Regulatory Factor X (RFX)-mediated transcriptional rewiring of ciliary genes in animals

Brian P. Piasecki, Jan Burghoorn, and Peter Swoboda[1]

Department of Biosciences and Nutrition, Center for Biosciences at NOVUM, Karolinska Institute, S-141 83 Huddinge, Sweden

Edited by Sean B. Carroll, University of Wisconsin, Madison, WI, and approved June 14, 2010 (received for review December 9, 2009)

Cilia were present in the last eukaryotic common ancestor (LECA) and were retained by most organisms spanning all extant eukaryotic lineages, including organisms in the Unikonta (Amoebozoa, fungi, choanoflagellates, and animals), Archaeplastida, Excavata, Chromalveolata, and Rhizaria. In certain animals, including humans, ciliary gene regulation is mediated by Regulatory Factor X (RFX) transcription factors (TFs). RFX TFs bind X-box promoter motifs and thereby positively regulate >50 ciliary genes. Though RFX-mediated ciliary gene regulation has been studied in several bilaterian animals, little is known about the evolutionary conservation of ciliary gene regulation. Here, we explore the evolutionary relationships between RFX TFs and cilia. By sampling the genome sequences of >120 eukaryotic organisms, we show that RFX TFs are exclusively found in unikont organisms (whether ciliated or not), but are completely absent from the genome sequences of all nonunikont organisms (again, whether ciliated or not). Sampling the promoter sequences of 12 highly conserved ciliary genes from 23 diverse unikont and nonunikont organisms further revealed that phylogenetic footprints of X-box promoter motif sequences are found exclusively in ciliary genes of certain animals. Thus, there is no correlation between cilia/ciliary genes and the presence or absence of RFX TFs and X-box promoter motifs in nonanimal unikont and in nonunikont organisms. These data suggest that RFX TFs originated early in the unikont lineage, distinctly after cilia evolved. The evolutionary model that best explains these observations indicates that the transcriptional rewiring of many ciliary genes by RFX TFs occurred early in the animal lineage.

cilia | transcription factor | X-box | Unikont

Cilia and flagella are homologous cellular structures that protrude from the cell surfaces of many different cell types in most eukaryotic organisms. Cilia are complex organelles comprised of more than 300 unique proteins that facilitate a complex array of both sensory- and motility-based functions (1). The axonemal core of the cilium is composed of nine doublet microtubules and is assembled and disassembled by an active process, termed intraflagellar transport (IFT) (2). In IFT, specialized motor proteins move IFT complexes and their cargo in both directions between the base and tip of the cilium (2–5).

Molecular phylogenetic analyses indicate that the eukaryotic domain consists of at least five distinct lineages, so-called supergroups, including the Unikonta, Archaeplastida, Chromalveolata, Excavata, and Rhizaria (6–9). However, the deep-rooting, monophyly, and defining characteristics of certain supergroups remain controversial (7, 8, 10). The Unikonta, which includes Amoebozoa, fungi, choanoflagellates, and animals, is likely a monophyletic supergroup (8, 10, 11). Cilia and ciliary genes have been identified in organisms spanning all major eukaryotic supergroups, indicating that the last eukaryotic common ancestor (LECA) was a ciliated unicellular organism (9, 12). Several theories have been proposed to explain how IFT and the first cilium evolved in protoeukaryotes, including both viral and autogenous origins (13, 14).

Unikont organisms show pronounced differences with respect to the evolution of cilia. For example, cilia and ciliary genes have been secondarily lost in the slime mold *Dictyostelium* (Amoebozoa) and in derived fungi (14). In certain animals, such as *Drosophila melanogaster* and *Caenorhabditis elegans*, ciliary motility-specific functions have been partially or completely lost, and sensory-specific functions of cilia and ciliary genes have been exploited for specialized neuronal functions (4, 5). In mammals, cilia and ciliary genes have been highly diversified to facilitate diverse roles in the development and homeostasis of most cells, tissues, and organs (4, 5, 15). The underlying mechanisms mediating this diversification remain largely unknown.

During the cell cycle and under certain physiological conditions (e.g., stress), ciliary genes are under direct transcriptional control in both unicellular (16) and multicellular (15, 17) unikont and nonunikont organisms. In multicellular organisms, some ciliary genes are differentially regulated for cell-specific expression patterns (18). Thus, the regulation of ciliary genes comprises a unique and complex transcriptional network, which possibly coevolved with multicellularity.

Gene expression in a given regulatory network involves interactions between transcriptional regulators (e.g., transcription factors) and *cis* regulatory elements (e.g., promoter motifs). Various genetic and genomic approaches have provided insights into how transcriptional networks have evolved or are "wired" and "rewired" to regulate the expression of genes involved in novel cellular and morphological pathways (19–22). These studies indicate that the successive variation and modification of preexisting genetic components contributes more to the underlying dynamics driving many evolutionary processes than the de novo acquisition of novel components (19–22). Though most functional analyses characterizing the evolution of transcriptional networks have been conducted in closely related species, much less is known about the mechanisms driving the evolution of transcriptional networks found across more distantly related organisms.

It was first demonstrated in the nematode worm *C. elegans* that RFX TFs are central regulators of ciliogenesis (23). Subsequently it has been shown that at least 50 and perhaps up to several hundred ciliary genes are part of the transcriptional network directly regulated by RFX TFs in *C. elegans*, *D. melanogaster*, mice, and humans (Table S1) (18, 23–29). All RFX TFs contain a highly conserved 76-residue DNA-binding domain, which directly interacts with a 13–15 DNA base-pair consensus sequence, the X-box promoter motif (30). Though only a single RFX gene, *daf-19*, has been identified in the *C. elegans* genome (23, 28), at least seven RFX TF gene paralogs are encoded in the human genome (29). The identities of the direct regulators mediating ciliary gene expression in nonanimal organisms are unknown.

Here, we show that RFX TFs were derived early in the unikont lineage, distinctly after cilia/ciliary genes evolved. Phylogenetic footprinting of ciliary gene promoters extracted from a broad range of eukaryotic genome sequences indicates that X-box promoter motifs are found exclusively in ciliary gene promoters of

Author contributions: B.P.P., J.B., and P.S. designed research; B.P.P. performed research; and B.P.P. and P.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: peter.swoboda@ki.se.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.0914241107/-/DCSupplemental.

animals. These data suggest that the transcriptional network regulating the formation of the cilium, a basal yet complex eukaryotic organelle, was genetically reprogrammed by RFX TFs early in the animal lineage. Understanding the detailed evolutionary dynamics mediating ciliary gene regulation will provide unique insights into how transcriptional networks, which govern more basal cellular processes, have diversified.

Results

RFX Transcription Factors (TFs) Originated Early in Only the Unikont Lineage. Parsing the publicly available eukaryotic genome sequences of >120 evolutionarily diverse organisms, including representative members of all eukaryotic phylogenetic supergroups that contain sequenced organisms, we found that the highly conserved RFX DNA-binding domain is present only in select eukaryotic organisms residing in the unikont lineage (Fig. 1, *SI Materials and Methods*, and Dataset S1). RFX homologs were identified in all animals, many fungi, and a single amoebozoan organism. For further comparisons, full-length RFX amino acid sequences were extracted from the annotated genomes of various unikont organisms, which were subsequently aligned and used to construct a phylogenetic tree (Fig. S1). These comparisons show a clear distinction between the three main groups that comprise the unikont lineage, including animals, fungi, and Amoebozoa. In addition to the DNA-binding domain, RFX TFs contain several other domains, including an A or activation domain, a dimerization domain, and domains B and C, both of unknown functions (Table S1) (29, 31). The activation domain is conserved almost exclusively in vertebrates, although some basal animals also contain a short four-residue sequence (VQYV) found in this domain. Both the dimerization domain and domain C were found in most RFX-containing organisms, except for in a few fungi and the basal amoebozoan *Acanthamoeba castellani*. Domain B was found in all organisms that contain RFX TFs. In contrast, we were unable to locate any of the different RFX TF domains in the genome of any of over 60 nonunikont organisms sampled, including members of Archaeplastida, Chromalveolata, Excavata, and Rhizaria (Fig. 1 and Dataset S1).

Ciliary genes have been extensively studied in several of the nonunikont model organisms sampled for this study, including *Chlamydomonas reinhardtii* (Archaeplastida) (1–4), *Tetrahymena thermophila* (Chromalveolata) (4, 32), and *Trypanosoma brucei* (Excavata) (4, 33). The collective data from these and other organisms has definitively shown that the structure and function of cilia and orthologous ciliary genes are highly conserved (Fig. S2). Moreover, ciliary genes are typically found in a single copy in the genomes of most ciliated organisms, and as such allow for the direct comparison of ciliary gene orthologs (1, 3–5). Because RFX TFs regulate the expression of >50 ciliary genes in animals, yet are completely absent from the genomes of all ciliated nonunikont organisms, the evolutionary relationships between RFX TFs and ciliary gene regulation were further examined.

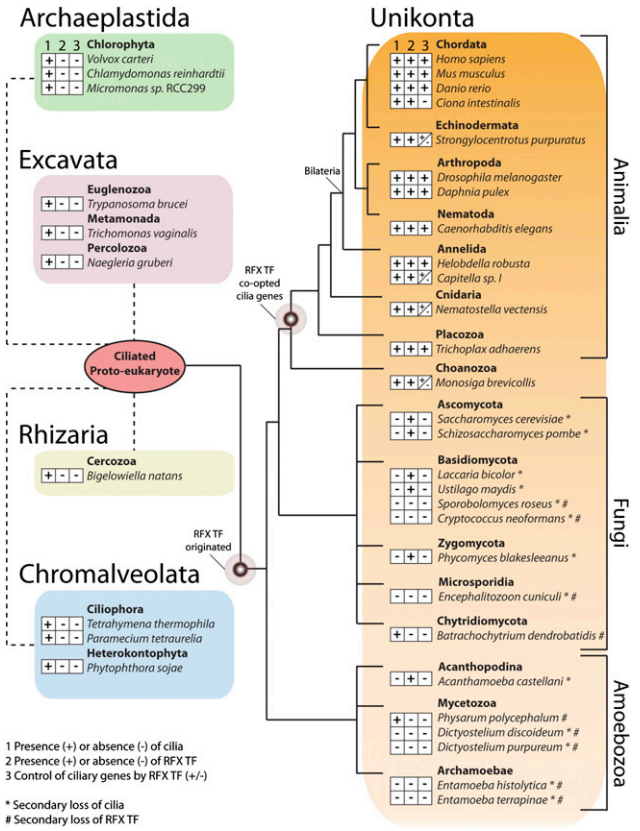

Fig. 1. Schematic diagram depicting the evolutionary conservation of cilia, RFX TFs, and the control of ciliary genes by RFX TFs. Representative organisms from all major eukaryotic phylogenetic supergroups are included. Dashed lines depict supergroups with an uncertain rooting. For each organism, the presence (+) or absence (–) of cilia, RFX TFs, and of ciliary genes that contain X-box DNA footprints or proven X-box promoter motifs, the binding site for RFX TFs, are indicated. The evolutionary loss of cilia is specified only for unikont organisms. All nonunikont organisms listed contain cilia.

RFX TFs Are Also Found in Unikonts Without Cilia. Cilia have been lost secondarily from several groups of organisms, most notably in derived fungi (unikonts) and flowering plants (nonunikonts) (14). Interestingly, RFX TFs are found in many nonciliated unikont organisms, including many fungi and a single amoebozoan organism. When present, these RFX TFs have been shown to bind X-box promoter motifs that are nearly identical to those found in the promoters of ciliary genes in animals. This has been demonstrated, for example, for the sole RFX ortholog found in the nonciliated yeast *Saccharomyces cerevisiae*, Crt1, which regulates a large battery of nonciliary genes (Table S1) (34, 35). A multiple-sequence alignment from representative unikont species illustrates the high degree of conservation observed in the DNA-binding domains of various RFX TFs from ciliated and nonciliated organisms, especially in the residues that make DNA contact or that comprise the hydrophobic core (Fig. 2). These observations could be supported by two possible explanations. They could indicate that RFX TFs diverged to serve some other cellular function when cilia were lost. Alternatively, RFX TFs may have not originally functioned in the regulation of ciliogenesis, but rather RFX regulation of cilia is a derived character state of this transcription factor family.

RFX TFs Are Not Essential for Ciliogenesis in Some Unikonts. The highly conserved RFX DNA-binding domain is absent from the genome sequences of some basal ciliated unikonts, indicating that RFX TFs are not essential for ciliary gene regulation in these groups. For example, the ciliated chytrid fungus *Batrachochytrium dendrobatidis* and the mycetozoan Amoebozoa *Physarum polycephalum* both contain ciliated cells but are missing the RFX DNA-binding domain. A more in-depth analysis revealed that all structural and functional domains of RFX are absent in the genomes of both of these organisms. Thus, some basal unikont organisms contain cilia in the absence of RFX TFs, whereas others contain RFX TFs but do not form cilia. These observations suggest that RFX TFs do not regulate ciliogenesis in more basal unikont organisms (Fig. 1 and Dataset S1). Further, these data collectively imply that RFX TFs secondarily acquired control over ciliary genes distinctly after cilia evolved (Fig. 1).

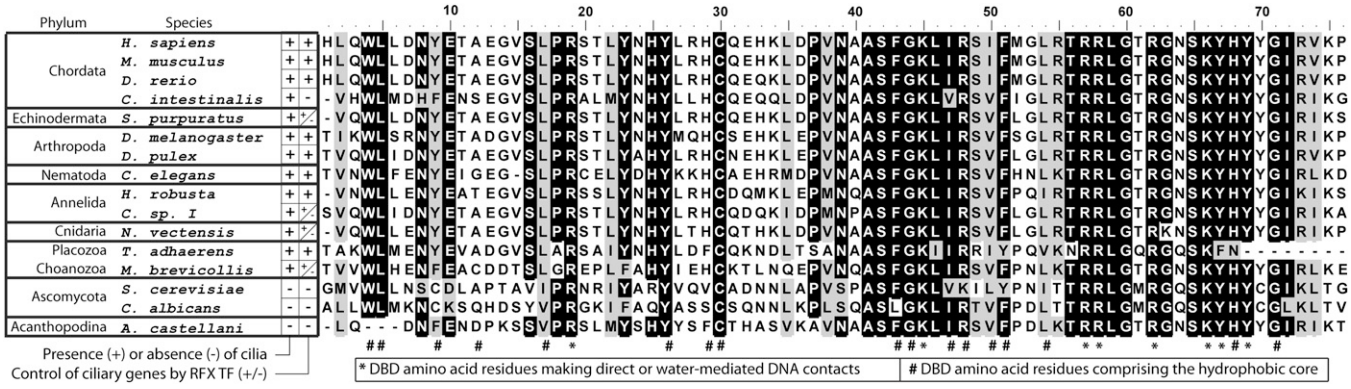

Fig. 2. Multiple-sequence alignment of the 76-residue DNA-binding domain (DBD) of RFX TFs from select unikont organisms. For each organism the presence (+) or absence (–) of cilia and of ciliary genes that contain X-box DNA footprints or proven X-box promoter motifs, the binding site for RFX TFs, are indicated. Structural data were derived from a multiwavelength anomalous dispersion model of human RFX1 (30).

Regulation of Ciliogenesis by RFX TFs Was Derived in Animals: Phylogenetic X-Box DNA Footprint Analyses. To confirm that the RFX regulation of ciliary genes is limited to only certain RFX-containing organisms within the unikonts, and to determine when the control of ciliary genes by RFX TFs was derived, an X-box DNA footprint/X-box promoter motif search was conducted in several ciliary gene promoters extracted from the genome sequences of 23 diverse unikont and nonunikont organisms (Fig. 3). For each organism sampled, BLASTP and T-BLASTN analyses were used to identify orthologs from up to 12 experimentally verified and highly conserved ciliary genes (1, 3–5, 23–27, 36). Because of variations in the availability of genome sequence data and because some structural and functional modifications to cilia have occurred during the evolution of certain organisms, not all of the 23 organisms investigated contained gene orthologs for every sampled gene. We extracted 2-kb promoter sequences from the 5′ upstream region at or near the translational start site for every sampled gene. Of the 12 ciliary genes selected for these analyses, eight of the genes contained X-box promoter motifs that are directly regulated by the DAF-19 RFX TF in *C. elegans*, which has been extensively characterized (18, 23, 24, 28, 36). These eight ciliary gene promoters are therefore expected to contain X-box DNA footprints in other organisms that likely also use RFX TFs for the regulation of ciliary genes. The remaining four ciliary genes either do not contain an X-box promoter motif in *C. elegans* (*KAP3* and *IFT74*) or are absent from the *C. elegans* genome (*SPAG6* and *CCDC147*). As such, these four genes comprise ciliary genes with unknown X-box regulation. The annotated gene ortholog of each of the sampled ciliary gene promoters is listed for humans and other ciliated model organisms (Fig. 3*A*). As controls, four similarly sampled promoters of nonciliary and nontarget genes of RFX TFs (Actin, *EF1α*, *HSP90*, and Ubiquitin) were extracted and analyzed in the same way. These last four genes are therefore not expected to contain X-box DNA footprints in any of the sampled organisms. To find X-box DNA footprints in the extracted gene-promoter sequences, a profile hidden Markov model (HMM) search was conducted (37) using a training set consisting of 17 known X-box motifs from ciliary gene promoters in humans, flies, and worms (Table S2). For additional comparison, the DNA sequence of each extracted gene promoter was subsequently scrambled and resampled identically, including all "expected," "unknown," and "not expected" promoter sequences. Promoter sequences were scrambled using two independent randomizations that separately preserved the mono and dinucleotide frequency rates of each sequence. To identify as many X-box DNA footprints as possible, different threshold values were examined. We found that a threshold value of greater than −1 was sufficient to identify all X-box DNA footprints that were conserved across species (cf. mouse and human *BBS5*; Dataset S2). Pilot experiments indicated that 1-kb promoter regions contained X-box DNA footprints that were conserved across species. Also, fewer random X-box hits can be expected when sampling shorter sequence stretches. Therefore, we trimmed all promoter regions to 1 kb for further analyses.

Many X-box DNA footprints were identified in the "expected" and "unknown" ciliary gene promoters of RFX-containing organisms, whereas very few, if any, X-box DNA footprints were identified in the corresponding gene promoters of non-RFX-containing organisms (Fig. 3*B* and Dataset S2). Because X-box promoter motifs are 13–15 nucleotides long and degenerate, a few X-box DNA footprints are always found at random in the sampled promoter sequences. As a control for false positives, the average background hit rate was determined for all sampled promoters after sequence scrambling. We found that of the 349 promoter regions sampled in these analyses, the sequence-scrambled false-positive hit rate was 6.6% for mononucleotide-preserved and 6.3% for dinucleotide-preserved sequences (Fig. 3*B*). The percentage of organisms in which X-box DNA footprints were identified from all "expected" and from all but one "unknown" ciliary gene promoters was well above the average false-positive rate of all sequence-scrambled promoters in all sampled RFX-containing organisms. In contrast, the percentage of organisms with X-box DNA footprints in the promoters of all sampled non-RFX-containing organisms was within the average false-positive rate of all sequence-scrambled promoters. Further, X-box DNA footprints were identified in only very few of the "not expected" group of gene promoters for both the RFX- and non-RFX-containing organisms. Of note, we observed some variation in the X-box hit rate even in ciliary genes from RFX-containing organisms. For example, many X-box DNA footprints were found in the *BBS1*, *IFT172*, and *SPAG6* ciliary gene promoters. By comparison, significantly fewer X-box DNA footprints were found in the *IFT81*, *KAP3*, and *IFT74* ciliary gene promoters. The successful identification of X-box DNA footprints nearly exclusively in the ciliary gene promoters of RFX-containing organisms within the unikonts suggests that not only are RFX TFs and ciliary genes with bona fide X-box promoter motifs present, but also that RFX TFs likely directly regulate ciliary gene expression in these organisms.

To provide a more in-depth analysis of the X-box DNA footprints identified at the organism-specific level, the percentage of "expected" and "unknown" ciliary gene promoters that contain X-box DNA footprints was determined for each RFX-containing unikont organism we studied (Fig. 3*C*). These data reveal that some organisms contain many ciliary genes with X-box DNA

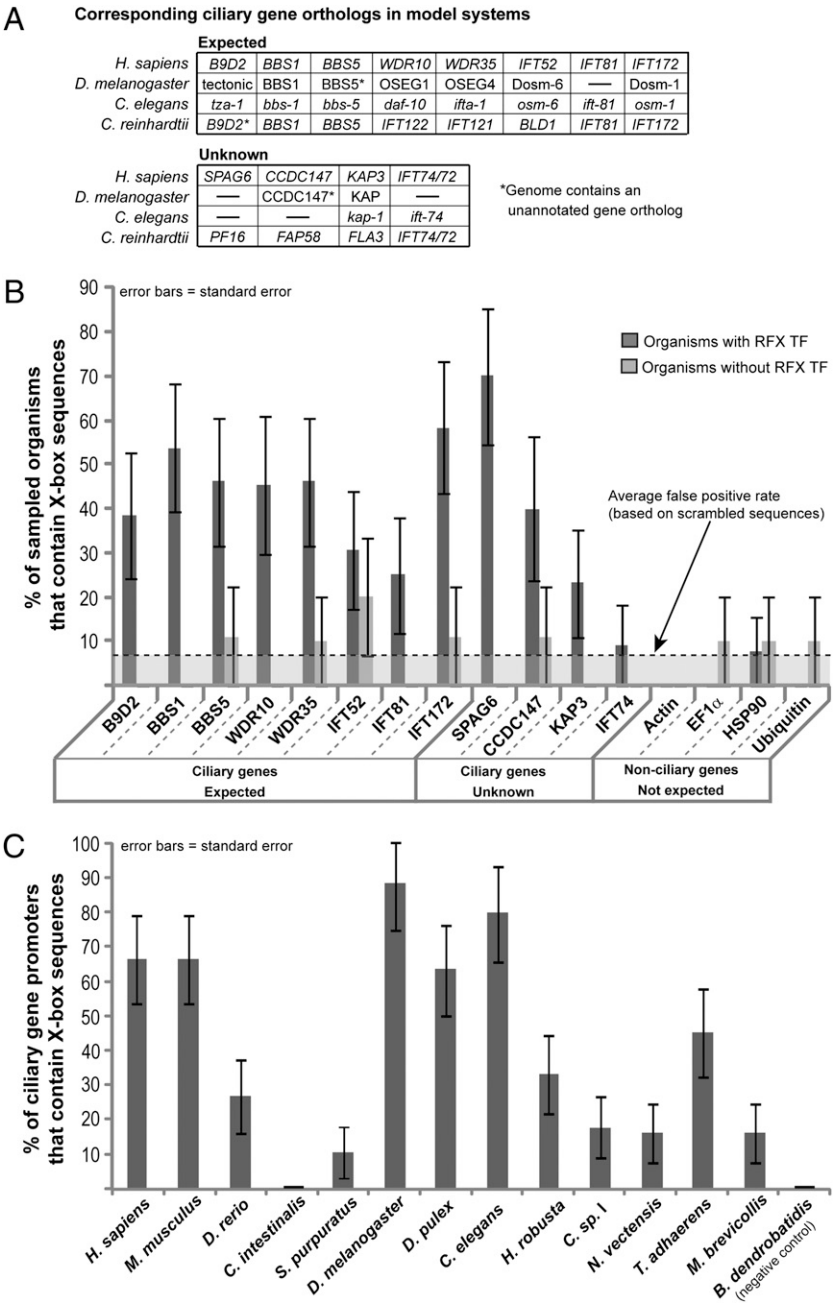

Fig. 3. Identification of X-box DNA footprints or proven X-box motifs in 1-kb promoter regions of evolutionarily conserved ciliary genes from select unikont and nonunikont organisms. (*A*) The annotated gene ortholog of every sampled ciliary gene is listed from humans and other ciliated model organisms. Dashes represent genes that are believed to have been lost from the genome of the respective organism. (*B*) Percent of sampled organisms that contain X-box DNA footprints or proven X-box motifs in ciliary and nonciliary gene promoters from organisms that contain and do not contain RFX TFs. "Expected" ciliary genes contain an experimentally proven X-box promoter motif in *C. elegans*. "Unknown" ciliary genes either do not contain an X-box promoter motif (*KAP3* and *IFT74*) or are absent from the genome (*SPAG6* and *CCDC147*) in *C. elegans*. "Not expected" genes are nonciliary and nontarget genes of RFX TFs. The false-positive rate (dashed line) was calculated by averaging the cumulative number of random X-box hits identified in a complete set of all sampled promoters after sequence scrambling. (*C*) Percent of all "expected" and "unknown" ciliary gene promoters that contain X-box DNA footprints or proven X-box promoter motifs in representative unikont organisms. No RFX TF ortholog was identified in *B. dendrobatidis* (negative control).

footprints, whereas others contain few or none of these footprints in the set of ciliary gene promoters we sampled. The chytrid fungus *B. dendrobatidis* represents a negative control because its genome encodes ciliary genes but not RFX TFs (Fig. 1). Many X-box DNA footprints were identified in *Trichoplax adhaerens*, *Helobdella robusta*, *C. elegans*, *Daphnia pulex*, *D. melanogaster*, *Mus musculus*, and *Homo sapiens*. In contrast, very few X-box DNA footprints were found in the ciliary gene promoters of certain other animals, including *Strongylocentrotus purpuratus*, *Capitella sp. I, Nematostella vectensis*, and the closely related choanoflagellate *Monosiga brevicollis*. However, it remains unclear whether these very few X-box DNA footprints identified were spurious hits or not. Surprisingly, the basal chordate *Ciona intestinalis* does not appear to contain any X-box DNA footprints in the set of ciliary gene promoters we sampled.

The marine placozoan *T. adhaerens* resides at a distinctly basal position within the animal lineage (38). Identification of both RFX TFs and X-box DNA footprints in the ciliary genes of this organism suggests that the last common ancestor of the placozoans, cnidarians, and bilaterians contained a similar ciliary gene regulatory mechanism (Fig. 1). Thus, the RFX-mediated regulation of ciliary genes was likely derived early in the animal lineage.

Statistical Comparisons of Endogenous and Sequence-Scrambled Promoters Confirm an Enrichment of X-Box DNA Footprints in Animals. To account for any organism-specific nucleotide bias in the sampled promoter sequences of each of the 23 unikont and nonunikont organisms, the average number of "expected" and "unknown" ciliary gene promoters that contain X-box DNA footprints was compared with a corresponding average number, when these same, "expected" and "unknown" promoter sequences had been scrambled. A comparison-of-means *t* test was used to determine how statistically different the number of occurrences of identified X-box footprints were, when comparing the endogenous and both mono- and dinucleotide preserved sequence-scrambled promoters of every organism (Table S3). Many, but not all, organisms that contain RFX TFs display statistically higher occurrences of X-box hits in endogenous promoter sequences than they do in scrambled promoter sequences. However, none of the non-RFX-containing organisms, including a single unikont and all nonunikont organisms, display any detectable and statistically significant difference in the number of X-box hits when comparing the endogenous and scrambled promoter sequences. These data confirm the enrichment of X-box DNA footprints exclusively in a variety of animals, including most chordates and all sampled vertebrates (Fig. 3). Further, they indicate that the candidate X-box DNA motifs identified in this study likely play a physiologically relevant role in the RFX-mediated regulation of ciliary gene expression exclusively in animals.

Discussion

RFX TFs are key regulators of ciliogenesis in worms, flies, mice, and humans (18, 23–29). In mammals, RFX3 governs the expression of at least six ciliary genes in the brain (26). The ciliary gene *ALMS1*, which when defective causes Alström syndrome in humans, is regulated by RFX1 (39). Though many direct downstream targets of RFX TFs have been revealed, very little is known about the upstream regulators of RFX TFs themselves. However, several recent studies indicate that RFX TFs may reside at the nexus between upstream patterning and specification pathways and downstream ciliary gene expression (i.e., cell differentiation), including during notochord development (e.g., mouse *Noto*) and during islet cell fate specification in pancreatic cells (e.g., mouse *Ngn3*) (40–42). Thus, RFX TFs play a crucial role in the transcriptional network regulating ciliary gene expression in animals (Fig. 1 and Table S1).

Cilia, which are found in organisms spanning all eukaryotic supergroups, were present in the last eukaryotic common ancestor (LECA) that likely evolved >800 million years ago (8). In contrast, RFX TFs that are associated with the regulation of ciliary gene expression (18, 23–29, 36), are found only in animals, suggesting that they originated distinctly after unikonts diverged from other ciliated organisms. Thus, though RFX TFs are key upstream regulators of ciliary gene expression in animals, they are dispensable for ciliogenesis in nonanimal organisms. In light of current evolutionary models, which predict that many transcriptional networks have evolved through modifications of preexisting genetic components (19–22), these observations raise the following questions. (*i*) How did RFX TFs evolve? (*ii*) When and how did RFX TFs co-opt, and which RFX TFs co-opted control of ciliary gene expression? (*iii*) Are there organism-specific differences with respect to the co-option of ciliary genes by RFX TFs?

The binding of DNA through a particular winged-helix domain is a defining characteristic of all RFX TFs (30). Certain RFX TFs do not regulate ciliary gene expression, and several non-RFX proteins with structurally very similar winged-helix domains are highly conserved in all eukaryotic organisms. For example, the yeasts *Schizosaccharomyces pombe* and *S. cerevisiae* are both non-ciliated unikont organisms that contain RFX TF gene orthologs (*SAK1* and *CRT1*, respectively), which function in cell-cycle regulation and progression (34, 43). Human RFX5, which has diverged significantly from other animal RFX TFs, regulates the expression of MHC class II genes (Table S1) (44). Non-RFX proteins encoding structurally very similar winged-helix domains are components of the origin of replication complex that facilitates DNA replication in all eukaryotic organisms (45). Interestingly, these DNA-binding proteins and their respective winged-helix domains are highly evolutionarily conserved, including in certain Archaea (46). Thus, it seems possible that the archaean and eukaryotic winged-helix-domain-containing proteins share a common ancestor. Perhaps the duplication of a gene encoding a basal eukaryotic winged-helix protein involved in DNA replication occurred early in the unikont lineage, which subsequently allowed for one member to diverge into an RFX TF exclusively in this lineage.

RFX TFs likely co-opted control over ciliary gene expression early in the animal lineage. *T. adhaerens*, currently the most basal animal with a fully sequenced genome, contained a statistically significant number of occurrences of identified X-box DNA footprints in the ciliary gene promoters sampled in this study. In comparison, a very low, albeit potentially relevant, number of X-box DNA footprints were identified in two *M. brevicollis* (Choanozoa) ciliary gene promoters. Choanoflagellates are the closest known extant unicellular relatives of animals and, as such, represent an important phylogenetic position for understanding how multicellularity evolved in animals (47). Though our data are most consistent with RFX TF co-option of ciliary genes beginning in early animals, at present we cannot exclude the possibility that the co-option of certain, likely fewer, ciliary genes occurred before the divergence of animals and choanoflagellates.

RFX TFs appear to have differentially co-opted control over many ciliary genes in various different animal groups. For example, cross-species comparisons revealed that several X-box DNA footprints identified in *T. adhaerens* are conserved in the corresponding human ciliary gene promoters (Fig. S3). Importantly, cross-species comparisons of all human X-box DNA footprints showed corresponding conservation in mouse ciliary gene promoters. Recently, the ciliary genes *Ift172* and *Spag6*, which were sampled in this study, were shown to be targets of the RFX regulatory pathway in mice (27, 48). In comparison, no X-box DNA footprints were found in the ciliary gene promoters sampled from the chordate *C. intestinalis*. Our observation of this potentially significant variation in the regulation of *Ciona* ciliary genes is consistent with other high-throughput studies showing that most *Ciona* genes are differentially regulated when compared with the corresponding genes of vertebrates (49). Further, these data may collectively indicate that there is a degree of plasticity found in the transcriptional regulation of ciliary genes by RFX TFs across different animal groups.

Modifications to transcriptional regulation are a key evolutionary force driving eukaryotic cellular diversification, particularly as it pertains to the derivation of multicellularity (19, 20, 22). However, most large-scale transcriptional network remodeling events appear evolutionarily neutral (50). Tuch et al. (21) proposed that transcriptional rewiring occurs through a series of evolutionarily neutral steps, resulting in the concurrent rewiring of many genes involved in a common pathway. This model was largely based on observations made from the large-scale transcriptional remodeling of genes controlled by the transcriptional regulator Mcm1 in various diverse fungi (51). Our observations that ciliary gene promoters have secondarily and differentially acquired RFX TF regulation exclusively in certain groups of animals suggest that the initial steps in the transcriptional rewiring of many ciliary genes may also have been neutral. Subsequent selective pressure may

then have favored RFX regulation of select ciliary genes in certain animals, including mammals. Understanding the more ancestral functions of RFX TFs, including functions that may still persist in all RFX-containing organisms, as well as why and how the transcriptional rewiring of so many ciliary genes occurred, will be the subject of intense future studies.

Materials and Methods

Protein sequence identifications were conducted by parsing publicly available genome sequences using T-BLASTN and BLASTP analyses. For phylogenetic comparisons, informative characters from full-length aligned amino acid sequences were identified and used to construct a maximum-likelihood phylogenetic tree. X-box DNA footprints were identified using HMMER 2.3.2 (http://hmmer.janelia.org/). Detailed information regarding protein sequence identifications, phylogenetic comparisons, and X-box DNA footprint identifications can be found in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Sandra Baldauf and Gemma Atkinson for assistance with phylogenetic comparisons and tree construction. John Archibald (Dalhousie University, Halifax, Nova Scotia, Canada) graciously provided unreleased sequencing data from the *Bigelowiella natans* genome. We thank Alexander Johnson, Oliver Homann, Thomas Bürglin, Johan Henriksson, Luca Jovine, Patrick Lemaire, and Mats Grahn for helpful discussions. Carolyn Silflow, Gabriele Senti, Elizabeth De Stasio, and Garry Wong critically evaluated the manuscript. Feedback provided by the anonymous reviewers greatly strengthened the manuscript. This work was supported by a Fulbright Fellowship and a Lars Hiertas Minne Foundation grant (to B.P.P.), a Carl Tryggers Foundation Fellowship (to J.B.), and grants from the Swedish Foundation for Strategic Research, Swedish Research Council, Marcus Borgström Foundation, and the NordForsk Nordic *C. elegans* Network (to P.S.).

1. Pazour GJ, Agrin N, Leszyk J, Witman GB (2005) Proteomic analysis of a eukaryotic cilium. *J Cell Biol* 170:103–113.
2. Kozminski KG, Johnson KA, Forscher P, Rosenbaum JL (1993) A motility in the eukaryotic flagellum unrelated to flagellar beating. *Proc Natl Acad Sci USA* 90: 5519–5523.
3. Cole DG, Snell WJ (2009) SnapShot: Intraflagellar transport. *Cell* 137:784–, e1.
4. Pedersen LB, Rosenbaum JL (2008) Intraflagellar transport (IFT) role in ciliary assembly, resorption and signalling. *Curr Top Dev Biol* 85:23–61.
5. Silverman MA, Leroux MR (2009) Intraflagellar transport and the generation of dynamic, structurally and functionally diverse cilia. *Trends Cell Biol* 19:306–316.
6. Baldauf SL (2003) The deep roots of eukaryotes. *Science* 300:1703–1706.
7. Keeling PJ, et al. (2005) The tree of eukaryotes. *Trends Ecol Evol* 20:670–676.
8. Cavalier-Smith T (2009) Megaphylogeny, cell body plans, adaptive zones: Causes and timing of eukaryotic basal radiations. *J Eukaryot Microbiol* 56:26–33.
9. Roger AJ, Simpson AG (2009) Evolution: Revisiting the root of the eukaryote tree. *Curr Biol* 19:R165–R167.
10. Rogozin IB, Basu MK, Csürös M, Koonin E (2009) Analysis of rare genomic changes does not support the unikont-bikont phylogeny and suggest cyanobacterial symbiosis as the point of primary radiation of eukaryotes. *Gen Biol Evol* 1:99–113.
11. Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290:972–977.
12. Cavalier-Smith T (2002) The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int J Syst Evol Microbiol* 52:297–354.
13. Satir P, Mitchell DR, Jékely G (2008) How did the cilium evolve? *Curr Top Dev Biol* 85: 63–82.
14. Jékely G, Arendt D (2006) Evolution of intraflagellar transport from coated vesicles and autogenous origin of the eukaryotic cilium. *Bioessays* 28:191–198.
15. Gerdes JM, Davis EE, Katsanis N (2009) The vertebrate primary cilium in development, homeostasis, and disease. *Cell* 137:32–45.
16. Chamberlain KL, Miller SH, Keller LR (2008) Gene expression profiling of flagellar disassembly in *Chlamydomonas reinhardtii*. *Genetics* 179:7–19.
17. Schneider L, et al. (2005) PDGFRalphaalpha signaling is regulated through the primary cilium in fibroblasts. *Curr Biol* 15:1861–1866.
18. Blacque OE, et al. (2005) Functional genomics of the cilium, a sensory organelle. *Curr Biol* 15:935–941.
19. Carroll SB (2005) Evolution at two levels: On genes and form. *PLoS Biol* 3:e245.
20. Prud'homme B, Gompel N, Carroll SB (2007) Emerging principles of regulatory evolution. *Proc Natl Acad Sci USA* 104 (Suppl 1):8605–8612.
21. Tuch BB, Li H, Johnson AD (2008) Evolution of eukaryotic transcription circuits. *Science* 319:1797–1799.
22. Shubin N, Tabin C, Carroll SB (2009) Deep homology and the origins of evolutionary novelty. *Nature* 457:818–823.
23. Swoboda P, Adler HT, Thomas JH (2000) The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in *C. elegans*. *Mol Cell* 5:411–421.
24. Efimenko E, et al. (2005) Analysis of *xbx* genes in *C. elegans*. *Development* 132: 1923–1934.
25. Laurençon A, et al. (2007) Identification of novel regulatory factor X (RFX) target genes by comparative genomics in *Drosophila* species. *Genome Biol* 8:R195.
26. El Zein L, et al. (2009) RFX3 governs growth and beating efficiency of motile cilia in mouse and controls the expression of genes involved in human ciliopathies. *J Cell Sci* 122:3180–3189.
27. Ashique AM, et al. (2009) The Rfx4 transcription factor modulates Shh signaling by regional control of ciliogenesis. *Sci Signal* 2:ra70.
28. Senti G, Swoboda P (2008) Distinct isoforms of the RFX transcription factor DAF-19 regulate ciliogenesis and maintenance of synaptic activity. *Mol Biol Cell* 19: 5517–5528.
29. Aftab S, Semenec L, Chu JS, Chen N (2008) Identification and characterization of novel human tissue-specific RFX transcription factors. *BMC Evol Biol* 8:226.
30. Gajiwala KS, et al. (2000) Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* 403:916–921.
31. Emery P, Durand B, Mach B, Reith W (1996) RFX proteins, a novel family of DNA binding proteins conserved in the eukaryotic kingdom. *Nucleic Acids Res* 24:803–807.
32. Rajagopalan V, et al. (2009) Analysis of properties of cilia using *Tetrahymena thermophila*. *Methods Mol Biol* 586:283–299.
33. Ralston KS, Kabututu ZP, Melehani JH, Oberholzer M, Hill KL (2009) The Trypanosoma brucei flagellum: Moving parasites in new directions. *Annu Rev Microbiol* 63:335–362.
34. Huang M, Zhou Z, Elledge SJ (1998) The DNA replication and damage checkpoint pathways induce transcription by inhibition of the Crt1 repressor. *Cell* 94:595–605.
35. Zaim J, Speina E, Kierzek AM (2005) Identification of new genes regulated by the Crt1 transcription factor, an effector of the DNA damage checkpoint pathway in *Saccharomyces cerevisiae*. *J Biol Chem* 280:28–37.
36. Chen N, et al. (2006) Identification of ciliary and ciliopathy genes in *Caenorhabditis elegans* through comparative genomics. *Genome Biol* 7:R126.
37. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14:755–763.
38. Srivastava M, et al. (2008) The *Trichoplax* genome and the nature of placozoans. *Nature* 454:955–960.
39. Purvis TL, et al. (2010) Transcriptional regulation of the Alström syndrome gene ALMS1 by members of the RFX family and Sp1. *Gene* 460:20–29.
40. Beckers A, Alten L, Viebahn C, Andre P, Gossler A (2007) The mouse homeobox gene Noto regulates node morphogenesis, notochordal ciliogenesis, and left right patterning. *Proc Natl Acad Sci USA* 104:15765–15770.
41. Smith SB, et al. (2010) Rfx6 directs islet formation and insulin production in mice and humans. *Nature* 463:775–780.
42. Soyer J, et al. (2010) Rfx6 is an Ngn3-dependent winged helix transcription factor required for pancreatic islet cell development. *Development* 137:203–212.
43. Wu SY, McLeod M (1995) The sak1+ gene of *Schizosaccharomyces pombe* encodes an RFX family DNA-binding protein that positively regulates cyclic AMP-dependent protein kinase-mediated exit from the mitotic cell cycle. *Mol Cell Biol* 15:1479–1488.
44. Reith W, LeibundGut-Landmann S, Waldburger JM (2005) Regulation of MHC class II gene expression by the class II transactivator. *Nat Rev Immunol* 5:793–806.
45. Duncker BP, Chesnokov IN, McConkey BJ (2009) The origin recognition complex protein family. *Genome Biol* 10:214.
46. Gaudier M, Schuwirth BS, Westcott SL, Wigley DB (2007) Structural basis of DNA replication origin recognition by an ORC protein. *Science* 317:1213–1216.
47. King N, et al. (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451:783–788.
48. Horvath GC, Kistler MK, Kistler WS (2009) RFX2 is a candidate downstream amplifier of A-MYB regulation in mouse spermatogenesis. *BMC Dev Biol* 9:63.
49. Sobral D, Tassy O, Lemaire P (2009) Highly divergent gene expression programs can lead to similar chordate larval body plans. *Curr Biol* 19:2014–2019.
50. Lynch M (2007) The evolution of genetic networks by non-adaptive processes. *Nat Rev Genet* 8:803–813.
51. Tuch BB, Galgoczy DJ, Hernday AD, Li H, Johnson AD (2008) The evolution of combinatorial gene regulation in fungi. *PLoS Biol* 6:e38.