The diversity and evolution of circadian clock proteins in fungi

Leonidas Salichos
Antonis Rokas

Department of Biological Sciences, Vanderbilt University, VU Station B 35-1634, Nashville, Tennessee 37235

Abstract: Circadian rhythms are endogenous cellular patterns that associate multiple physiological and molecular functions with time. The *Neurospora* circadian system contains at least three oscillators: the FRQ/WC-dependent circadian oscillator (FWO), whose core components are the FRQ, WC-1, WC-2, FRH, and FWD-1 proteins; the WC-dependent circadian oscillator (WC-FLO); and one or more FRQ/WC-independent oscillators (FLO). Little is known about the distribution of homologs of the *Neurospora* clock proteins or about the molecular foundations of circadian rhythms across fungi. Here, we examined 64 diverse fungal proteomes for homologs of all five *Neurospora* clock proteins and retraced their evolutionary history. The FRH and FWD-1 proteins were likely present in the fungal ancestor. WC-1 and WC-2 homologs are absent from the early diverging chytrids and Microsporidia but are present in all other major clades. In contrast to the deep origins of these four clock proteins FRQ homologs are taxonomically restricted within Sordariomycetes, Leotiomycetes and Dothideomycetes. The large number of FRH and FWD-1 homologs identified and their lack of concordance with the fungal species phylogeny indicate that they likely underwent multiple rounds of duplications and losses. A notable exception is the 10 FRQ-like proteins in *Fusarium oxysporum*, which resulted from nine duplication events. Our results suggest that the machinery required for FWO oscillator function is taxonomically restricted within Ascomycetes. Although the WC proteins are widely distributed, the functional diversity of the few non-*Neurospora* circadian oscillators suggests that a WC-FLO oscillator is unlikely to fully explain the observed rhythms. The contrast between the diversity of circadian oscillators and the conservation of most of their machinery is likely best explained by considering the centrality of noncircadian functions in which RNA helicase (FRH), F-box (FWD-1), WC-1 and WC-2 (light-sensing) proteins participate in fungi and eukaryotes.

Key words: circadian rhythm, homology, FRQ/WC-based oscillator (FWO), FRQ-less oscillator (FLO), *Neurospora*, phylogeny, WC-FLO

INTRODUCTION

Fungi contain a number of cellular mechanisms that enable them to track time and adjust to environmental changes (Bell-Pedersen, Garceau, Loros 1996). Many of these rhythms are not dependent on environmental cues (Murray et al. 2001, Murray, Klevecz, Lloyd 2003, Dunlap and Loros 2006, Tu and McKnight 2007, Lloyd 2008) and are typically limited to a single or a few species (Ingold 1971, Bunning and Moser 1973). In contrast, rhythms driven by endogenous circadian clocks are present across life, thus enabling organisms to anticipate regular environmental changes and adjust their biological activities accordingly. A rhythm can be classified as circadian if it satisfies three basic criteria: (i) it has a ∼24 h period, (ii) its period length is temperature compensated and (iii) it can be reset by environmental cues such as light and temperature (Liu and Bell-Pedersen 2006).

degradation of FRQ results in the release of the WWC inhibition and thus frq reactivation.


Rhythmic events independent of FRQ also have been identified in *Neurospora*, suggesting the existence of additional FRQ-less oscillators (FLO) (Loros and Feldman 1986, Aronson et al. 1994, Luo, Loros, Dunlap 1998, Greene et al. 2003, Dunlap and Loros 2004, Lakin-Thomas and Brody 2004, de Paula et al. 2006, Liu and Bell-Pedersen 2006). For example, both the activity of nitrate reductase (Christensen et al. 2004) and levels of diacylglycerol (Ramsdale and Lakin-Thomas 2000) have been shown to oscillate in frq mutant and null strains. Although the molecular components of several of these FLO are largely unknown (Liu and Bell-Pedersen 2006), their function is FRQ/WC independent (Christensen et al. 2004). Furthermore, recent work has identified a novel FLO circadian oscillator that requires WC-1 and WC-2 but not FRQ (Correa et al. 2005, de Paula et al. 2006) (WC-FLO). Thus, the *Neurospora* system contains at least three distinct oscillators: FWO, WC-FLO and one or more FRQ/WC-independent FLO (de Paula et al. 2006).

Circadian rhythms have been shown in few representatives from most fungal phyla (Ascomycetes, Basidiomycetes, Zygomycetes), but in contrast to the well studied *Neurospora*, relatively little is known about their molecular foundations (Bell-Pedersen, Garceau, Loros 1996). One approach to better understand circadian rhythms in fungi is to examine the distribution and evolutionary history of the known key *Neurospora* circadian clock proteins across the kingdom (Merrow and Dunlap 1994, Lewis and Feldman 1996, Lewis, Morgan, Feldman 1997, Lombardi and Brody 2005, Dunlap and Loros 2006) and functional complementation tests with these homologs have been performed (Merrow and Dunlap 1994), but a systematic evolutionary analysis across the fungal kingdom of all *Neurospora* clock components so far has been lacking. Here, we investigate the presence of homologs of five *Neurospora* circadian clock proteins (FRQ, WC-1, WC-2, FRH and FWD-1) in the proteomes of 64 fungi spanning the diversity of the lineage. For each protein, we examine the distribution of its homologs and functional domains across fungal proteomes and reconcile its evolutionary history within the context of fungal evolution. Finally, we discuss the implications of our findings for understanding the evolution of the molecular basis of circadian rhythms in Fungi.

MATERIALS AND METHODS

Our dataset consisted of 666,385 proteins from 64 fungal proteomes spanning the fungal kingdom. The circadian clock proteins from *Neurospora crassa* were retrieved from GenBank (accession numbers: FRQ: XP_959909, 989 amino acids; WC-1: CAA63964, 1167 amino acids; WC-2: CAA70336, 530 amino acids; FRH: AAW32908, 1106 amino acids; FWD-1: AAT94285, 1010 amino acids).

To identify fungal homologs we performed a similarity search of the *Neurospora* circadian clock proteins across the 64 fungal proteomes with the BLAST algorithm (Altschul et al. 1997). We selected homologous sequences by three filtering criteria: the BLAST e-value was less than or equal to $1 \times 10^{-5}$, the length of the recovered homolog was at least half the length of the query sequence and the BLAST length of the recovered homologous sequence was at least half the length of the query sequence. All sequences that satisfied these criteria were analyzed further. The values for our filtering criteria were chosen to be sufficiently lenient so as not to exclude any unambiguous homologs but sufficiently strict so as not to include proteins that were distantly related and of ambiguous homology. Distant or ambiguous homologs are difficult to verify and study in a phylogenetic framework because their similarity to the query sequences is low and they are difficult to align (see below).

To independently confirm our strategy for identifying homologs we also searched all fungal proteomes for the functional domains of each *Neurospora* circadian clock protein found in the PFAM database (Finn et al. 2006) with HMMER software (Eddy 1998). Those used were the FRQ domain (PF09421) contained within the FRQ protein, the DEAD (PF00270), Helicase_C (PF00271) and DSHCT (PF08148) domains contained within the FRH protein, the PAS_3 (PF08447) and GATA (PF00320) domains contained within each of the WC proteins, and the F-box (PF00646) and WD-40 (PF00400) domains contained within the FWD-1 protein. Proteins were considered as hits if they contained all domains from a given *Neurospora* circadian clock protein in the same relative order and the e-value of each domain hit was equal or less than $1 \times 10^{-5}$. 
To study the evolutionary history of the homologs of the *Neurospora* circadian clock proteins, we first aligned all sequences that met the filtering criteria with ClustalW (Thompson, Higgins, Gibson 1994). ClustalW produces multiple sequence alignments by calculating a distance score (rewarding amino acid matches and penalizing amino acid mismatches and gaps) between all pairs of sequences and used this distance score to progressively align sequence pairs of increasing distance (Thompson, Higgins, Gibson 1994). We next removed columns of uncertain alignment using the Gblocks software (Castresana 2000). Gblocks automatically excludes poorly aligned or divergent regions, a procedure that has been shown to improve phylogenetic accuracy (Talavera and Castresana 2007). We excluded amino acid columns in which more than half the sequences contained a gap, and left all other Gblocks parameters to their default values. We manually edited the alignment for the FWD-1 protein because removal of columns of uncertain alignment by Gblocks resulted in a short alignment (28 amino acid columns).

For the phylogenetic analyses we used the optimality criteria of maximum likelihood (ML) and Bayesian inference (BI), as implemented in PHYML (Guindon and Gascuel 2003) and MrBayes (Ronquist and Huelsenbeck 2003, Altekar et al. 2004), respectively. For the ML analysis we selected the parameters of the evolutionary models that best described the evolutionary history of each protein with ProtTest software (Abascal, Zardoya, Posada 2005), and assessed robustness in inference by analyzing 100 bootstrap replicates. For the BI analysis we ran two independent analyses assuming a mixture of amino acid substitution models. Each analysis was run with four chains (one cold and three hot) for 2000000 generations. Trees were sampled every 1000 generations and the first 2000 sampled trees were discarded as burn-in, by which point stationarity was already reached.

To estimate protein loss and gain events we reconciled each ML-generated protein phylogenetic tree with the fungal species tree with Notung software (Durand, Halldorsson, Vernot 2006, Vernot et al. 2008). Notung estimates losses and duplications under the parsimony optimality criterion and allows the fitting of a nonbinary protein tree to a binary (fully resolved) species tree. The latter capability is important because the limited amount of data contained in single gene phylogenies can result in spurious inference (Rokas et al. 2003). To mitigate the effects of spurious protein tree inference in our estimations of loss and duplication events we collapsed all clades in our protein trees that were supported by bootstrap values less than 70% (Hillis and Bull 1993) before reconciliation with the species tree. The species tree of the 64 fungal genomes was synthesized with information from several published multi-gene phylogenies (Fitzpatrick et al. 2006, Geiser et al. 2006, James et al. 2006) (SUPPLEMENTARY MATERIAL).

**RESULTS**

To identify the distribution of *Neurospora* clock protein homologs across fungi we performed a BLAST analysis on the 64 proteomes. Implementation of our three filtering criteria (BLAST e-value and alignment length, and homolog length) resulted in the identification of 314 homologs for all five circadian clock proteins. There were respectively 26, 40, 40, 98 and 110 homologs for FRQ, WC-1, WC-2, FRH and FWD-1 (Fig. 1). The distribution of protein homologs is largely consistent with the distribution of their protein domains across fungi (SUPPLEMENTARY MATERIAL).

Homologs of all five proteins appear to be present only in the Ascomycete clades of Sordariomycetes, Leotiomycetes and Dothideomycetes (Fig. 1). Homologs of WC-1, WC-2, FRH and FWD-1 are present in Ascomycetes, Basidiomycetes and Zygomycetes. In contrast, members of the Saccharomycetes clade as well as the single chytrid representative appear to contain only the FWD-1 and FRH protein homologs. The single Microsporidian contains only an FRH homolog (Fig. 1), although a protein with significant hits to the protein domains present in the FWD-1 protein also was detected in the domain analysis (SUPPLEMENTARY MATERIAL). WC-1 and WC-2 appear to have been lost in Saccharomycetes, although a WC-2 homolog is found in *Yarrowia lipolytica*, the earliest branching species of the clade (but note that the domain analysis failed to find significant hits to the WC-associated domains; SUPPLEMENTARY MATERIAL). The lack of WC-1 and WC-2 in chytrids and Microsporidia should be treated with caution due to the use of a single representative proteome from each of the two clades.

The protein phylogenies of FRQ, WC-1 and WC-2 were largely concordant with the fungal species phylogeny, in agreement with Lewis and Feldman (1996). However, whereas most species contain only one copy of the FRQ protein, *Fusarium oxysporum* appears to contain 10 intact and presumably functional copies (Fig. 2). Reconciliation of the FRQ tree with the species tree suggested the occurrence of at least nine duplications and two loss events (Fig. 3A). Eight of the nine duplication events appear to have occurred within the *Fusarium oxysporum* lineage, whereas the last one is inferred to have occurred at the ancestral branch of the *Fusarium-Nectria* clade, with two of the four species in the clade (*Fusarium verticillioides* and *Fusarium graminearum*) subsequently losing these duplicated copies (Fig. 3A). Of interest, the majority of the FRQ copies found in the *F. oxysporum* genome appear to be flanked by transposases and reverse transcription, suggesting that the duplications might have occurred via the action of transposable elements. For the WC-1 protein, reconciliation analysis identified seven duplications and 13 losses, whereas the same analysis for
Fig. 1. The distribution of the five proteins participating in the Neurospora circadian oscillator across 64 diverse fungal genomes and a proposed model for their evolution. The species phylogeny shows the classes of all 64 fungal species whose proteomes were analyzed in this study. The phylogeny was synthesized with information from several published multigene phylogenies (Fitzpatrick et al. 2006, Geiser et al. 2006, James et al. 2006) (SUPPLEMENTARY TABLE). The number
the WC-2 protein revealed six duplications and seven losses. Of note, a number of duplications and losses in both WC-1 and WC-2 appear to have taken place in the common ancestor of the Zygomycetes clade composed of *Rhizopus* and *Phycomyces* (Fig. 3B, C). In contrast to the FRQ, WC-1 and WC-2 proteins, the evolutionary history of the FWD-1 and FRH homologs is characterized by extensive duplications and losses. We estimated 20 duplications and 138 losses for FWD-1 and 13 duplications and 59 losses for FRH, including four basal duplications in FWD-1 and two duplications in FRH.

**DISCUSSION**

We have traced the evolution of homologs of the five *Neurospora* clock proteins across the fungal kingdom. Pinpointing exactly when some of the proteins likely originated is challenging due to the limited number of early branching fungal genomes available (e.g. only and distribution of FRQ, WC-1, WC-2, FRH and FWD-1 protein homologs across all fungal proteomes examined are shown on the table on the right of the phylogeny. FRH and FWD-1 are likely ancestral to all fungi, but FWD-1 subsequently was lost from Microsporidia, according to this model. The WC-1 and WC-2 proteins likely were gained in the ancestor of Zygomycetes, Basidiomycetes and Ascomycetes but subsequently were lost in the ancestor of the Saccharomycetes clade. The FRQ protein was likely gained last, in the common ancestor of Sordariomycetes, Leotiomycetes and Dothideomycetes.
Fig. 3. Reconciled trees depicting the evolutionary history of FRQ (panel A), WC-1 (panel B) and WC-2 (panel C) homologs. Filled black circles represent duplication events, while gray taxon labels represent loss events. Species names have been abbreviated to the acronyms (in Fig. 1). The identity of nodes where proteins losses are inferred by reconciliation analyses to have occurred are shown (Supplementary Fig. 1). Note that all branches with bootstrap support < 70% have been collapsed. The protein and species phylogenies used in these reconciliations are shown (Supplementary Material).
one Chytrid and one Microsporidian proteome are available). Taxon sampling aside, we have found that FRH was likely present in the fungal ancestor, a result consistent with the distribution of RNA helicases across eukaryotes (Dangel et al. 1995, Lee et al. 1995, Dunlap and Loros 2006). FWD-1 also is found across fungi, with the exception of Microsporidia (Fig. 1). However FWD-1 is a member of the F-box family of proteins, which is widely distributed across eukaryotes (Jonkers and Rep 2009). Thus, the lack of FWD-1 in Microsporidia is probably best explained as a lineage-specific loss (Fig. 1) instead of as a gain in a fungal ancestor following its divergence of Microsporidia from the rest of the fungal lineage. Additional support for a lineage-specific loss comes from consideration of the obligate intracellular parasitic lifestyle of Microsporidia and their highly reduced genome (Katinka et al. 2001) as well as from the detection of a protein with significant matches to the FWD-1 protein domains (SUPPLEMENTARY MATERIAL). WC-1 and WC-2 likely evolved in the common ancestor of Zygomycetes, Basidiomycetes and Ascomycetes, although they both subsequently were lost from Saccharomycetes. Of interest, FRQ was likely gained within Ascomycetes in the ancestor of Sordariomycetes, Leotiomycetes and Dothideomycetes, well after the emergence of the other clock proteins.

A minimum of three distinct oscillators have been identified in Neurospora (de Paula et al. 2006, Liu and Bell-Pedersen 2006). The first is the core FRQ/WC-based circadian oscillator (FWO), which requires all five Neurospora clock proteins and drives the asexual sporulation circadian rhythm (Loros and Dunlap 2001, Liu and Bell-Pedersen 2006). The second is the WC-FLO circadian oscillator, which requires WC-1 and WC-2 but not FRQ and drives the ccg-16 mRNA expression daily rhythm (Correa et al. 2003, de Paula et al. 2006). One or more FRQ/WC independent oscillators (FLO) also are active, but their exact number and molecular components are largely unknown (Liu and Bell-Pedersen 2006). The distribution and evolutionary history of the Neurospora circadian clock homologs, coupled with the knowledge of demonstrated circadian rhythms in a few phylogenetically diverse fungal species, raise a number of interesting hypotheses with respect to the evolution of FWO-like and FLO-like oscillators in fungi (Fig. 1), which we discuss below.

We have found that all five circadian clock proteins participating in the Neurospora FWO-oscillator are present only in the clades of Sordariomycetes, Leotiomycetes and Dothideomycetes (Fig. 1). The taxonomically restricted distribution of FRQ strongly suggests that only members of these clades are likely to exhibit circadian rhythms driven by the FWO oscillator. This is consistent with functional complementation experiments showing that the circadian output of frq-null Neurospora strains can be rescued through transformation of non-Neurospora Sordariomycete frq homologs (Merrow and Dunlap 1994).

We have found homologs of WC-1, WC-2, FRH and FWD-1 but not FRQ in most representatives of Zygomycetes, Basidiomycetes and Ascomycetes. Circadian rhythms have been demonstrated in certain species from all three phyla. Examples include the Ascomycetes Aspergillus flavus and Aspergillus nidulans (Greene et al. 2003), whose proteomes are included in this study (Fig. 1), the Zygomycete dung fungus Pilobolus (Bruce, Weight, Pittendrigh 1960), a genus closely related to Rhizopus and Phycymycetes, the two Zygomycetes in our study (White et al. 2006), and the Basidiomycete Sphaerobolus stellatus (Engel and Friederichen 1964, Bell-Pedersen, Garceau, Loros 1996). It is tempting to speculate that the circadian rhythms in at least some of these species are driven by FRQ-independent, WC-FLO-type oscillators. However, the circadian rhythms demonstrated in the two Aspergillus species differ significantly in their manifestation (Greene et al. 2003). The rhythm in A. flavus is with respect to sclerotia development, whereas the rhythm in A. nidulans is with respect to the mRNA expression of the gpdA (glyceraldehyde-3-phosphate dehydrogenase) locus (Greene et al. 2003). Of note, gpdA is apparently rhythmic in A. nidulans strains lacking the wc-1 ortholog (Greene et al. 2003).

All but one species examined from the Saccharomycetes clade appear to have lost FRQ, WC-1 and WC-2, most likely as a consequence of the dramatic genome size reduction that took place in the Saccharomycetes ancestor. The only two oscillators known to drive circadian rhythms in Neurospora, FWO and WC-FLO require either all three proteins (FWO) or the WC proteins (WC-FLO) for function. Consistent with these findings, to date no study has provided evidence in support of the existence of circadian oscillation in any of the species from this clade. Therefore the FLO rhythmicity that has been observed in ultradian rhythms in Saccharomycetes (Murray et al. 2001, Murray, Klevecz, Lloyd 2003, Tu et al. 2005, Tu and McKnight 2007, Lloyd 2008) likely is driven by components independent from those driving the FWO or WC-FLO oscillators.

The molecular characterization of the circadian oscillators in Neurospora (Loros and Dunlap 2001) and Aspergillus (Correa et al. 2003, de Paula et al. 2006) suggests that the mechanisms by which circadian rhythms are generated are likely not conserved across fungi. This is not surprising because one would expect that maintenance of oscillators...
likely would be strongly coupled with organismic life strategy. For example experimental competitive assays of cyanobacterial strains with and without functional circadian oscillators have shown that rhythmicity is adaptive in cyclic environments but not in constant ones (Woelfle et al. 2004). However, the results of our study suggest that, with the exception of FRQ, a large part of the molecular machinery participating in circadian oscillator function seems remarkably conserved (Fig. 1). This difference in degree of conservation is likely best explained by considering the functional properties of the conserved proteins in the oscillator machinery. Two of the four conserved proteins, FWD-1 and FRH, are members of protein families that are widely conserved across eukaryotes and with fundamental roles in several molecular and cellular functions (Dangel et al. 1995, Lee et al. 1995, Jonkers and Rep 2009). The other two conserved proteins, WC-1 and WC-2, are instrumental in fungal light sensing, whose molecular mechanism appears widely conserved across fungi (Idnurm and Heitman 2005, Terashima et al. 2005, Idnurm et al. 2006, Silva, Torres-Martinez, Garre 2006, Sanz et al. 2009). Thus the conservation of all four proteins also could be explained through their participation in other, noncircadian, processes that are conserved across fungi.

ACKNOWLEDGEMENTS

We thank Yi Liu, Carl Johnson, Brian Robertson, Tetsuya Mori, Associate Editor Manfred Binder and two anonymous reviewers for insightful comments on an earlier draft of the manuscript, Yi Liu for providing the sequences of the five Neurospora circadian clock proteins, and Daniel M. Erickson for conducting preliminary evolutionary analyses of the distribution of the Neurospora circadian clock proteins in fungal genomes. This work was conducted in part with the resources of the Advanced Computing Center for Research and Education at Vanderbilt University and the Genome Technology Core at Vanderbilt University Medical Center. LS was financially supported by the Graduate Program in Biological Sciences at Vanderbilt University. Research in AR’s lab was supported by the Searle Scholars Program and the National Science Foundation (DEB-0844968).

LITERATURE CITED


de Paula RM, Lewis ZA, Greene AV, Seo KS, Morgan IW, Vitalini MW, Bennett L, Gomer RH, Bell-Pedersen D. 2006. Two circadian timing circuits in Neurospora crassa...


