PLOS ONE

RESEARCH ARTICLE

Evolutionary History of the Photolyase/ Cryptochrome Superfamily in Eukaryotes

Qiming Mei[1], **Volodymyr Dvornyk**[2,3]*

1 Key Laboratory of Vegetation Restoration and Management of Degraded Ecosystems, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, People's Republic of China, **2** School of Biological Sciences, the University of Hong Kong, Pokfulam Rd., Hong Kong SAR, People's Republic of China, **3** Department of Life Sciences, College of Science and General Studies, Alfaisal University, Riyadh, Kingdom of Saudi Arabia

* dvornyk@hku.hk

Citation: Mei Q, Dvornyk V (2015) Evolutionary History of the Photolyase/Cryptochrome Superfamily in Eukaryotes. PLoS ONE 10(9): e0135940. doi:10.1371/journal.pone.0135940

Editor: Nicholas S Foulkes, Karlsruhe Institute of Technology, GERMANY

Received: April 2, 2015

Accepted: July 28, 2015

Published: September 9, 2015

Data Availability Statement: All sequences are available from the Genbank database, please see supplementary S1 Table for accession numbers.

Funding: This study was supported by The University of Hong Kong PhD Fellowship to Mr. Qiming Mei. 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.

Abstract

Background

Photolyases and cryptochromes are evolutionarily related flavoproteins, which however perform distinct physiological functions. Photolyases (PHR) are evolutionarily ancient enzymes. They are activated by light and repair DNA damage caused by UV radiation. Although cryptochromes share structural similarity with DNA photolyases, they lack DNA repair activity. Cryptochrome (CRY) is one of the key elements of the circadian system in animals. In plants, CRY acts as a blue light receptor to entrain circadian rhythms, and mediates a variety of light responses, such as the regulation of flowering and seedling growth.

Results

We performed a comprehensive evolutionary analysis of the CRY/PHR superfamily. The superfamily consists of 7 major subfamilies: CPD class I and CPD class II photolyases, (6–4) photolyases, CRY-DASH, plant PHR2, plant CRY and animal CRY. Although the whole superfamily evolved primarily under strong purifying selection (average $\omega = 0.0168$), some subfamilies did experience strong episodic positive selection during their evolution. Photolyases were lost in higher animals that suggests natural selection apparently became weaker in the late stage of evolutionary history. The evolutionary time estimates suggested that plant and animal CRYs evolved in the Neoproterozoic Era (~1000–541 Mya), which might be a result of adaptation to the major climate and global light regime changes occurred in that period of the Earth's geological history.

Introduction

Photolyases are light-dependent DNA repair enzymes. They are activated by blue light and repair UV induced DNA damage by removing pyrimidine dimers. Three types of PHRs have been identified: CPD photolyases repair cyclobutane pyrimidine dimers, (6–4) photolyases

repair (6–4) pyrimidine pyrimidone, and cryptochrome-DASHs exhibit a variety of physiological functions including single-strand DNA photolyase activity [1,2], transcriptional regulation in *Synechocystis* [3] and light-dependent regulation of metabolism in *Fusarium* [4].

Photolyases are evolutionary old proteins found in many species from bacteria to vertebrates [1,5]. Recent studies suggested that DNA repair might have a common evolutionary origin with circadian rhythmicity [6]. Circadian rhythmicity is a roughly 24-hour cycle of biochemical, physiological, and behavioral processes. It was found in both prokaryotes and eukaryotes [7]. Circadian rhythms enhance fitness of organisms in both constant and changing environments [8]. A circadian clock system consists of a central oscillator, input and output pathways [9]. The central oscillator is able to maintain the rhythmic output in the absence of the external stimuli [10]. Unicellular organisms rely on a single independent circadian oscillator, whereas organisms with differentiated tissues may have multi-oscillator systems to coordinate with different rhythms [7]. In animals, the central oscillator resides in the brain, which controls the circadian behavior of the whole organism and synchronizes peripheral clocks in other organs [11].

The circadian oscillators of eukaryotes have been extensively studied in fruit fly and mammals [12–14]. One of the key elements of the circadian system in animals and plants is CRY; in plants it acts as a blue light receptor to entrain circadian rhythms [15].

Cryptochromes are flavoproteins, which are homologous to photolyases but lack the DNA repair activity [16]. Cryptochromes and photolyases form the photolyase/cryptochrome superfamily [17]. Cryptochromes are ubiquitous in plants and animals [11]. Photolyases and cryptochromes have two conserved domains, a DNA photolyase related domain and a FAD binding domain. In addition, plant and animal cryptochromes possess a C-terminal domain of a variable length, which is absent in the CRY-DASH and photolyase proteins [18] (Fig 1, the conserved domains were identified by the Conserved Domains Database (CDD) tool [19]). The variation in the length of the C-terminal domains results in functional diversity within the cryptochrome family [20]. The C-terminus of the mammalian cryptochrome possesses a nuclear localization domain required for CRY's nuclear localization; deletion of the C-terminus prevents mammalian CRY from negatively regulating the transcription of other circadian components. The C-terminal domain of the *Arabidopsis* CRY is essential for mediating the signaling mechanism by responding to the light [21].

Cryptochromes possess two chromophores: pterin (in the form of 5, 10-methenyltetrahydrofolate, MTHF) and flavin (in the form of flavin adenine dinucleotide, FAD); they bind to the DNA photolyase related region and FAD binding domain respectively as cofactors that absorb light [22]. Photolyases also have two cofactors, one of them is FAD, and another can be either 5,10-methenyl-tetrahydrofolate (MTHF) or 8-hydroxy-5-deazaflavin (8-HDF) [22]. The

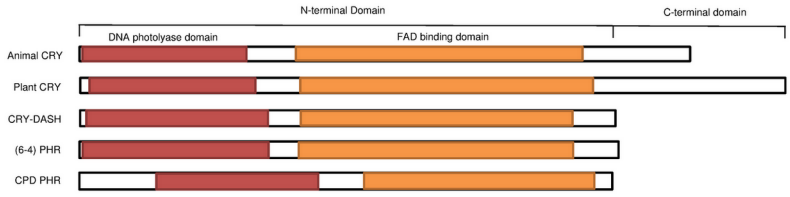

Fig 1. Domain architecture of the CRY/PHR superfamily. The conserved domains were identified by the Conserved Domains Database (CDD) tool [19]. The reference sequences are as follows. Animal CRY: *Homo sapiens* CRY1 (NP_004066); plant CRY: *Arabidopsis thaliana* CRY1 (NP_567341); CRY-DASH: *Xenopus laevis* (NP_001084438); (6–4) PHR: *Danio rerio* (NP_571863); CPD PHR: *D. rerio* CPD class II PHR (NP_957358). The DNA photolyase related domain and a FAD binding domain are shared by both CRY and PHR; whereas the C-terminal variable domain is only present in CRY but not in PHY.

doi:10.1371/journal.pone.0135940.g001

3D-architectures of the photolyase/cryptochrome superfamily members are similar. All of them fold into 2 domains, an α/β domain and a helical domain. These 2 domains are connected by a variable loop and the 2 lobes of the helical domain form a groove, which is called FAD-access cavity. FAD embeds in this "molecular pocket", and may be resolved at the bottom [23].

Studies on molecular evolution of the CRY/PHR superfamily are somewhat limited. The pioneering study of Kanai et al. [24] was conducted on the relatively small number of sequences and thus lacked generalization. Lucas-Lledo and Lynch [5] performed more comprehensive phylogenomic analysis of photolyases focused primarily at the gene gain-loss events and mutation rates. Our study presents the results of the much more comprehensive analysis of the occurrence, phylogeny, selection, conservation and evolutionary time scale of the photolyase superfamily in eukaryotes.

Materials and Methods

DNA and protein sequences

DNA and protein sequences of CRY and PHR were retrieved from the GenBank; only the sequences from fully sequenced eukaryotic genomes were used for the analyses. We used BLASTP and TBLASTN [25] to search the protein database. The following protein sequences were selected as the queries for respective groups of organisms: *Homo sapiens* (NP_004066 and NP_066940) and *Aedes aegypti* (XP_001648498 and XP_001655778) for animal CRY; *Xenopus laevis* (NP_001081421), *Drosophila melanogaster* (NP_724274), *Monosiga brevicollis* (XP_001747506), *Arabidopsis thaliana* (NP_566520) and *Verticillium albo-atrum* (XP_003006428) for (6–4) PHR; *Salpingoeca* sp. ATCC 50818 (XP_004989008), *Arabidopsis thaliana* (NP_568461) and *Verticillium albo-atrum* (XP_003009023) for CRY-DASH; *Xenopus laevis* (NP_001089127), *Drosophila melanogaster* (NP_523653) *Monosiga brevicollis* (XP_001746666) and *Arabidopsis thaliana* (NP_849651) for CPD class II PHR; *Verticillium albo-atrum* (XP_002999933) for fungal CPD class I PHR; *Arabidopsis thaliana* (NP_567341 and NP_171935) for plant CRY and *Arabidopsis thaliana* (NP_182281) for plant PHR2. Bit score of 200 was applied as a lower threshold of sequence selection. Also, we conducted the BLASTP and TBLASTN search in different taxonomy groups of prokaryotes using the same set of queries, to clarify the existence of CRY/PHR proteins in bacteria and archaea. One PHR sequence from each bacteria (*Cronobacter sakazakii* CPD Class I PHR YP_001438714; *Gloeobacter violaceus* (6–4) PHR NP_924695; *Spirosoma linguale* CRY-DASH YP_003390944 and *Geobacter sulfurreducens* CPD Class II PHR NP_953872) and archaea (*Halorhabdus utahensis* CRY-DASH YP_003131490 and CPD Class I PHR YP_003131773; *Methanosarcina barkeri* CPD Class I PHR YP_304088) was included to represent the prokaryotic CRY/PHR groups. Finally, a total of 762 sequences were selected for the analysis (S1 Table).

RNA polymerase II subunit RPB2 was used for the comparative taxonomical analysis [26]. The *Homo sapiens* RPB2 (NP_000929) protein sequence was retrieved using the following sequence as a probe. DNA-directed RNA polymerase subunit B of 2 archaea (*Halorhabdus utahensis* and *Methanosarcina barkeri*) and 4 bacteria (*Cronobacter sakazakii*, *Geobacter sulfurreducens*, *Gloeobacter violaceus* and *Spirosoma linguale*) were used as outgroups (S2 Table). The taxonomical analysis involved smaller number of species (206) because RPB2 sequences from some species were not available in the databases.

Sequence editing and aligning

The protein sequences were aligned by MUSCLE v. 3.8.31 [27]; the nucleotide sequences were aligned according to the protein alignment using Rev-Trans v.1.4 [28] available at http://www.cbs.dtu.dk/services/RevTrans/. The aligned sequences were trimmed manually in Bioedit

v.7.0.9 [29] by removing poorly aligned terminal regions. The final alignment of the CRY/PHR and RBP2 proteins included 489 and 1172 positions, respectively. The sequences utilized in this study are listed in online supporting information S1 and S2 Tables. Also, the alignment of CRY/PHR which is utilized for further analyses was uploaded as supporting information.

Phylogenetic reconstruction

The phylogenetic reconstruction was performed using the protein sequences. The most appropriate model of amino acid substitutions for the data set was determined according to the Akaike information criterion (AIC) and using ProtTest v.3.0 [30,31]. Based on the test, the LG model with proportion of invariable sites, gamma distribution and equilibrium frequencies (LG+I+G+F, p-inv = 0.046, $\alpha = 0.723$) [32] was used for the phylogenetic analysis of the RPB2 proteins. For the CRY/PHR proteins, the best fitting model was WAG+G ($\alpha = 1.061$). Two phylogenetic algorithms were utilized to infer the tree. First, the maximum likelihood (ML) phylogenetic tree was constructed using PhyML v.3.0 [33]. The approximate likelihood-ratio test (aLRT) [34] was applied to estimate a statistical support for individual nodes. Second, we used the Bayesian relaxed clock as implemented in BEAST v.2.1.3 [35]. The length of the MCMC chain was set for 10 million with trees sampled every 1000 steps. The maximum clade credibility tree was determined using TreeAnnotator v.2.1.2 from the BEAST software package.

Analysis of selection

Gene duplication is an important mechanism for generating novel functional proteins, because the redundant homologs are free to accumulate substitutions. However, whether the new function of a gene evolves under positive selection remains controversial [36]. We analyzed positive selection using the ML approach as implemented in HyPhy software package v.2.1.2 [37]. The ratio of nonsynonymous to synonymous substitutions (dN/dS or ω) was used to measure the strength of selection on the CRY/PHR genes. The set of tests was conducted: (1) the basic model (M0), which estimates uniform ω ratio among all sites, was used to calculate a representative estimate for the whole dataset [38]; (2) the site models including M1 (nearly neutral), M2 (selection), M3 (discrete), M7 (beta distribution, $\omega > 1$ disallowed) and M8 (beta distribution, $\omega > 1$ allowed) [39–41]. The likelihood-ratio test (LRT) was performed between the following pairs of the models: (1) M3 vs. M0; (2) comparisons of site models including M2 vs. M1 and M7 vs. M8. Then χ^2 tests were performed with degrees of freedom (*df*) between two compared models.

Also, we utilized the recently developed method, branch-site random effects likelihood (REL) [42] implemented in HyPhy v.2.1.2 [37], to detect episodic diversifying selection. This LRT-based approach identified all lineages in the phylogeny with a proportion of positive selected sites, without making priori assumptions ("foreground" and "background" branches) that may lead to high rates of false positive or negative selection [42]. The branch-site REL method has been successfully applied to infer positive selection in several research publications [43–45].

Identification of conserved residues

Evolutionarily conserved residues and motifs are hypothesized to be functionally important [46]. We utilized ConSurf (http://consurf.tau.ac.il/) to identify probable functionally important residues in the CRY/PHR proteins [47]. The analysis was conducted using the Bayesian algorithm and the LG model with the parameters as specified above. The same sequence alignment used in DIVERGE v.2.0 was uploaded to the server. Degrees of conservation in the protein subfamilies visualized with Chimera v.1.6.2 [48].

Estimating evolutionary time of gene duplications and gain-loss events

The inferred RPB2 tree was tested for the presence of molecular clock using HyPhy v.2.1.2 [37]. Based on the test results, the model with local clock was utilized for the further analysis. Six internal calibration points (CP1-CP6) were used for time estimates. CP1-CP4 indicate the origins of main groups of animals with minimum and maximum time estimates constrained by respective biostratigraphic evidence [49]: CP1 corresponds to the origin of eutherians (113–95.3 Mya); CP2 is the divergence of birds and crocodile (250.4–235 Mya); CP3 is the split of the ray-finned fishes and tetrapods (421.75–416.1 Mya) and CP4 corresponded to the divergence of flies and mosquitos (295.4–238.5 Mya). CP5 was inferred from the phylogenetic study of three genes (*rbcL*, *atpB*, 18S rDNA) in 560 angiosperms, which estimated the origin time of Angiospermae between 179 and 158 Mya [50]. CP6 is the origin of Ascomycota about 500–650 Mya [51].

The computations were conducted using the ML approach as implemented in PAML v. 4.4 [52]. The substitution model for the data was determined by jModelTest v.0.1.1 [53]. The GTR +I+G (p-inv = 0.031 and $\alpha = 0.455$) model turned to fit our data best [54]. We also used the Bayesian relaxed clock as implemented in BEAST v.1.6.2 to estimate the dates of various events in evolution of the superfamily [35].

Results

Occurrence and phylogeny of the CRY/PHR superfamily in eukaryotes

Members of the CRY/PHR superfamily were found in genomes of species across all kingdoms; however, their occurrence greatly varied among taxa (Table 1). The most ubiquitous groups in eukaryotes are CPD class II photolyases, (6–4) photolyases, and CRY-DASH, which are present in the majority of the studied taxa. The other types of photolyases are more taxon-specific. CPD class I photolyases occur only in prokaryotes, fungi and basal eukaryotes (Ciliophora and Euglenozoa). Among fungi, Ascomycetes and Basidiomycetes possess only CPD class I but no CPD class II PHR, while unicellular *Nosema ceranae* has only CPD class II PHR. The CPD class II photolyases are ubiquitous in archaea and eukaryotes but rare in bacteria (found only in Methanomicrobia and Methanobacteria) and lost in placental mammals. Three subfamilies are specific either to plants or animals: plant cryptochromes, plant PHR2 (photolyase/blue-light receptor 2), and animal cryptochromes (Table 1).

Both the ML and Bayesian trees feature 6 main clades with significant statistical support, which correspond to animal CRY and (6–4) photolyases, CRY-DASH, plant PHR2, plant CRY and 2 classes of CPD photolyases (CPD class I and class II) (Fig 2).

Selection in the genes of the CRY/PHR superfamily

We utilized 6 models implemented in PAML to estimate the selective forces after several duplications having occurred in the CRY/PHR superfamily (Table 2). According to the results, the global ω estimated by M0 were low (0.0168), indicating that cryptochromes and photolyases have experienced strong purifying selection. Site model M2a is significantly better than M1a ($2\Delta\ln L = 3996989.1508$, $df = 2$, $p < 0.01$), and M8 manifests a higher likelihood as compared to M7 ($2\Delta\ln L = 1998514.0002$, $df = 2$, $p < 0.01$). Therefore, the alternative hypotheses (selection) cannot be rejected. However, no positively selected sites were identified by the site models.

The branch-site REL method [42] determined about 35.67% (270/757) of the branches to evolve under episodic diversifying selection (S3 Table and S1 Fig). Some of them had sites with very large ω^+ ($>> 1.0$) but the proportion of the sites under positive selection was generally less than 10% in most branches ranging between 0.021–0.082 (e.g., Node99, $\omega^+ = 5230.128$, weight $p^+ = 0.043$). On the other hand, some lineages, e.g., plant CRY in higher plants (Node9),

Table 1. Occurrence of the CRY/PHR genes in main taxa.

Taxa	Class I CPD *Phr*	Class II CPD *Phr*	(6–4) *Phr*	*Cry-DASH*	Plant *Cry*	Plant *PHR2*	Animal *Cry*
Prokaryotes							
Archaea	+	+[a]		+[b]			
Bacteria	+	+	+[c]	+			
Alveolata							
Ciliophora	+						
Perkinsea		+					
Apicomplexa		+					
Stramenopiles							
Bacillariophyta		+	+	+			
Oomycetes			+	+			
Eustigmatophyceae			+	+			
Other basal eukaryotes							
Euglenozoa	+						
Amoebozoa		+					
Heterolobosea				+			
Rhodophyta				+			
Cryptophyta				+			
Choanoflagellates		+	+	+			
Fungi							
Microsporidia		+[d]					
Ascomycetes	+		+	+			
Basidiomycetes	+		+	+			
Viridiplantae							
Chlorophyta		+	+	+	+	+	
Bryophyta		+	+	+	+	+	
Lycopodiophyta		+	+	+	+	+	
Amborellales		+	+	+	+	+	
Liliopsida		+	+	+	+	+	
Eudicotidae		+	+	+	+	+	
Animalia							
Nematoda		+					
Cnidaria		+	+	+			+
Mollusca		+	+	+			+
Crustacea		+	+	+			+
Insecta		+	+				+
Echinodermata		+	+	+			+
Cephalochordata [e]							+
Chondrichthyes [e]							+
Actinopterygii		+	+	+			+
Sarcopterygii [e]							+
Amphibia		+	+	+			+
Testudines		+	+	+			+
Lepidosauria		+	+	+			+
Archosauria		+	+				+
Aves		+	+	+			+
Metatheria		+					+
Monotremata							+
Eutheria							+

[a]: Found only in Methanomicrobia and Methanobacteria

[b]: Found only in Halobacteria

[c]: Found only in *Gloeobacter violaceus*

[d]: Found only in *Nosema ceranae*

[e]: Absent of CRY/PHR homologs may be due to insufficient annotation

doi:10.1371/journal.pone.0135940.t001

indicated quite a large proportion of branch length (weight $p^+ = 0.163$) with positive selection (S3 Table and S1 Fig).

Identification of conserved amino acid residues

In the CRY/PHR superfamily, clusters of conserved sites were located in DNA photolyase related domain (positions 3–159 of the alignment) and FAD binding domain (position 201–477) (Fig 3). The FAD binding domains of plant PHR2 proteins are shorter (~ 95 aa) and share fewer conserved residues than in the other subfamilies. In addition, we identified two highly conserved residues (with conservation scores ≥ 8) shared by all 7 subfamilies: the Arg locates at position 12 of the alignment and the Ser at position 241 (Fig 3).

Time estimates of the events in CRY/PHR superfamily evolution

Time estimates of the key events in evolution of the CRY/PHR superfamily are given in Table 3 and Fig 4. Generally, the ML analysis yielded the smaller values as compared to the Bayesian

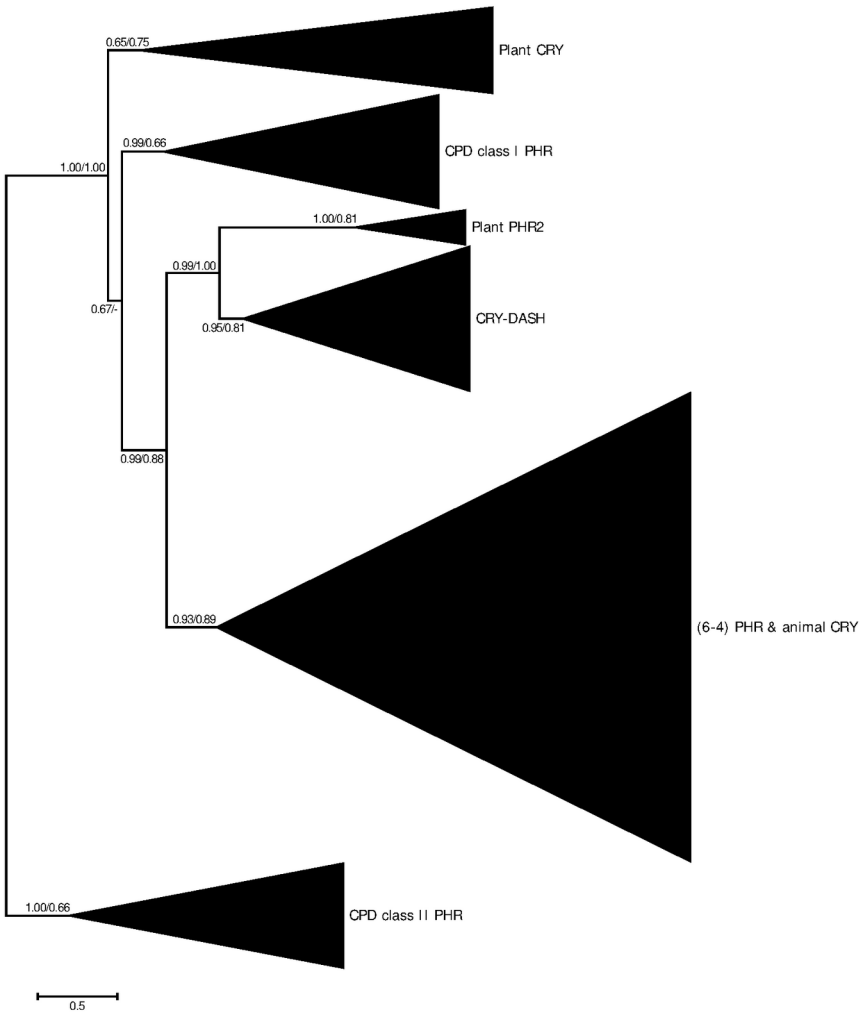

Fig 2. Unrooted maximum-likelihood tree of the CRY/PHR homologs. Maximum-likelihood probabilities and Bayesian posteriors of the node support below 0.5 are not shown. The values are the likelihoods and posteriors respectively. The CRY/PHR superfamily include 6 main subfamilies: animal CRY and (6–4) photolyases, CRY-DASH, plant PHR2, plant CRY and 2 classes of CPD photolyases (CPD class I and class II).

doi:10.1371/journal.pone.0135940.g002

Table 2. Results of the codon based positive selection tests by HyPhy.

Model	Parameters
Model 0 (one ratio)	lnL = -507700.8700
	$\omega = 0.0168$
Model 1 (Neutral)	lnL = -397024795.0599
	$\omega = 0.0168 \pm 0.0000$
	$P_0 = 1.0000$, $\omega_0 = 0.0168$
Model 2 (Selection)	lnL = -395026300.4845
	$\omega = 0.0184 \pm 0.0000$
	$P_1 = 1.0000$, $\omega_1 = 0.0184$
	$P_2 = 0.6005$, $\omega_2 = 4.0276$
LRT [a] (M1/M2, *df* = 2)	2ΔlnL = 3996989.1508 **
Model 3 (Discrete)	lnL = -398024989.2332ω = 0.1823 ± 0.0252
	$P_1 = 0.7792$, $R_1 = 0.2038$, $\omega_1 = 0.4802$, $P_2 = 1.0000$, $R_2 = 1.4261$
LRT (M0/M3, *df* = 4)	2ΔlnL < 0
Model 7 (Beta)	lnL = -398023356.0916
	$\omega = 0.0338 \pm 0.0012$
	$\beta_P = 0.6156$, $\beta_Q = 17.5868$
Model 8 (Beta & ω)	lnL = -397024099.0915
	$\omega = 0.0650 \pm 0.0034$
	$\beta_P = 0.8055$, $\beta_Q = 11.5904$, $\omega_1 = 1$, P = 1.0000
LRT (M7/M8, *df* = 2)	2ΔlnL = 1998514.0002 **

[a]: Likelihood ratio test was estimated by 2ΔlnL and followed by a χ^2 test

*: $p < 0.05$

**: $p < 0.01$).

LRT = $-2(\ln L_0 - \ln L_A)$

doi:10.1371/journal.pone.0135940.t002

analysis. On the other hand, the ML estimates fall within the 95% HPD range of the Bayesian estimates. The aggregated results from both analyses suggest that several major gene duplication and losses in the superfamily occurred between 700 and 400 Mya (Table 3), which correspond to four geological periods: Devonian, Silurian, Ordovician, and Cambrian. The origin of animal CRY is dated back to the Neoproterozoic Era (~1000–541 Mya). The plant CRY is evolutionary younger and evolved in Paleozoic Era (~541–252) (Table 3).

Discussion

Origin and evolutionarily ancient diversification of photolyases

The phylogenetic tree of the CRY/PHR proteins inferred in the present study was similar to those from the previous studies [5,11]. The protein subfamilies form several main clades, including animal CRY & (6–4) photolyases, CRY-DASH, plant CRY and 2 classes of CPD photolyases (CPD class I and class II). Also, we identified a new group of photolyases, PHR2, which is specific to plants. The PHR2 homologs formed a clade phylogenetically related to CRY-DASH.

The pioneering phylogenetic study of cryptochrome/photolyase superfamily [23] involved limited number of sequences. In that study, the CPD class I PHR photolyases were classified into two types according to their cofactors, MTHF and 8-HDF [23]. The split of these two types was thought to occur in the early stage of the evolution of the CRY/PHR superfamily (before the divergence of prokaryotic and eukaryotic organisms) [23].

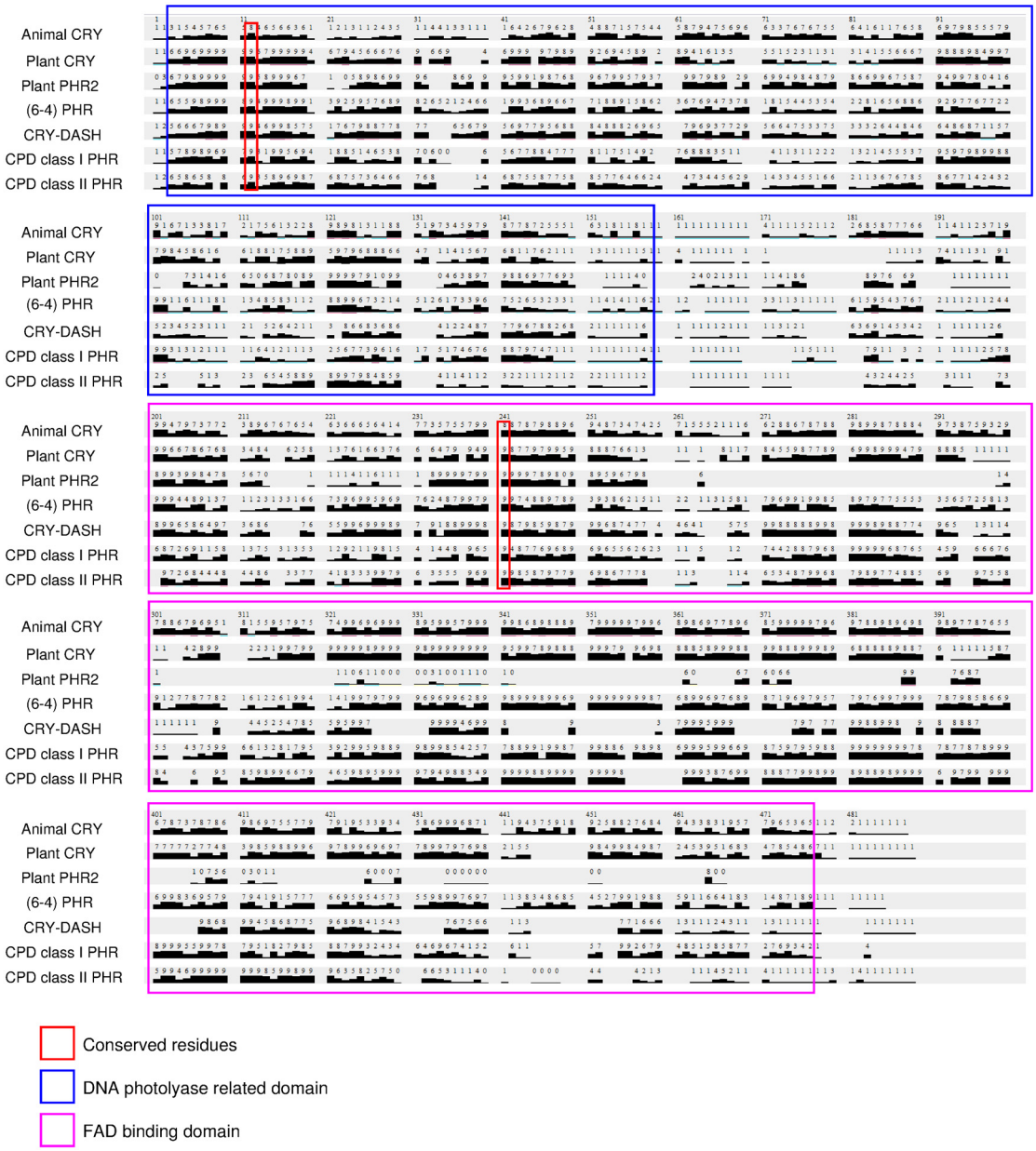

Fig 3. Group conserved residues identified by ConSurf. Degrees of conservation in the protein subfamilies visualized with Chimera v.1.6.2 [48]. The numbers of upper row and black bars indicate the level of conservation (0–9), 0 being the lowest.

doi:10.1371/journal.pone.0135940.g003

The occurrence of CRY/PHR proteins are taxon-specific (Table 1). The distribution of the CRY/PHR proteins in protists reflects paraphyly and evolutionary diversification of this group of eukaryotes. The evolutionary oldest protists, *Trypanosoma brucei*, *Tetrahymena thermophile* and *Leishmania major* have only CPD class I. The evolutionary younger protists lack CPD class I and may possess members of the other three types of photolyases (CPD class II, (6–4) PHR and CRY-DASH).

Another interesting fact is that mammals possess no CPD class II photolyases, except for marsupials (*Monodelphis domestica* and *Sarcophilus harrisii*). Recent research demonstrated that the CPD photolyase in a marsupial *Potorous tridactylus* was able to act as a cryptochrome,

Table 3. Maximum-likelihood and Bayesian time estimates for the nodes (Mya) (Fig 4).

Node	Maximum Likelihood	Bayesian [b]	Evolutionary Events
1	812.23 ± 142.23	961.33 (755.41–1359.82)	Loss of CPD class I *Phr* in plants and animals
2	528.21 ± 168.73	422.30 (285.08–679.26)	Origin of plant *Cry*
3	606.62 ± 122.43	698.45 (500.62–769.12)	Origin of animal *Cry*
4	-[a]	320.27 (269.12–408.18)	Loss of *Cry-DASH* in insects
5	544.06 ± 120.83	598.85 (482.22–638.85)	Origin of *Cry4*
CP1	113–95.3		Loss of photolyases, *Cry-DASH* and *Cry4* in placental mammals [47]
CP2	86.5–66		Loss of *Cry-DASH* and (6–4) *Phr* in birds [47]
CP3	421.75–416.1		Duplication of vertebrate *Cry* [47]
CP4	295.4–238.5		Loss of insect *Cry2* in fly [47]
CP5	179–158		Duplication of plant *Cry* [48,87]
CP6	500–650		Origin of Angiospermae [49]

[a]: Cannot be calculated due to the lack of the DNA sequence of the *Daphnia pulex* RPB2 (EFX81055)

[b]: Posterior mean (95% HPD)

doi:10.1371/journal.pone.0135940.t003

suggesting that ancestral CRY/PHR proteins were likely manifested both DNA repair and circadian clock function [55].

The PHR2 (photolyase/blue receptor 2) subfamily previously reported from green algae *Chlamydomonas reinhardtii* [56] is common in higher plants (S1 and S2 Figs). However, they are absent from all other algae (S1 Table). The PHR2 protein from *Chlamydomonas* manifests CPD photolyase repairing activity for both chloroplast and nuclear DNA [56]. However, PHR2 in *Arabidopsis thaliana* shows no chloroplast DNA repair activity [57]. At the same time, the CPD class II PHR of higher plants exhibits the CPD repair function in DNA of all cellular genomes (nuclear, chloroplasts and mitochondrial) [58]. The property of photolyases to target different organelles is thought to be associated with some functional motif, but the exact mechanism of organelle-targeting is yet to be determined [58]. Among eukaryotes, plants have the largest set of various types of photolyases (Table 1). Some of them have similar functions, as in the above example. Such functional redundancy may be an evolutionary adaptation to the relatively high, as compared to the other eukaryotes, exposure to sunlight and, respectively, harmful UV radiation. This higher exposure prompts for more efficient DNA repairing mechanism.

Although the physiological function of PHR2 is similar to that of the CPD class I and CPD class II proteins, their sequence homology is weak. Our results suggest that PHR2 likely evolved from an ancient CRY-DASH gene (Fig 2). This split was followed by significant functional divergence: CRY-DASH encodes a single-strand DNA (ssDNA) CPD photolyase, whereas plant PHR2 repairs CPD dimers of double-strand DNA (dsDNA) [1]. The ssDNA binding property of CRY-DASH is closely related to its structure, the CPD-binding cavity of the *Arabidopsis* CRY--DASH is unable to stably bind CPD from the dsDNA since the binding is less energetic [59].

Based on their roles in the circadian clock, two groups of functionally different animal CRY proteins were identified [60]. A *Drosophila*-like type 1 CRY is a UV-A/blue light receptor in the circadian oscillator, while a vertebrate-like type 2 CRY is thought to be a negative regulator of the clock's transcriptional feedback loop [61]. Whether the type 2 cryptochromes also have photoreceptor function is still debated [11]. There are several transcription factors involved in animal circadian feedback loop, including Period (PER), Timeless (TIM), brain and muscle Arnt-like protein-1 (BMAL1), CYC (Cycle) and circadian locomotor output cycles kaput (CLOCK or CLK) [62–66]. The molecular mechanisms of the fly and mammalian circadian clocks are different. The fly CRY binds to TIM that results in degradation of the latter and

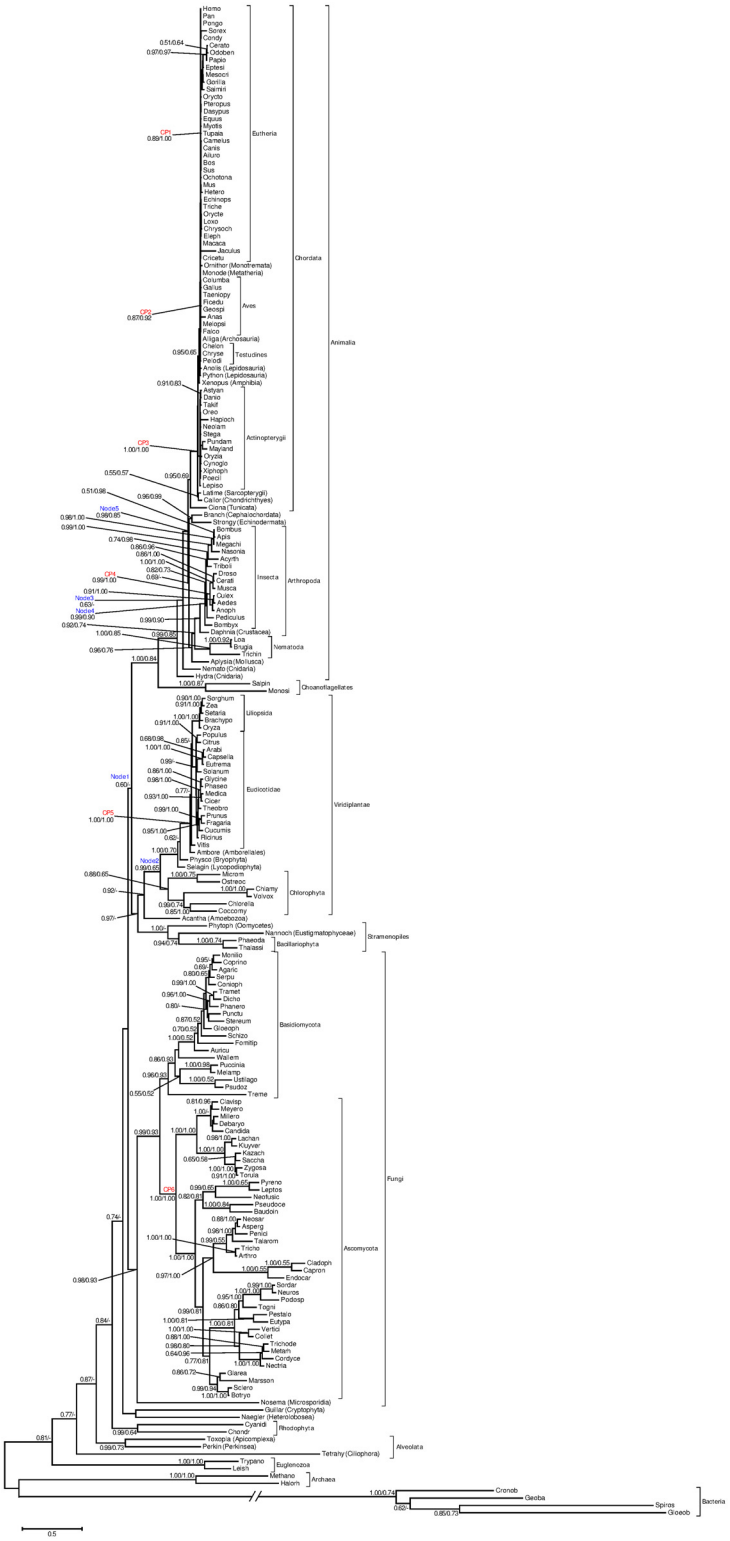

Fig 4. Maximum-likelihood tree with local clock of the RPB2 homologs. Maximum-likelihood probabilities and Bayesian posteriors of the node support below 0.5 are not shown. The values are the likelihoods and posteriors respectively. The internal calibration points (CP1-CP6): CP1 corresponds to the origin of eutherians (113–95.3 Mya); CP2 is the divergence of birds and crocodile (250.4–235 Mya); CP3 is the split of the ray-finned fishes and tetrapods (421.75–416.1 Mya) and CP4 corresponded to the divergence

of flies and mosquitos (295.4–238.5 Mya). CP5 estimated the origin time of Angiospermae (179–158 Mya) [48,87]. CP6 is the origin of Ascomycota about 500–650 Mya [49]. In addition, time of five evolutionary events were estimated (Node11-5). Time estimates was showed in Table 3.

doi:10.1371/journal.pone.0135940.g004

subsequent inhibition of the activity of a PER/TIM dimer. Without CRY, the PER/TIM dimer is able to enter nucleus and repress the transcription of other clock genes. In mammals, three period (PER 1, 2 and 3) and two cryptochrome (CRY 1 and 2) proteins form a cytoplasmic heterodimer, which enters the nucleus and then inhibits transcription of CLK and BMAL1 [67]. In plants, circadian clocks are entrained by red light receptor phytochromes (PHY) as well as blue light receptor cryptochromes, but the signal transduction pathways were not comprehensively studied [68].

The obtained phylogenetic tree (S1 and S2 Figs) indicates that there are two insect *Cry* paralogs, *Cry1* and *Cry2*. Unlike the evolutionary older CRY1, CRY2 is not photosensitive and has a transcriptional repressive function similarly to vertebrate CRY [69]. Among the studied insect species, only few possess both paralogs, the others lack either one. Interestingly, the majority of insects lost photosensitive CRY1 and has only CRY2. It is still unknown how insects lacking CRY1 sense light. It may be possible that those species have developed some compensating mechanism for photosensitivity.

The duplication of the insect CRY ancestor probably occurred well before the origin of insects, maybe even prior to the split of deuterostomes and protostomes, because homologous cryptochromes were found in *Nematostella vectensis* (Nemato1, Nemato1b, Nemato1c and Nemato1d), *Strongylocentrotus purpuratus* (Strongy1 and Strongy2) and *Daphnia pulex* (DaphniaM and DaphniaD, S1 Fig). The duplications of cytochromes also occurred in vertebrates and plants, but there is no evidence in the available literature that the paralogs of mammalian and plant CRYs have experienced functional divergence.

The genome of zebra fish *Danio rerio* possesses the largest known number of the cryptochrome genes, 8. Four of them (NP_001070765, NP_571865, NP_571866 and NP_571867; annotated as *Cry1a*, *1b*, *2a* and *2b*, respectively) are very similar and appear in the same clade with vertebrate *Cry1*; one (BAA96850) annotated as *Cry3* falls in the clade with vertebrate *Cry2*, one (NP_571862) is *Cry4*, one (NP_991249) is *Cry-DASH* and one (XP_009291670) is plant-like *Cry* [70]. Recently, a plant-like *Cry* was identified in *D. rerio* (XP_009291670), which may act as a circadian photoreceptor [70]. The maximum number of cryptochrome gene copies in other animal taxa (including birds, reptiles and amphibians) is usually 2 or 3. The extra *Cry* homologs in zebra fish were likely generated by ancient polyploidy events [71]. In the evolution history of vertebrates, their genomes were duplicated twice (occurred in the early evolution of deuterostomes), and a third genome duplication, which is named the fish-specific genome duplication (FSGD), occurred in the basal group of ray-finned fishes (Actinopterygii) (~ 350 Mya) [72]. The *Cry4* genes are apparently evolutionarily oldest among animal cryptochromes. A primitive Cephalochordate *Branchiostoma floridae* possesses homologs of *Cry4* (XP_002609503 and XP_002595074), thus suggesting that *Cry4* might emerge in Chordata, and then be lost in mammals. However, CRY4 was reported to exhibit neither (6–4) PHR nor circadian functions [73,74].

Cryptochromes of *Drosophila* (insect CRY1) and plants exhibit similar physiological functions, both of them play roles in light sensing and entrainment of circadian oscillator, but are not essential for the self-regulation of the clock [75,76]. However, the evolutionary history of plant and animal cryptochromes is quite different: they evolved from the different ancestral photolyase genes soon after the plant—animal divergence [77]. Animal cryptochromes originated from (6–4) photolyases, while plant cryptochromes are evolved from CPD photolyases

(Fig 2). Given that cryptochromes (except for CRY-DASH) are absent in fungi and prokaryotes, this may suggest that the plant and animal cryptochromes might appear soon after the origin of these taxa [77]. The origin of animal *Cry* was accompanied by coevolution of other circadian components including PER, BMAL1, CYC and CLK [77].

Photolyases, including CPD PHR, (6–4) PHR and single-strand photolyase CRY-DASH, were lost in higher animals, including placental mammals (Fig 2). The loss of PHR genes in higher animals was thought to occur due to weak natural selection [5]. In the absence of photolyases, placental mammals rely on more complex and less efficient nucleotide excision system consisting of DNA glycosylases, nucleases and DNA polymerases to eliminate pyrimidine dimers [5,78]. It was hypothesized that the loss of photolyases in higher eukaryotes was associated with the reduced UV stress [5]. Strong UV radiation was diminished by accumulation of oxygen since Proterozoic (~ 2500–540 Mya) [79].

Although eukaryotes and some prokaryotes (cyanobacteria) display circadian rhythms, the input signals to the clock are not always controlled by cryptochromes. In plants, circadian rhythms are activated through the red light receptor phytochrome (PHY) and blue light receptor cryptochrome [80]. The generic feedback loop of circadian rhythm in fungi is different: blue light is absorbed by the flavoprotein white collar-1 (WC-1) [81]. Cyanobacteria are the simplest organisms to exhibit circadian rhythmicity [82]; an environmental signal is transduced to the endogenous clock by the circadian input kinase A (CikA) [83], and the whole circadian system is controlled by the *kaiABC* gene cluster [84]. There had been no evidence for common ancestor of eukaryotic and prokaryotic circadian genes [85], until it was found that the cryptochromes have a common ancestor with the prokaryotic photolyase [17]. However, it is still unclear what circadian function is performed by CRY-DASH in prokaryotes. Another open question is how the distinctive circadian mechanisms emerged in different groups of eukaryotes. This prompts for further extensive studies of circadian genes in prokaryotes and eukaryotes to solve the above problems.

Episodic positive selection and conservation of photolyase and cryptochrome

The results of the selection analysis are in an agreement with the previously reported data, indicating that CRY/PHR genes have evolved primarily under strong purifying selection (Table 2) [5,86]. On the other hand, positive selection likely occurred during some periods of cryptochromes' evolution: the members of the superfamily experienced multiple duplications and neofunctionalizations, which are usually accompanied by strong episodic positive selection [87]. The site-specific models are conservative in estimating positive selection, especially for protein superfamilies with long evolutionary histories [88]. Therefore, we utilized the branch-site model to detect episodic diversifying selection.

Based on the results of REL analysis, episodic diversifying selection was apparently quite common in the evolution of the PHR/CRY superfamily (S3 Table and S1 Fig). In most cases, episodic diversifying selection operated on speciation events, e.g., plant CRYs of Angiospermae. Angiosperms are an evolutionary young (~179–158 Mya) and species-rich group, which experienced fast diversification and dominated almost all environments on Earth [89]. The speciation processes were likely followed by episodic diversifying selection in circadian genes, which might help these species to adapt to various ecological niches. Similar scenario was reported for molecular evolution of other eukaryotic circadian genes, particularly plant phytochromes (PHY) [90]. This might imply coevolution of cryptochromes and phytochromes. Indeed, phytochromes directly or indirectly interact with CRY and perform similar circadian functions (red/far-red light receptors), thus PHY and CRY might evolve under similar selection pressure [91]. On the

other hand, strong positive selection played a role in functional divergence of PHR/CRY proteins, e.g., Node 320 (divergence of CRY-DASH and plant PHR2) (S1 Fig).

The overall sequence similarity between the two homologous domains, DNA photolyase related domain (positions 3–159 of the alignment) and FAD binding domain (position 201–477), is generally high (Fig 3). The observed high conservation of PHR/CRY sequences corresponds to the fact that PHR and CRY proteins maintain a conserved 3D structure [11]. On the other hand, only two conserved residues shared by all subfamilies (Fig 3). One of them, Arg, is located at position 12 of the alignment, which was reported within the binding site of Cl^- in *Arabidopsis* CRY-DASH (R51 of PDB: 2VTB) [92]. As to S241, no functional or structural importance of this residue was reported previously. S241 is adjacent to α helices; we hypothesize that this residue is essential for maintaining the secondary structure of α helix and, respectively, for FAD binding (e.g., S260 and α14 of *Arabidopsis* CRY1, PDB: 1U3D [93]). In addition, the FAD binding domains of plant PHR2 are shorter and less conserved as compare to the other subfamilies. It is yet to be studied how plant PHR2 maintains its function with such an "incomplete" domain.

The key events in CRY/PHR evolution and major changes in the global light regime

The phylogenetic analysis along with the biostratigraphy made it possible to estimate the approximate time of key events in evolution of the CRY/PHR superfamily. According to the time estimates based on the ML and Bayesian approaches, plant, insect and vertebrate cryptochromes originated in Neoproterozoic Era (~1000–541 Mya). Studies of fossil record and geological patterns suggest that, during that period, the day length steadily increased from 18 h at 900 Mya [94] to 21 h at 600 Mya [95] and to approximately 22 hours at 650–600 Mya [96]. An ancestor of the vertebrate *Cry1* and *Cry2* duplicated in Silurian-Devonian Period (~443.4–358.9 Mya) (Table 3 and Fig 4). At that time, the climate of Earth became stable and warm, the concentration of oxygen increased and the level of the harmful UV radiation lowered [97]. New groups of living organisms evolved and spread in this era, such as lobe-finned fish and amphibians [98]. As one of the main adaptation mechanisms, the endogenous circadian system increases Darwinian fitness through synchronizing the metabolic and other biological rhythms of an organism with environmental light/dark cycle [99]. Therefore, the circadian system in eukaryotes might have experienced certain evolutionary changes to adjust to the increase of day length. These changes might include, among the others, functionally important substitutions in the existing circadian genes, the origin of clock genes *de novo* or co-option of non-circadian genes to perform circadian function. In these terms, the general direction of the circadian system evolution in eukaryotes is similar to that in prokaryotes (cyanobacteria) [100]. Furthermore, the time estimates for the major events in evolution of cryptochromes and the cyanobacterial circadian system [101] correspond to each other and to the major environmental changes in the global light regime. For example, the origin of animal cryptochromes (elements of the circadian input pathway) occurred about the same time (~700–600 Mya) when the *bona fide* circadian system of some cyanobacteria lost *kaiA*, also an important element of the circadian input [101], In turn, this period corresponds to the suggested upper time limit of the last of the three periods proposed to describe the role of UV radiation in the evolution of cyanobacteria [102].

Circadian rhythmicity and photo-activated DNA repair were suggested to have a common evolutionary origin [103]. Escape from sunlight represented a major selective force for development of circadian rhythms [104]. Geological studies indicated that in Precambrian times (~3800–544 Mya) atmosphere contained little oxygen and primitive organisms were exposed

to high ultraviolet radiation during the daytime [105]. There are 2 main strategies for organisms to avoid the harmful effects of UV radiation [6]. The first one is repairing the UV-induced DNA damage which is the physiological function of photolyase. The other one is to avoid being irradiated, such as migrate to deeper water. These movements were observed by the diel vertical migrations of zooplankton, which initiated and controlled by light [106]. Such migrations also occur in other marine and freshwater organisms such as water flea *Daphnia magna* [107], and sensitivity was related closely to the UV photoreceptors in its compound eye [107]. These diel vertical migrations may help to understand the coevolution of photoreception and circadian rhythms, and the coevolution of their respective controlling genes [6].

Supporting Information

S1 Fig. Unrooted maximum-likelihood tree of the CRY/PHR homologs. Maximum-likelihood probabilities of the node support below 0.5 are not shown. The values are the likelihoods and bootstraps respectively. Internal branch marks in blue font (e.g. Node14) correspond to branches evolved under episodic diversifying selection; marks in violet font (e.g. Node6) represent positively selected branches (mean $\omega > 1$) (S3 Table). The bar below the tree indicates 0.5 substitutions per site.
(JPG)

S2 Fig. Unrooted Bayesian tree of the CRY/PHR homologs. The node labels are the posterior values. The branches width and color correspond to the substitution rate. The bar below the tree indicates 0.5 substitutions per site. Relative evolutionary rates are illustrated by color and width of the branches. Thinner-thicker and lighter-darker indicate lower-higher rates, respectively.
(JPG)

S1 Table. List of the CRY and PHR sequences used in the study.
(DOCX)

S2 Table. List of the RPB2 sequences used in the study.
(DOCX)

S3 Table. Results of episodic diversifying selection test by REL. Mean ω was obtained using local MG94 model (no site-to-site variation). Values of ω^+ and weight ω^+ reflect the strength of selection and the proportion of the total branch length of positive selection. Estimates of the uncorrected *p*-value were generated by the mixture of distributions, corrected *p* was the probability obtained after Holm's correction for multiple testing. Branches with mean $\omega > 1$ are shaded. Infinite mean ω for branches may be resulted by lack of synonymous substitution. As a result, they were not considered to be positively selected branches.
(DOCX)

Author Contributions

Conceived and designed the experiments: VD. Performed the experiments: QM. Analyzed the data: QM. Contributed reagents/materials/analysis tools: VD QM. Wrote the paper: VD QM.

References

1. Selby CP, Sancar A (2006) A cryptochrome/photolyase class of enzymes with single-stranded DNA-specific photolyase activity. Proceedings of the National Academy of Sciences of the United States of America 103: 17696–17700. PMID: 17062752

2. Essen LO, Klar T (2006) Light-driven DNA repair by photolyases. Cellular and Molecular Life Sciences 63: 1266–1277. PMID: 16699813
3. Brudler R, Hitomi K, Daiyasu H, Toh H, Kucho K, Ishiura M, et al. (2003) Identification of a new cryptochrome class. Structure, function, and evolution. Molecular Cell 11: 59–67. PMID: 12535521
4. Castrillo M, Garcia-Martinez J, Avalos J (2013) Light-dependent functions of the *Fusarium fujikuroi* CryD DASH cryptochrome in development and secondary metabolism. Applied and Environmental Microbiology 79: 2777–2788. doi: 10.1128/AEM.03110-12 PMID: 23417004
5. Lucas-Lledo JI, Lynch M (2009) Evolution of mutation rates: phylogenomic analysis of the photolyase/cryptochrome family. Molecular Biology and Evolution 26: 1143–1153. doi: 10.1093/molbev/msp029 PMID: 19228922
6. Gehring W, Rosbash M (2003) The coevolution of blue-light photoreception and circadian rhythms. Journal of Molecular Evolution 57 Suppl 1: S286–289. PMID: 15008426
7. Bell-Pedersen D, Cassone VM, Earnest DJ, Golden SS, Hardin PE, Thomas TL, et al. (2005) Circadian rhythms from multiple oscillators: lessons from diverse organisms. Nature Reviews Genetics 6: 544–556. PMID: 15951747
8. Paranjpe DA, Sharma VK (2005) Evolution of temporal order in living organisms. Journal of Circadian Rhythms 3: 7. PMID: 15869714
9. Crosthwaite SK, Loros JJ, Dunlap JC (1995) Light-induced resetting of a circadian clock is mediated by a rapid increase in frequency transcript. Cell 81: 1003–1012. PMID: 7600569
10. Wenderoth N, Bock O (1999) Load dependence of simulated central tremor. Biological Cybernetics 80: 285–290. PMID: 10326243
11. Lin C, Todo T (2005) The cryptochromes. Genome Biology 6: 220. PMID: 15892880
12. Emery P, So WV, Kaneko M, Hall JC, Rosbash M (1998) CRY, a *Drosophila* clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. Cell 95: 669–679. PMID: 9845369
13. Stanewsky R, Kaneko M, Emery P, Beretta B, Wager-Smith K, Kay SA, et al. (1998) The *cry*^{b} mutation identifies cryptochrome as a circadian photoreceptor in *Drosophila*. Cell 95: 681–692. PMID: 9845370
14. Thresher RJ, Vitaterna MH, Miyamoto Y, Kazantsev A, Hsu DS, Petit C, et al. (1998) Role of mouse cryptochrome blue-light photoreceptor in circadian photoresponses. Science 282: 1490–1494. PMID: 9822380
15. Cashmore AR (2003) Cryptochromes: enabling plants and animals to determine circadian time. Cell 114: 537–543. PMID: 13678578
16. Thompson CL, Sancar A (2002) Photolyase/cryptochrome blue-light photoreceptors use photon energy to repair DNA and reset the circadian clock. Oncogene 21: 9043–9056. PMID: 12483519
17. Ahmad M, Cashmore AR (1993) HY4 gene of *A. thaliana* encodes a protein with characteristics of a blue-light photoreceptor. Nature 366: 162–166. PMID: 8232555
18. Todo T (1999) Functional diversity of the DNA photolyase/blue light receptor family. Mutation Research 434: 89–97. PMID: 10422537
19. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Research 39: D225–229. doi: 10.1093/nar/gkq1189 PMID: 21109532
20. Chaves I, Yagita K, Barnhoorn S, Okamura H, van der Horst GT, Tamanini F. (2006) Functional evolution of the photolyase/cryptochrome protein family: importance of the C terminus of mammalian CRY1 for circadian core oscillator performance. Molecular and Cellular Biology 26: 1743–1753. PMID: 16478995
21. Yang HQ, Wu YJ, Tang RH, Liu D, Liu Y, Cashmore AR. (2000) The C termini of *Arabidopsis* cryptochromes mediate a constitutive light response. Cell 103: 815–827. PMID: 11114337
22. Malhotra K, Kim ST, Batschauer A, Dawut L, Sancar A (1995) Putative blue-light photoreceptors from *Arabidopsis thaliana* and *Sinapis alba* with a high degree of sequence homology to DNA photolyase contain the two photolyase cofactors but lack DNA repair activity. Biochemistry 34: 6892–6899. PMID: 7756321
23. Park HW, Kim ST, Sancar A, Deisenhofer J (1995) Crystal structure of DNA photolyase from *Escherichia coli*. Science 268: 1866–1872. PMID: 7604260
24. Kanai S, Kikuno R, Toh H, Ryo H, Todo T (1997) Molecular evolution of the photolyase-blue-light photoreceptor family. Journal of Molecular Evolution 45: 535–548. PMID: 9342401
25. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403–410. PMID: 2231712

26. Sidow A, Thomas WK (1994) A molecular evolutionary framework for eukaryotic model organisms. Current Biology 4: 596–603. PMID: 7953533

27. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792–1797. PMID: 15034147

28. Wernersson R, Pedersen AG (2003) RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. Nucleic Acids Research 31: 3537–3539. PMID: 12824361

29. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic acids symposium series 41: 95–98.

30. Akaike H (1981) A new look at the statistical model identification. IEEE Transactions on Automatic Control 19: 716–723.

31. Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27: 1164–1165. doi: 10.1093/bioinformatics/btr088 PMID: 21335321

32. Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. Molecular Biology and Evolution 25: 1307–1320. doi: 10.1093/molbev/msn067 PMID: 18367465

33. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52: 696–704. PMID: 14530136

34. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Systematic Biology 55: 539–552. PMID: 16785212

35. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7.

36. Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. Proceedings of the Royal Society of London Series B: Biological Sciences 256: 119–124. PMID: 8029240

37. Kosakovsky Pond SL, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics 21: 676–679. PMID: 15509596

38. Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Molecular Biology and Evolution 11: 725–736. PMID: 7968486

39. Nielsen R, Yang ZH (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148: 929–936. PMID: 9539414

40. Yang Z (2000) Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. Journal of Molecular Evolution 51: 423–432. PMID: 11080365

41. Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. Molecular Biology and Evolution 22: 1107–1118. PMID: 15689528

42. Kosakovsky Pond SL, Murrell B, Fourment M, Frost SDW, Delport W, Scheffler K. (2011) A random effects branch-site model for detecting episodic diversifying selection. Molecular Biology and Evolution 28: 3033–3043. doi: 10.1093/molbev/msr125 PMID: 21670087

43. Cadar D, Csagola A, Kiss T, Tuboly T (2013) Capsid protein evolution and comparative phylogeny of novel porcine parvoviruses. Molecular Phylogenetics and Evolution 66: 243–253. doi: 10.1016/j.ympev.2012.09.030 PMID: 23044400

44. de Oliveira AS, Melo FL, Inoue-Nagata AK, Nagata T, Kitajima EW, Resende RO. (2012) Characterization of bean necrotic mosaic virus: a member of a novel evolutionary lineage within the genus *Tospovirus*. PLOS One 7: e38634. doi: 10.1371/journal.pone.0038634 PMID: 22715400

45. Lau SK, Li KS, Tsang AK, Shek CT, Wang M, Choi GK, et al. (2012) Recent transmission of a novel alphacoronavirus, bat coronavirus HKU10, from Leschenault's rousettes to pomona leaf-nosed bats: first evidence of interspecies transmission of coronavirus between bats of different suborders. Journal of Virology 86: 11906–11918. doi: 10.1128/JVI.01305-12 PMID: 22933277

46. Graur D, Li WH (2000) Fundamentals of molecular evolution, 2nd edition. Sunderland, MA: Sinauer Associates. 81–81 p.

47. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Research 38: W529–W533. doi: 10.1093/nar/gkq399 PMID: 20478830

48. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. (2004) UCSF chimera—a visualization system for exploratory research and analysis. Journal of Computational Chemistry 25: 1605–1612. PMID: 15264254

49. Benton MJ, Donoghue PCJ (2007) Paleontological evidence to date the tree of life. Molecular Biology and Evolution 24: 26–53. PMID: 17047029

50. Sanderson MJ, Thorne JL, Wikstrom N, Bremer K (2004) Molecular evidence on plant divergence times. American Journal of Botany 91: 1656–1665. doi: 10.3732/ajb.91.10.1656 PMID: 21652315

51. Lucking R, Huhndorf S, Pfister DH, Plata ER, Lumbsch HT (2009) Fungi evolved right on track. Mycologia 101: 810–822. PMID: 19927746

52. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24: 1586–1591. PMID: 17483113

53. Posada D (2008) jModelTest: Phylogenetic model averaging. Molecular Biology and Evolution 25: 1253–1256. doi: 10.1093/molbev/msn083 PMID: 18397919

54. Rodriguez F, Oliver JL, Marin A, Medina JR (1990) The general stochastic model of nucleotide substitution. Journal of Theoretical Biology 142: 485–501. PMID: 2338834

55. Chaves I, Nijman RM, Biernat MA, Bajek MI, Brand K, da Silva AC, et al. (2011) The *Potorous* CPD photolyase rescues a cryptochrome-deficient mammalian circadian clock. PLOS One 6.

56. Petersen JL, Lang DW, Small GD (1999) Cloning and characterization of a class II DNA photolyase from *Chlamydomonas*. Plant Molecular Biology 40: 1063–1071. PMID: 10527430

57. Chen JJ, Jiang CZ, Britt AB (1996) Little or no repair of cyclobutyl pyrimidine dimers is observed in the organellar genomes of the young *Arabidopsis* seedling. Plant Physiology 111: 19–25. PMID: 12226273

58. Takahashi M, Teranishi M, Ishida H, Kawasaki J, Takeuchi A, Yamaya T, et al. (2011) Cyclobutane pyrimidine dimer (CPD) photolyase repairs ultraviolet-B-induced CPDs in rice chloroplast and mitochondrial DNA. Plant J 66: 433–442. doi: 10.1111/j.1365-313X.2011.04500.x PMID: 21251107

59. Huang Y, Baxter R, Smith BS, Partch CL, Colbert CL, Deisenhofer J. (2006) Crystal structure of cryptochrome 3 from Arabidopsis thaliana and its implications for photolyase activity. Proceedings of the National Academy of Sciences of the United States of America 103: 17701–17706. PMID: 17101984

60. Zhu HS, Yuan Q, Briscoe AD, Froy O, Casselman A, Reppert SM. (2005) The two CRYs of the butterfly. Current Biology 15: R953–R954. PMID: 16332522

61. VanVickle-Chavez SJ, Van Gelder RN (2007) Action spectrum of *Drosophila* cryptochrome. Journal of Biological Chemistry 282: 10561–10566. PMID: 17284451

62. McDonald MJ, Rosbash M (2001) Microarray analysis and organization of circadian gene expression in *Drosophila*. Cell 107: 567–578. PMID: 11733057

63. Shirai H, Oishi K, Ishida N (2006) Bidirectional CLOCK/BMAL1-dependent circadian gene regulation by retinoic acid in vitro. Biochemical and Biophysical Research Communications 351: 387–391. PMID: 17069763

64. Langmesser S, Tallone T, Bordon A, Rusconi S, Albrecht U (2008) Interaction of circadian clock proteins PER2 and CRY with BMAL1 and CLOCK. BMC Molecular Biology 9: 41. doi: 10.1186/1471-2199-9-41 PMID: 18430226

65. Rutila JE, Suri V, Le M, So WV, Rosbash M, Hall JC. (1998) CYCLE is a second bHLH-PAS clock protein essential for circadian rhythmicity and transcription of *Drosophila* period and timeless. Cell 93: 805–814. PMID: 9630224

66. Bae K, Lee C, Sidote D, Chuang KY, Edery I (1998) Circadian regulation of a *Drosophila* homolog of the mammalian clock gene: PER and TIM function as positive regulators. Molecular and Cellular Biology 18: 6142–6151. PMID: 9742131

67. Jin X, Shearman LP, Weaver DR, Zylka MJ, de Vries GJ, Reppert SM. (1999) A molecular mechanism regulating rhythmic output from the suprachiasmatic circadian clock. Cell 96: 57–68. PMID: 9989497

68. McClung CR (2006) Plant circadian rhythms. Plant Cell 18: 792–803. PMID: 16595397

69. Yuan Q, Metterville D, Briscoe AD, Reppert SM (2007) Insect cryptochromes: gene duplication and loss define diverse ways to construct insect circadian clocks. Molecular Biology and Evolution 24: 948–955. PMID: 17244599

70. Oliveri P, Fortunato AE, Petrone L, Ishikawa-Fujiwara T, Kobayashi Y, Todo T, et al. (2014) The cryptochrome/photolyase family in aquatic organisms. Marine Genomics 14: 23–37. doi: 10.1016/j.margen.2014.02.001 PMID: 24568948

71. Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. PLOS Biology 3: e314. PMID: 16128622

72. Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y (2003) Genome duplication, a trait shared by 22,000 species of ray-finned fish. Genome Research 13: 382–390. PMID: 12618368

73. Liu C, Hu J, Qu C, Wang L, Huang G, Niu P, et al. (2015) Molecular evolution and functional divergence of zebrafish (*Danio rerio*) cryptochrome genes. Scientific Reports 5: 8113. doi: 10.1038/srep08113 PMID: 25630924

74. Kobayashi Y, Ishikawa T, Hirayama J, Daiyasu H, Kanai S, Toh H, et al. (2000) Molecular analysis of zebrafish photolyase/cryptochrome family: two types of cryptochromes present in zebrafish. Genes to Cells 5: 725–738. PMID: 10971654

75. Ceriani MF, Darlington TK, Staknis D, Mas P, Petti AA, Weitz CJ, et al. (1999) Light-dependent sequestration of TIMELESS by CRYPTOCHROME. Science 285: 553–556. PMID: 10417378

76. Devlin PF, Kay SA (2000) Cryptochromes are required for phytochrome signaling to the circadian clock but not for rhythmicity. Plant Cell 12: 2499–2509. PMID: 11148293

77. Cashmore AR, Jarillo JA, Wu YJ, Liu D (1999) Cryptochromes: blue light receptors for plants and animals. Science 284: 760–765. PMID: 10221900

78. Wood RD, Mitchell M, Sgouros J, Lindahl T (2001) Human DNA repair genes. Science 291: 1284–1289. PMID: 11181991

79. Canfield DE, Habicht KS, Thamdrup B (2000) The Archean sulfur cycle and the early history of atmospheric oxygen. Science 288: 658–661. PMID: 10784446

80. Somers DE, Devlin PF, Kay SA (1998) Phytochromes and cryptochromes in the entrainment of the *Arabidopsis* circadian clock. Science 282: 1488–1490. PMID: 9822379

81. Bell-Pedersen D, Garceau N, Loros JJ (1996) Circadian rhythms in fungi. Journal of Genetics 75: 387–401.

82. Kondo T, Ishiura M (2000) The circadian clock of cyanobacteria. Bioessays 22: 10–15. PMID: 10649285

83. Schmitz O, Katayama M, Williams SB, Kondo T, Golden SS (2000) CikA, a bacteriophytochrome that resets the cyanobacterial circadian clock. Science 289: 765–768. PMID: 10926536

84. Ishiura M, Kutsuna S, Aoki S, Iwasaki H, Andersson CR, Tanabe A, et al. (1998) Expression of a gene cluster kaiABC as a circadian feedback process in cyanobacteria. Science 281: 1519–1523. PMID: 9727980

85. Rosbash M (2009) The implications of multiple circadian clock origins. PLOS Biology 7: 421–425.

86. Layeghifard M, Rabani R, Pirhaji L, Yakhchali B (2008) Evolutionary mechanisms underlying the functional divergence of duplicate genes involved in vertebrates' circadian rhythm pathway. Gene 426: 65–71. doi: 10.1016/j.gene.2008.08.014 PMID: 18804153

87. Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Molecular Biology and Evolution 15: 568–573. PMID: 9580986

88. Lu A, Guindon S (2014) Performance of standard and stochastic branch-site models for detecting positive selection among coding sequences. Molecular Biology and Evolution 31: 484–495. doi: 10.1093/molbev/mst198 PMID: 24132121

89. Wikstrom N, Savolainen V, Chase MW (2001) Evolution of the angiosperms: calibrating the family tree. Proceedings of the Royal Society B: Biological Sciences 268: 2211–2220. PMID: 11674868

90. Mei QM, Dvornyk V (2014) Evolution of PAS domains and PAS-containing genes in eukaryotes. Chromosoma 123: 385–405. doi: 10.1007/s00412-014-0457-x PMID: 24699836

91. Wade HK, Bibikova TN, Valentine WJ, Jenkins GI (2001) Interactions within a network of phytochrome, cryptochrome and UV-B phototransduction pathways regulate chalcone synthase gene expression in Arabidopsis leaf tissue. Plant Journal 25: 675–685. PMID: 11319034

92. Pokorny R, Klar T, Hennecke U, Carell T, Batschauer A, Essen LO. (2008) Recognition and repair of UV lesions in loop structures of duplex DNA by DASH-type cryptochrome. Proceedings of the National Academy of Sciences of the United States of America 105: 21023–21027. doi: 10.1073/pnas.0805830106 PMID: 19074258

93. Brautigam CA, Smith BS, Ma ZQ, Palnitkar M, Tomchick DR, Machius M, et al. (2004) Structure of the photolyase-like domain of cryptochrome 1 from *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America 101: 12142–12147. PMID: 15299148

94. Sonett CP, Kvale EP, Zakharian A, Chan MA, Demko TM (1996) Late Proterozoic and Paleozoic tides, retreat of the moon, and rotation of the earth. Science 273: 100–104. PMID: 8688061

95. Zahnle K, Walker JC (1987) A constant daylength during the Precambrian era? Precambrian Research 37: 95–105. PMID: 11542096

96. Schmidt PW, Williams GE (1995) The Neoproterozoic climatic paradox: equatorial palaeolatitude for Marinoan glaciation near sea level in South Australia. Earth and Planetary Science Letters 134: 107–124.

97. Canfield DE (2005) The early history of atmospheric oxygen: Homage to Robert A. Garrels. Annual Review of Earth and Planetary Sciences 33: 1–36.

98. Niedzwiedzki G, Szrek P, Narkiewicz K, Narkiewicz M, Ahlberg PE (2010) Tetrapod trackways from the early Middle Devonian period of Poland. Nature 463: 43–48. doi: 10.1038/nature08623 PMID: 20054388

99. Dodd AN, Salathia N, Hall A, Kevei E, Toth R, Nagy F, et al. (2005) Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. Science 309: 630–633. PMID: 16040710

100. Dvornyk V (2009) The circadian clock gear in cyanobacteria: assembled by evolution. In: Ditty JL, Mackey SR, Johnson CH, editors. Bacterial circadian programs: Springer Berlin Heidelberg. pp. 241–258.

101. Baca I, Sprockett D, Dvornyk V (2010) Circadian input kinases and their homologs in cyanobacteria: evolutionary constraints versus architectural diversification. Journal of Molecular Evolution 70: 453–465. doi: 10.1007/s00239-010-9344-0 PMID: 20437037

102. Garcia-Pichel F (1998) Solar ultraviolet and the evolutionary history of cyanobacteria. Origins of Life and Evolution of the Biosphere 28: 321–347. PMID: 9611770

103. Sancar A (2000) Cryptochrome: The second photoactive pigment in the eye and its role in circadian photoreception. Annual Review of Biochemistry 69: 31–67. PMID: 10966452

104. Pittendrigh CS (1993) Temporal organization: reflections of a Darwinian clock-watcher. Annual Review of Physiology 55: 16–54. PMID: 8466172

105. Schopf JW, Hayes JM, Walter MR, editors (1983) Evolution of the earth's earliest ecosystems: recent progress and unsolved problems. Princeton, NJ Priceton University Press.

106. Nybakken JW (2001) Marine biology: An ecological approach 5th ed: Benjamin Cummings.

107. Smith KC, Macagno ER (1990) UV photoreceptors in the compound eye of *Daphnia magna* (Crustacea, Branchiopoda). A fourth spectral class in single ommatidia. Journal of Comparative Physiology A 166: 597–606.