sample t-test comparing protein evolutionary conservation (estimated by the dN/dS ratio) between rhythmic and non-rhythmic genes with, or without controlling for the effect of gene expression level (triangles).**

(XLSX)

**S1 File. Complementary information.**

(PDF)

## Acknowledgments

We thank Johanna Kraemer (CIG, UNIL, Lausanne Switzerland) for her helpful advice on datasets in plants. We also thank Qingyao Huang (DMLS, UZH, Zürich, Switzerland) and Manfredo Quadroni (PAF, UNIL, Lausanne, Switzerland) for their helpful advice on interpretation of proteomics data.

## Author Contributions

**Conceptualization:** David Laloum, Marc Robinson-Rechavi.

**Data curation:** David Laloum.

**Formal analysis:** David Laloum.

**Funding acquisition:** Marc Robinson-Rechavi.

**Investigation:** David Laloum, Marc Robinson-Rechavi.

**Supervision:** Marc Robinson-Rechavi.

**Visualization:** David Laloum.

**Writing – original draft:** David Laloum.

**Writing – review & editing:** David Laloum, Marc Robinson-Rechavi.

## References

1. Graham B. The basics of selection. Chapman & Hall. 1997;.
2. Hubbard J, Frisk MK, Ruppert E, Tsai JW, Fuchs F, Robin-Choteau L, et al. Melanopsin-dependent direct photic effects are equal to clock-driven effects in shaping the nycthemeral sleep-wake cycle. *bioRxiv*. 2020;.
3. Yoo SH, Yamazaki S, Lowrey PL, Shimomura K, Ko CH, Buhr ED, et al. PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. *Proceedings of the National Academy of Sciences*. 2004; 101(15):5339–5346. PMID: [14963227](https://pubmed.ncbi.nlm.nih.gov/14963227/)
4. Boothroyd CE, Wijnen H, Naef F, Saez L, Young MW. Integration of Light and Temperature in the Regulation of Circadian Gene Expression in *Drosophila*. *PLOS Genetics*. 2007; 3:1–16. <https://doi.org/10.1371/journal.pgen.0030054> PMID: [17411344](https://pubmed.ncbi.nlm.nih.gov/17411344/)
5. Nagoshi E, Saini C, Bauer C, Laroche T, Naef F, Schibler U. Circadian Gene Expression in Individual Fibroblasts: Cell-Autonomous and Self-Sustained Oscillators Pass Time to Daughter Cells. *Cell*. 2004; 119(5):693–705. <https://doi.org/10.1016/j.cell.2004.11.015> PMID: [15550250](https://pubmed.ncbi.nlm.nih.gov/15550250/)
6. Gerber A, Saini C, Curie T, Emmenegger Y, Rando G, Gosselin P, et al. The systemic control of circadian gene expression. *Diabetes, Obesity and Metabolism*. 2015; 17(S1):23–32. <https://doi.org/10.1111/dom.12512> PMID: [26332965](https://pubmed.ncbi.nlm.nih.gov/26332965/)

7. Saini C, Suter DM, Liani A, Gos P, Schibler U. The Mammalian Circadian Timing System: Synchronization of Peripheral Clocks. *Cold Spring Harbor Symposia on Quantitative Biology*. 2011; 76:39–47. <https://doi.org/10.1101/sqb.2011.76.010918> PMID: 22179985
8. Policarpo M, Fumey J, Lafargeas P, Naquin D, Thermes C, Naville M, et al. Contrasting gene decay in subterranean vertebrates: insights from cavefishes and fossorial mammals. *Molecular Biology and Evolution*. 2020;.
9. Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. A circadian gene expression atlas in mammals: Implications for biology and medicine. *Proceedings of the National Academy of Sciences*. 2014; 111(45):16219–16224. <https://doi.org/10.1073/pnas.1408886111> PMID: 25349387
10. Boyle G, Richter K, Priest HD, Traver D, Mockler TC, Chang JT, et al. Comparative Analysis of Vertebrate Diurnal/Circadian Transcriptomes. *PLOS ONE*. 2017; 12(1):1–18. <https://doi.org/10.1371/journal.pone.0169923> PMID: 28076377
11. Korenčič A, Košir R, Bordyugov G, Lehmann R, Rozman D, Herzel H. Timing of circadian genes in mammalian tissues. *Scientific Reports*. 2014; 4(1):5782. <https://doi.org/10.1038/srep05782> PMID: 25048020
12. Wang GZ, Hickey SL, Shi L, Huang HC, Nakashe P, Koike N, et al. Cycling Transcriptional Networks Optimize Energy Utilization on a Genome Scale. *Cell Reports*. 2015; 13(9):1868–1880. <https://doi.org/10.1016/j.celrep.2015.10.043> PMID: 26655902
13. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Correction: Corrigendum: Global quantification of mammalian gene expression control. *Nature*. 2013; 495(7439):126–127. <https://doi.org/10.1038/nature11848> PMID: 23407496
14. Laloum D, Robinson-Rechavi M. Methods detecting rhythmic gene expression are biologically relevant only for strong signal. *PLOS Computational Biology*. 2020; 16(3):1–23. <https://doi.org/10.1371/journal.pcbi.1007666> PMID: 32182235
15. Wagner A. Energy Constraints on the Evolution of Gene Expression. *Molecular Biology and Evolution*. 2005; 22(6):1365–1374. <https://doi.org/10.1093/molbev/msi126> PMID: 15758206
16. Lynch M, Marinov GK. The bioenergetic costs of a gene. *Proceedings of the National Academy of Sciences*. 2015; 112(51):15690–15695. <https://doi.org/10.1073/pnas.1514974112> PMID: 26575626
17. Mauvoisin D, Wang J, Jouffe C, Martin E, Atger F, Waridel P, et al. Circadian clock-dependent and -independent rhythmic proteomes implement distinct diurnal functions in mouse liver. *Proceedings of the National Academy of Sciences*. 2014; 111(1):167–172. <https://doi.org/10.1073/pnas.1314066111> PMID: 24344304
18. Baerenfaller K, Massonnet C, Walsh S, Baginsky S, Bühlmann P, Hennig L, et al. Systems-based analysis of Arabidopsis leaf growth reveals adaptation to water deficit. *Molecular systems biology*. 2012; 8:606–606. <https://doi.org/10.1038/msb.2012.39> PMID: 22929616
19. Thattai M, van Oudenaarden A. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences*. 2001; 98(15):8614–8619. <https://doi.org/10.1073/pnas.151588598> PMID: 11438714
20. Hausser J, Mayo A, Keren L, Alon U. Central dogma rates and the trade-off between precision and economy in gene expression. *Nature Communications*. 2019; 10(1):68. <https://doi.org/10.1038/s41467-018-07391-8> PMID: 30622246
21. Liu J, Frochoux M, Gardeux V, Deplancke B, Robinson-Rechavi M. Inter-embryo gene expression variability recapitulates the hourglass pattern of evo-devo. *BMC Biology*. 2020; 18(1):129. <https://doi.org/10.1186/s12915-020-00842-z> PMID: 32950053
22. Van Oss SB, Carvunis AR. De novo gene birth. *PLOS Genetics*. 2019; 15(5):1–23. <https://doi.org/10.1371/journal.pgen.1008160> PMID: 31120894
23. Neme R, Tautz D. Phylogenetic patterns of emergence of new genes support a model of frequent de novo evolution. *BMC Genomics*. 2013; 14(1):117. <https://doi.org/10.1186/1471-2164-14-117> PMID: 23433480
24. Wall DP, Hirsh AE, Fraser HB, Kumm J, Giaever G, Eisen MB, et al. Functional genomic analysis of the rates of protein evolution. *Proceedings of the National Academy of Sciences*. 2005; 102(15):5483–5488. <https://doi.org/10.1073/pnas.0501761102> PMID: 15800036
25. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. Why highly expressed proteins evolve slowly. *Proceedings of the National Academy of Sciences*. 2005; 102(40):14338–14343. <https://doi.org/10.1073/pnas.0504070102> PMID: 16176987
26. Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*. 2004; 21(5):650–659. <https://doi.org/10.1093/bioinformatics/bti042> PMID: 15388519

27. Kryuchkova-Mostacci N, Robinson-Rechavi M. A benchmark of gene expression tissue-specificity metrics. *Briefings in Bioinformatics*. 2016; 18(2):205–214.
28. Barroso GV, Puzovic N, Dutheil JY. The Evolution of Gene-Specific Transcriptional Noise Is Driven by Selection at the Pathway Level. *Genetics*. 2018; 208(1):173–189. <https://doi.org/10.1534/genetics.117.300467> PMID: 29097405
29. Consortium TM. Robject files for tissues processed by Seurat. 2018;.
30. Efroni I, Ip PL, Nawy T, Mello A, Birnbaum KD. Quantification of cell identity from single-cell gene expression profiles. *Genome Biology*. 2015; 16(1):9. <https://doi.org/10.1186/s13059-015-0580-x> PMID: 25608970
31. Ouyang Y, Andersson CR, Kondo T, Golden SS, Johnson CH. Resonating circadian clocks enhance fitness in cyanobacteria. *Proceedings of the National Academy of Sciences*. 1998; 95(15):8660–8664. <https://doi.org/10.1073/pnas.95.15.8660> PMID: 9671734
32. Dodd AN, Salathia N, Hall A, Kévei E, Tóth R, Nagy F, et al. Plant Circadian Clocks Increase Photosynthesis, Growth, Survival, and Competitive Advantage. *Science*. 2005; 309(5734):630–633. <https://doi.org/10.1126/science.1115581> PMID: 16040710
33. Cheng Y, Chi Y, Zhang L, Wang GZ. A single factor dominates the behavior of rhythmic genes in mouse organs. *BMC Genomics*. 2019; 20(1):879. <https://doi.org/10.1186/s12864-019-6255-3> PMID: 31747875
34. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, et al. Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science*. 2010; 329(5991):533–538. <https://doi.org/10.1126/science.1188308> PMID: 20671182
35. Schmiedel JM, Carey LB, Lehner B. Empirical mean-noise fitness landscapes reveal the fitness impact of gene expression noise. *Nature Communications*. 2019; 10(1):3180. <https://doi.org/10.1038/s41467-019-11116-w> PMID: 31320634
36. Metzger BPH, Yuan DC, Gruber JD, Duveau F, Wittkopp PJ. Selection on noise constrains variation in a eukaryotic promoter. *Nature*. 2015; 521(7552):344–347. <https://doi.org/10.1038/nature14244> PMID: 25778704
37. Geva-Zatorsky N, Rosenfeld N, Itzkovitz S, Milo R, Sigal A, Dekel E, et al. Oscillations and variability in the p53 system. *Molecular Systems Biology*. 2006; 2(1):2006.0033. <https://doi.org/10.1038/msb4100068> PMID: 16773083
38. Acar M, Mettetal JT, van Oudenaarden A. Stochastic switching as a survival strategy in fluctuating environments. *Nature Genetics*. 2008; 40(4):471–475. <https://doi.org/10.1038/ng.110> PMID: 18362885
39. Wolf L, Silander OK, van Nimwegen E, Golding I. Expression noise facilitates the evolution of gene regulation. *eLife*. 2015; 4:e05856. <https://doi.org/10.7554/eLife.05856> PMID: 26080931
40. Thattai M, van Oudenaarden A. Stochastic Gene Expression in Fluctuating Environments. *Genetics*. 2004; 167(1):523–530. <https://doi.org/10.1534/genetics.167.1.523> PMID: 15166174
41. Schmutzer M, Wagner A. Gene expression noise can promote the fixation of beneficial mutations in fluctuating environments. *PLOS Computational Biology*. 2020; 16(10):1–24. <https://doi.org/10.1371/journal.pcbi.1007727> PMID: 33104710
42. Duveau F, Hodgins-Davis A, Metzger BP, Yang B, Tryban S, Walker EA, et al. Fitness effects of altering gene expression noise in *Saccharomyces cerevisiae*. *eLife*. 2018; 7:e37272. <https://doi.org/10.7554/eLife.37272> PMID: 30124429
43. Liu J, Martin-Yken H, Bigey F, Dequin S, François JM, Capp JP. Natural Yeast Promoter Variants Reveal Epistasis in the Generation of Transcriptional-Mediated Noise and Its Potential Benefit in Stressful Conditions. *Genome Biology and Evolution*. 2015; 7(4):969–984. <https://doi.org/10.1093/gbe/evv047> PMID: 25762217
44. Jouffe C, Cretenet G, Symul L, Martin E, Atger F, Naef F, et al. The Circadian Clock Coordinates Ribosome Biogenesis. *PLOS Biology*. 2013; 11(1):1–17. <https://doi.org/10.1371/journal.pbio.1001455> PMID: 23300384
45. Noya SB, Colameo D, Brüning F, Spinnler A, Mircsof D, Opitz L, et al. The forebrain synaptic transcriptome is organized by clocks but its proteome is driven by sleep. *Science*. 2019; 366 (6462). <https://doi.org/10.1126/science.aav2642> PMID: 31601739
46. Chang J, Garva R, Pickard A, Yeung CYC, Mallikarjun V, Swift J, et al. Circadian control of the secretory pathway maintains collagen homeostasis. *Nature Cell Biology*. 2020; 22(1):74–86. <https://doi.org/10.1038/s41556-019-0441-z> PMID: 31907414
47. Dudek M, Angelucci C, Ruckshanthi JPD, Wang P, Mallikarjun V, Lawless C, et al. Circadian time series proteomics reveals daily dynamics in cartilage physiology. *bioRxiv*. 2019;.

48. Krahmer J, Hindle M, Perby LK, Nielsen TH, VanOoijen G, Halliday KJ, et al. Circadian protein regulation in the green lineage II. The clock gene circuit controls a phospho-dawn in *Arabidopsis thaliana*. *bioRxiv*. 2019;.
49. Bläsing OE, Gibon Y, Günther M, Höhne M, Morcuende R, Osuna D, et al. Sugars and Circadian Regulation Make Major Contributions to the Global Regulation of Diurnal Gene Expression in *Arabidopsis*. *The Plant Cell*. 2005; 17(12):3257–3281. <https://doi.org/10.1105/tpc.105.035261> PMID: 16299223
50. Noordally ZB, Hindle MM, Martin SF, Seaton DD, Simpson TI, Bihan TL, et al. Circadian protein regulation in the green lineage I. A phospho-dawn anticipates light onset before proteins peak in daytime. *bioRxiv*. 2018;.
51. Moulager M, Monnier A, Jesson B, Bouvet R, Mosser J, Schwartz C, et al. Light-Dependent Regulation of Cell Division in *Ostreococcus*: Evidence for a Major Transcriptional Input. *Plant Physiology*. 2007; 144(3):1360–1369. <https://doi.org/10.1104/pp.107.096149> PMID: 17535824
52. Guerreiro ACL, Benevento M, Lehmann R, van Breukelen B, Post H, Giansanti P, et al. Daily rhythms in the cyanobacterium *synechococcus elongatus* probed by high-resolution mass spectrometry-based proteomics reveals a small defined set of cyclic proteins. *Molecular & cellular proteomics: MCP*. 2014; 13(8):2042–2055. <https://doi.org/10.1074/mcp.M113.035840> PMID: 24677030
53. Ito H, Mutsuda M, Murayama Y, Tomita J, Hosokawa N, Terauchi K, et al. Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in *Synechococcus elongatus*. *Proceedings of the National Academy of Sciences*. 2009; 106(33):14168–14173. <https://doi.org/10.1073/pnas.0902587106> PMID: 19666549
54. Christopher B, Gill S, Melkani G, Panda S. type; 2015. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE64108>. GSE64108.
55. Mure LS, Le HD, Benegiamo G, Chang MW, Rios L, Jillani N, et al. Diurnal transcriptome atlas of a primate across major neural and peripheral tissues. *Science*. 2018; 359(6381). <https://doi.org/10.1126/science.aao0318> PMID: 29439024
56. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhari J, et al. Ensembl 2018. *Nucleic Acids Research*. 2017; 46(D1):D754–D761. <https://doi.org/10.1093/nar/gkx1098>
57. Consortium TU. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*. 2018; 47(D1): D506–D515.
58. Ahdesmaki M, Fokianos K, Strimmer K. GeneCycle: Identification of Periodically Expressed Genes; 2012. Available from: <https://CRAN.R-project.org/package=GeneCycle>.
59. Ahdesmäki M, Lähdesmäki H, Pearson R, Huttunen H, Yli-Harja O. Robust detection of periodic time series measured from biological systems. *BMC Bioinformatics*. 2005; 6(1):117. <https://doi.org/10.1186/1471-2105-6-117> PMID: 15892890
60. Akashi H, Gojobori T. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proceedings of the National Academy of Sciences*. 2002; 99(6):3695–3700. <https://doi.org/10.1073/pnas.062526999> PMID: 11904428