BMC
Evolutionary Biology

RESEARCH ARTICLE

Convergent evolution of RFX transcription factors and ciliary genes predated the origin of metazoans

Jeffrey SC Chu, David L Baillie and Nansheng Chen*

Abstract

Background: Intraflagellar transport (IFT) genes, which are critical for the development and function of cilia and flagella in metazoans, are tightly regulated by the Regulatory Factor X (RFX) transcription factors (TFs). However, how and when their evolutionary relationship was established remains unknown.

Results: We have identified evidence suggesting that RFX TFs and IFT genes evolved independently and their evolution converged before the first appearance of metazoans. Both ciliary genes and RFX TFs exist in all metazoans as well as some unicellular eukaryotes. However, while RFX TFs and IFT genes are found simultaneously in all sequenced metazoan genomes, RFX TFs do not co-exist with IFT genes in most pre-metazoans and thus do not regulate them in these organisms. For example, neither the budding yeast nor the fission yeast possesses cilia although both have well-defined RFX TFs. Conversely, most unicellular eukaryotes, including the green alga *Chlamydomonas reinhardtii*, have typical cilia and well conserved IFT genes but lack RFX TFs. Outside of metazoans, RFX TFs and IFT genes co-exist only in choanoflagellates including *M. brevicollis*, and only one fungus *Allomyces macrogynus* of the 51 sequenced fungus genomes. *M. brevicollis* has two putative RFX genes and a full complement of ciliary genes.

Conclusions: The evolution of RFX TFs and IFT genes were independent in pre-metazoans. We propose that their convergence in evolution, or the acquired transcriptional regulation of IFT genes by RFX TFs, played a pivotal role in the establishment of metazoan.

Background

All metazoans and many unicellular eukaryotes have functional cilia (also known as flagella) [1]. Both motile and immotile cilia (also known as sensory or primary cilia) hold many receptors for sensing environmental signals. Cilia may offer competitive advantages to ciliated organisms by allowing them to avoid predation and also to track nutritionally rich resources [2]. It is thus not surprising that cilia and most ciliary genes are deeply conserved, both in structure and function, in the "tree of life". Such high levels of conservation suggest a common evolutionary origin [1]. Ciliary defects have been associated with defective development in the nematode *Caenorhabditis elegans* [3] as well as a growing list of devastating human genetic disease conditions collectively called ciliopathies, including polycystic kidney disease (PKD), Bardet-Biedl syndrome (BBS), Alstrome syndrome, Jou-

bert syndrome, Meckel-Gruber syndrome, and primary ciliary dyskinesia [4,5]. In mammals, cilia are found on essentially all cell types, highlighting the critical role cilia play [6]. One essential cellular process in cilia is the intraflagellar transport (IFT) that is responsible for the assembly and maintenance of eukaryotic cilia. The IFT machinery consists of four basic molecular modules: (a) motors, (b) Complex A, (c) Complex B, and (d) BBS complex [7,8].

How IFT genes are regulated at the transcriptional level remained largely unknown until this century when Swoboda and colleagues discovered in *C. elegans* that many IFT genes are regulated by DAF-19, a RFX type transcription factor [3]. Mutations in *daf-19* resulted in defects in cilia development and constitutive dauer formation [3]. DAF-19 binds to X-box motif, which is a highly conserved *cis*-regulatory element first discovered in mammals [3,9]. Ciliary genes in *C. elegans* often contain one or more putative X-box motifs 100 bp - 250 bp upstream of the coding sequences [3,4,10,11]. In addition, ciliary genes and cilia development in the fruit fly *Drosophila*

* Correspondence: chenn@sfu.ca

1 Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada

Full list of author information is available at the end of the article

melanogaster were also suggested to be regulated by RFX TFs [12]. Two RFX genes dRFX[13] and dRFX2 [14] have been identified in *D. melanogaster*. dRFX was identified through a homology search for the RFX DNA binding domain (DBD) and dRFX2 was identified through yeast-one-hybrid (Y1H) screening for transcription factors that bind to a putative promoter sequence [13,14]. Notably, dRFX2 has not been found in the *D. melanogaster* genome sequences, suggesting that it is likely located within the heterochromatin regions (William Gelbart, *personal communication*).

RFX TFs were first identified in mammals as binding proteins of the X-box motif [15]. Through bioinformatics searches and molecular characterization, seven RFX genes--RFX1-7 have been found in mammals [16,17]. Different mammalian RFX genes show differential but overlapping expression patterns [16], suggesting that they have complementary and cooperative roles in regulating genes in many different biological pathways. Indeed, mammalian RFX TFs have been shown to interact with each other and with many additional co-factors [16]. Accumulating evidence confirms that RFX genes regulate development and function of cilia in mammals as well. For instance, RFX3 knockout in mice led to abnormal cilia development in both brain [18] and pancreas [19].

Outside of metazoans, however, there is no evidence suggesting that IFT genes are regulated by RFX TFs. No RFX TFs have been reported in the green alga *Chlamydomonas reinhardtii*, a popular model organism for studying cilia biology. Conversely, RFX TFs exist in organisms including the budding yeast *Saccharomyces cerevisiae* and the fission yeast *Schizosaccharomyces pombe* that do not have cilia [17], suggesting that RFX TFs do not regulate ciliary genes in these organisms. Based on these observations, we hypothesize that IFT genes and RFX TFs evolved independently and that their evolution converged at some point. To test this hypothesis, we have identified and examined IFT genes and RFX TFs in hundreds of fully sequence genomes that have become available recently.

Results

Molecular evolution of ciliary genes

Cilia have been observed to exist in many organisms including mammals, fruit flies, and *C. elegans*. Here, we examine the conservation of cilia by examining the ciliary components identified through searches for human orthologs. In total, we have examined the sequenced genomes of 153 species ranging from metazoans to fungi and plants. The ciliary components examined here include: (1) Five genes from the Motor module (DYNC2H1, K1FAP3, KIF17, KIF3B, and KIF3A); (2) Four from the Complex A module (IFT122, IFT140, WDR35, and WDR19); (3) Nine from the Complex B

module (IFT88, IFT80, IFT172, IFT57, CLUAP1, IFT52, IFT20, IFT81, and IFT74); (4) Six from the BBS complex (BBS5, TTC8, BBS2, ARL6, BBS1, and BBS7) [7,8] (Figure 1). We present results from 31 representative species in Figure 1. Most of the ciliary genes examined are strongly conserved in all metazoans ranging from the sea anemone (*Nematostella vectensis*) to human (*Homo sapiens*) (Figure 1). The unicellular choanoflagellate *Monosiga brevicollis*, which have been regarded as the closest extant relative of the last unicellular ancestor of metazoans [20], also have well conserved ciliary genes. Many ciliated protists, including *Paramecium tetraurelia*, *Tetrahymena thermophila*, and *Phytophthora ramorum* have most of the ciliary genes, consistent with previous reports [21]. Also in agreement with previous reports [22,23], we have identified conserved ciliary genes in the unicellular algae *Chlamydomonas reinhardtii* and its closely related multicellular organism *Volvox carteri*. Protists *Giardia lamblia* and *Physarum polycephalum* have ciliary features that are similar to cilia development in mammals [24-27]. However, we observe reduced similarity for all ciliary genes in these two species, suggesting that these ciliary genes in protozoa are fast evolving [28]. The apicomplexan parasite *Plasmodium falciparum* lacks many ciliary genes, consistent to the idea that the apicomplexan parasites may have an entirely different ciliary assembly mechanism [29]. Among the 51 sequenced fungi, we found only two species, *Allomyces macrogynus* and *Batrachochytrium dendrobatidis*, have conserved IFT genes (Figure 1). Interestingly, both species lack most components of the BBS complex. These observations are consistent with previous proposal that cilia were lost independently in many fungal species in evolution [30,31]. Taken together, our comparative identification and analysis of IFT genes suggest that IFT genes are deeply conserved and can be found in all metazoans, most unicellular eukaryotes, and some fungi, but they do not exist in plants such as *Arabidopsis thaliana* and prokaryotes [1,21] (Figure 1).

Molecular evolution of RFX TFs

Using well defined RFX DBD peptide sequences (76 amino acids long) (Figure 2) from human [16], *C. elegans* [3], *D. Melanogaster* [13], and *S. Cerevisae* [32] as queries, we searched the genomes of the same 153 species for RFX TFs. Because the known RFX DBDs in yeast as well as humans show very high similarity, we used very stringent criteria to look for new RFX TFs. We only consider proteins whose putative RFX DBD show at least 40% percentage identity (PID) to the queries (see Methods). RFX DBD has been shown to contain nine residues that have direct contact with DNA sequences (X-box motifs) [33]. All nine residues are highly conserved in all known RFX DBDs (Figure 2). Therefore we also required that the

Figure 1 The conservation of RFX TFs and ciliary IFT components in selected species. These species were selected to provide a wide sampling of the "tree of life". The phylogenetic relationship between each species was derived from the "Tree of Life Web Project" [44]. Species indicated with "*" have ciliated cells based on published evidence. The 'RFX #' column shows the number of putative RFX TFs identified in this project or reported previously. The grey scale table shows the sequence conservation of individual ciliary components in each species. Darker shade represents higher sequence similarity and conservation. The numbers in each box indicate the percent identity revealed by the alignments between IFT genes and their corresponding human orthologs.

Species	RFX #	Motors					Complex A				Complex B									BBS Complex					
		DYNC2H1	KIFAP3	KIF17	KIF3B	KIF3A	IFT122	IFT140	WDR35	WDR19	IFT88	IFT80	IFT172	IFT57	CLUAP1	IFT52	IFT20	IFT81	IFT74	BBS5	TTC8	BBS2	ARL6	BBS1	BBS7
Homo sapiens*	7			81			86	82	92	89	88					64					90			92	92
Mus musculus*	7	80	35	81	83	91	76	70	86	82	80	84	81	72	74	78	37	78	62	90	83	79	39	0	78
Gallus gallus*	7	72	89	24	73	93	70	57	82	74	73	76	55	72	72	80	61	75	69	87	76	79	55	66	81
Xenopus tropicalis*	7	62	76	45	80	92	30	58	79	76	74	74	51	65	77	51	70	44	90	90	26	74	88	62	76
Danio rerio*	9	55	54	38	67	70	56	47	60	57	52	56	62	57	51	63	32	47	41	64	62	57	66	43	58
Ciona intestinalis*	4																								
Drosophila melanogaster*	4	29	47	25	46	58	43	33	49	41	32	35	46	32	26	34	24	21	20	48	29	0	55	30	0
Strongylocentrotus purpuratus*	6	34	63	17	49	51	53	36	68	61	61	62	20	58	60	51	35	57	38	75	68	55	50	54	58
Caenorhabditis elegans*	1	40	32	31	38	54	34	31	31	37	45	40	42	37	37	38	17	23	24	46	40	29	44	27	33
Nematostella vectensis*	5	62	63	40	66	63	63	45	63	56	51	43	57	53	54	57	14	36	40	65	43	57	44	76	55
Monosiga brevicollis*	2	46	40	33	50	53	45	36	48	44	45	35	51	39	46	53	11	42	36	45	44	24	39	28	30
Saccharomyces cerevisiae	1	17	10	17	27	28	5	0	3	3	8	8	0	15	20	0	14	23	21	0	12	0	39	0	0
Schizosaccharomyces pombe	1	17	4	17	27	28	3	2	3	3	14	9	0	13	16	0	0	21	21	0	12	0	39	0	0
Cryptococcus neoformans grubii	0	23	9	18	29	32	3	0	0	2	10	11	3	10	16	0	14	20	22	0	12	0	41	0	0
Batrachochytrium dendrobatidis*	0	44	35	28	48	48	44	26	44	33	44	50	51	35	42	52	16	36	32	0	9	0	47	4	0
Allomyces macrogynus*	1	25	27	29	43	49	33	19	37	23	33	32	45	31	38	46	18	30	26	0	7	0	40	0	0
Physarum polycephalum*	0	7	8	22	34	34	8	11	11	2	5	7	2	13	18	19	13	19	19	45	19	10	39	16	15
Paramecium tetraurelia*	0	35	22	25	41	44	32	25	40	33	35	45	40	28	36	46	18	28	27	30	41	23	41	25	23
Tetrahymena thermophila*	0	41	24	26	43	46	27	23	41	35	36	43	40	27	28	27	29	42	31	41	42	28	41	42	28
Phytophthora ramorum*	0	44	28	27	46	45	44	37	50	39	36	48	46	31	41	48	16	34	29	60	48	41	54	32	36
Plasmodium falciparum*	0	23	0	15	27	30	4	0	0	3	7	4	0	15	21	0	0	20	22	0	0	0	36	0	0
Leishmania infantum*	0	38	0	25	30	32	34	24	39	33	37	36	39	25	32	39	24	29	26	41	35	32	44	31	22
Trypanosoma brucei*	0	38	0	20	30	34	31	25	40	32	36	38	42	29	38	40	14	29	27	41	38	32	44	30	20
Giardia lamblia*	0	19	16	23	37	38	13	6	22	9	23	8	17	19	25	27	0	22	22	20	15	0	38	0	0
Chlamydomonas reinhardtii*	0	42	37	25	50	47	46	34	53	37	36	44	43	38	20	28	25			51	49	39	49	39	29
Volvox carteri*	0	37	38	25	51	49	46	34	53	39	36	47	45	34	35	46	20	28	25	52	51	35	52	39	0
Arabidopsis thaliana	0	3	6	14	31	32	4	2	3	4	10	10	4	13	20	0	0	22	21	0	14	0	39	0	0
Methaococcus jannaschii	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sulfolobus solfataricus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Escherichia coli	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bacillus subtilis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

DBDs of candidate RFX TFs contain all of these nine conserved residues.

We found candidate RFX TFs in all sequenced metazoan genomes (Figure 1). In addition to the RFX TFs that have been reported previously, including seven RFX TFs found in mammals [16], DAF-19 in *C. elegans* [3], and dRFX [13], we found many RFX genes that have not been described previously. We have identified seven RFX genes (RFX1-7) in all vertebrate genomes except fish genomes, which have nine putative RFX genes (RFX1-9). We have also identified four RFX genes in *Ciona intesttinalis*, six in the purple sea urchin (*Strongylocentrotus purpuratus*), and five in the sea anemone (*Nematastella vectensis*). In *D. melanogaster*, in addition to the two RFX genes reported previously--dRFX and dRFX2, we have identified a novel RFX TF, which we named *dRFX1*. Interestingly, among all metazoans examined, nematodes


```
hRFX1      : TVQWLLDNYETAEGVSLPRSTLYCHYLLHCQEQK-LEPVNAASFGKLIRSVFMGLRTRRLGTRGNSKYHYYGLRIKA--
hRFX2      : HLQWLLDNYETAEGVSLPRSSLYNHYLRHCQEHK-LDPVNAASFGKLIRSVFMGLRTRRLGTRGNSKYHYYGIRLKP--
hRFX3      : HLQWLLDNYETAEGVSLPRSTLYNHYLRHCQEHK-LDPVNAASFGKLIRSIFMGLRTRRLGTRGNSKYHYYGIRVKP--
hRFX4      : TLQWLEENYEIAEGVCIPRSALYMHYLDFCEKND-TQPVNAASFNGKLIRQQFPQLTTRRLGTRGQSKYHYYGIAVKE--
hRFX5      : AYRWIRNHLEEHTDTCLPKQSVVDAYRKYCESLACCRPLSTANFGKLIRELPDIKARRLGGRGQSKYCYSGIRRKT--
hRFX6      : TLQWLEENYIVCEGVCLPRCILYAHYLDFCRKEK-LEPACAATFGKTIRQKFPLLTTRRLGTHGKSKYHYYGIGIKE--
hRFX7      : AFSWIRNTLEEHPETSLPKQEVYDEYKSYCDNLG-YHPLSAADFGKIMKNVFPNMKARRLGTRGKSKYCYSGLRKKA--
DAF-19     : TVNWLFENYEIGEG-SLPRCELYDHYKKHCAEHR-MDPVNAASFGKLIRSVFHNLKTRRLGTRGNSKYHYYGIRLKD--
sRFX1      : ALLWLMKNCKSQHDSYVPRGKIFAYYASSCSQNN-LKPLSQASLGKLIRTVFPDLTTRRLGMRGQSKYHYCGLKLTVNE
dRFX       : TIKWLSRNYETADGVSLPRSTLYNHYMQHCSEHK-LEPVNAASFGKLIRSVFSGLRTRRLGTRGNSKYHYYGIRIKP--
Mbre_cRFX1 : TVVWLHENFEACDDTSLGREPLFAHYIEHCKTLN-QEPVNQASFGKLIRSVFPNLKTRRLGTRGNSKYHYYGIRLKE--
Mbre_cRFX2 : ---WIHEHYELKEAACVLRSSLYENYVKFCELTS-QEPTNAANFGKLIRQQFPGLKTRRLGTRGQSKYHYYGLRLK--
```

Figure 2 DBDs of RFX TFs are highly conserved. Representative DBDs from humans (hRFX1-7), *C. elegans* (DAF-19), *D. melanogaster* (dRFX), *S. serevisiae* (sRFX1), and *M. brevicollis* (Mbre_cRFX1 and Mbre_cRFX2). DBDs from different species show high similarity at the peptide level. Nine residues of DBD that directly contact DNA (indicated by arrows) are essentially identical for all RFX TFs.

including *C. elegans* are the only organisms that possess just one RFX gene.

RFX TFs are also found in some non-metazoans. Of the 51 fungus species examined, we identified single RFX TFs in 44 species, including the budding yeast *Saccharomyces cerevisiae* and the fission yeast *Schizosaccharomyces pombe*, as previously reported [17], as well as a ciliated fungus *Allomyces macrogynus*, whose genome was recently sequenced by the Fungal Genome Initiative of the Broad Institute http://www.broadinstitute.org/annotation/fungi/fgi/. All unicellular organisms we have examined possess either one RFX gene (fungi) or none except for the choanoflagellates. For example, *M. brevicollis*, which was recently sequenced [20], contain two genes (Mbre_cRFX1 and Mbre_cRFX2) with well-defined RFX DBDs.

RFX DBD sequences are the defining features of all known RFXs and show high similarity (>40% PID) to each other. However, there are a small number of additional proteins that contain domains that show weaker similarity (<30% PID) to known RFX DBDs. In particular, a gene (ARID2) in the human genome contains a RFX-like domain that shows 29% PID to the human RFX1 DBD. Among the nine residues that have direct contact with DNA sequences, five can be found in the RFX-like domain found in ARID2. ARID2, whose function as a transcription factor has not been well studied, has orthologs in all mammals as well as other vertebrates (data not shown). Additionally, a gene in *M. brevicollis* also shows weak similarity (27%) to known RFX DBDs (five of the nine residues that have direct contact with DNA are conserved). We name this novel gene Mbre_cRFX3. Because of their low similarity to known RFX DBDs, these RFX like genes--ARID2 and Mbre_cRFX3--are not regarded as RFX TFs in this project and thus are not examined further. No RFX genes have been found in any bacteria, ancient bacteria, or plants (Figure 1).

DBDs in the two putative RFX TFs in *M. brevicollis* are essentially indistinguishable from the DBDs in previously characterized RFX TFs with ~70% PID at the peptide level. All nine residues that make direct contacts with X-box motifs are conserved [33] (Figure 2, residues indicated with arrows). In addition to the DBDs, Mbre_cRFX1 also shares other functional domains within known RFX TFs including the dimerization domains (DD), and the extended dimerization domains (B and C domains), which exist in all mammalian RFX TFs except RFX5 and RFX7 [16,17] (Figure 3). Aligning Mbre_cRFX1 to human RFX1-3 shows clear alignment for conserved DBD, DD, and extended dimerization domains (B and C domains) (Figure 4). None of the *M. brevicollis* RFX TFs have readily identifiable activation domains (AD). The lack of typical AD in RFX TFs in *M. brevicollis, C. elegans,*

D. melanogaster, and sea anemone (Figure 3) suggests that AD might have been acquired later in metazoan evolution. Alternatively, their ADs have yet to be identified and characterized. Mbre_cRFX2 has a readily identifiable DBD but lacks other conserved domains, which is similar to the human RFX5 and RFX7 that lack other domains (Figure 3). The presence of DBD (in both Mbre_cRFX1 and Mbre_cRFX2) and other conserved protein domains (in Mbre_cRFX1) suggest that they may function in transcriptional regulation of gene expression in *M. brevicollis.* However, their target genes remain to be identified.

To further examine the relationship between the *M. brevicollis* RFX TFs and those identified in mammals and other species, we constructed a phylogenetic tree that contains all known and putative RFX TFs based on the similarity between the DBD domains (Figure 5). Sequences outside of the DBDs are excluded from analysis since they are often very diverse and are not readily alignable. Previous analysis of mammalian RFX TFs revealed three groups: RFX1-3, RFX4-RFX6, and RFX5-RFX7 groups [16], which is generally consistent with this phylogenetic tree with newly identified members (Figure 5). The tree shown in Figure 5 contains an additional clade (shown in black), which contains RFX TFs identified in fungal genomes and, interestingly, *dRFX2* in *D. melanogaster* [14]. Fungus RFX RFs (members in the Fungus clade) and RFX5-RFX7 TFs show similar domain compositions with all members lacking B, C, and D domains, which are found in the RFX1-3 and RFX4-RFX6 TFs (Figure 3). The inferred phylogenetic tree clearly shows that the Mbre_cRFX1 fits into the RFX1-3 group, while Mbre_cRFX2 fits into the RFX4-6 group. Mbre_cRFX3, which show weaker similarity to known DBDs, clusters closer to DBDs of the RFX5-7 groups. However, as mentioned before, we did not include Mbre_cRFX3 in the phylogenetic tree. The phylogenetic relationship between *M. brevicollis* and previously identified RFX TFs suggest that these three RFX TFs families were established before the split between choanoflagellates and metazoans.

In the inferred phylogenetic tree, the nematodes are the only metazoans that have only one RFX TF--DAF-19, which groups together with the mammalian RFX1-3 group (Figure 5). It was proposed previously that prior to the complete sequencing of the *C. elegans* genome, more RFX TFs should exist in *C. elegans* [17]. However, exhaustive searches of the completed *C. elegans* genome revealed no traces of additional RFX genes, suggesting that RFX genes corresponding to other RFX groups (RFX4-6 and RFX5-7) were lost in the last common ancestor of the nematode species. In fact, none of the seven sequenced nematode genomes have more than one RFX TF (Additional file 1).
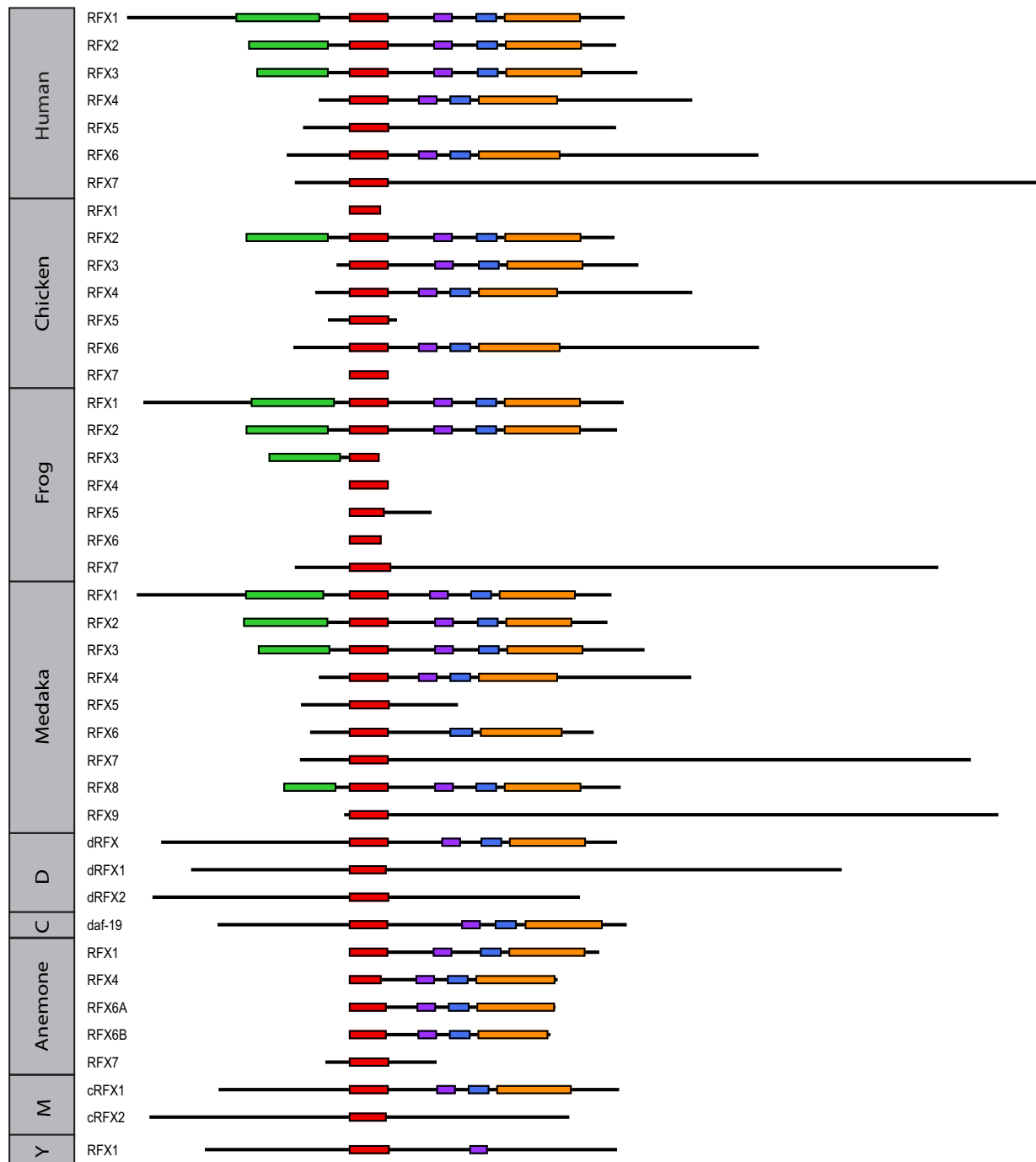
Figure 3 Predicted protein domains of RFX TFs in representative species. The defining domain of all RFX TFs--DBD--is shown in red. Other domains including the activation domain (green), the B domain (purple), C domain (blue), and D domain (orange) are not present in all RFX TFs. In the left column, Y stands for the budding yeast *S. cerevisiae*, M for *Monosiga brevicollis*, C for *C. elegans*, and D for *D. melanogaster*.
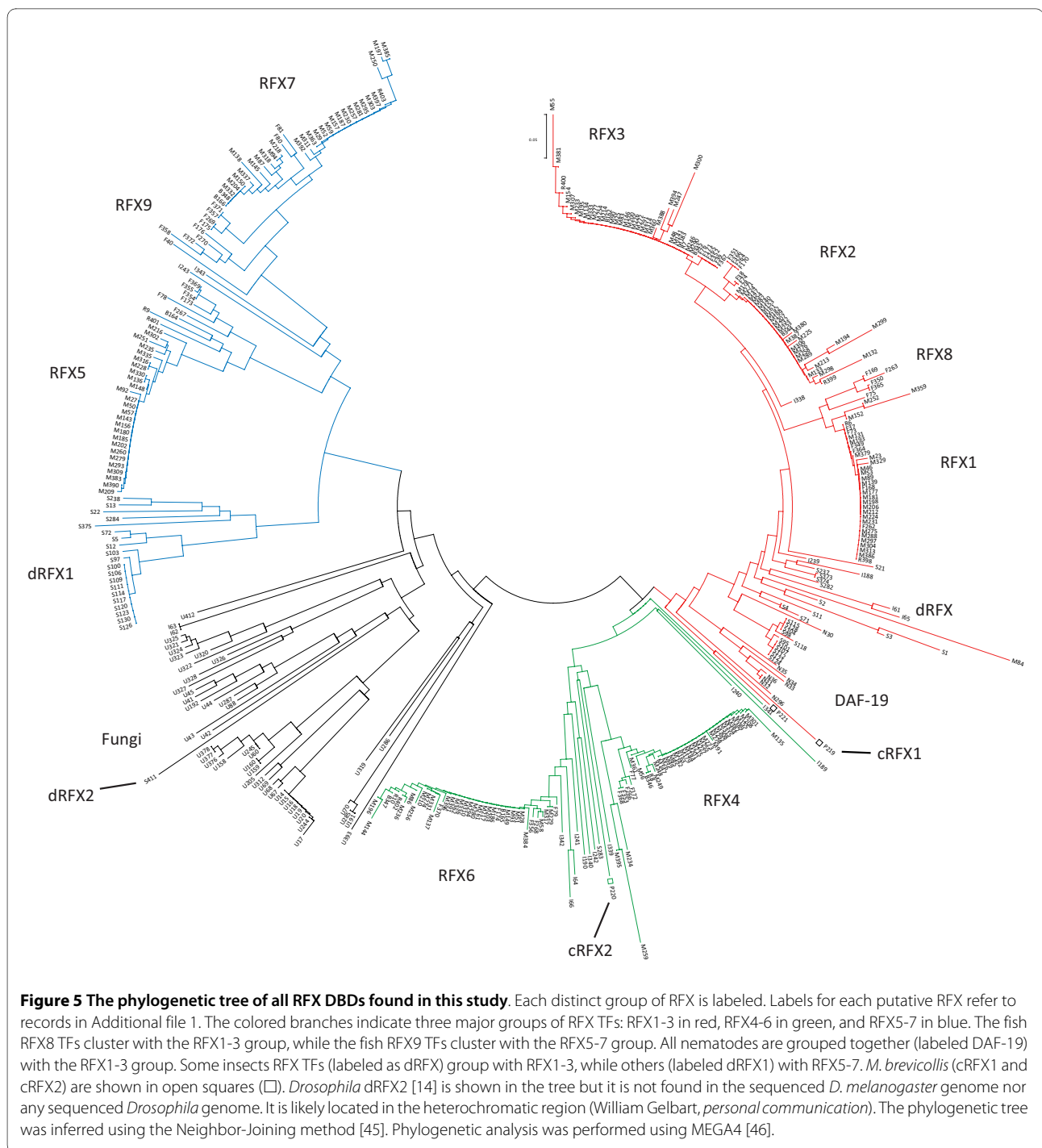
Evolutionary relationship between ciliary genes and RFX TFs

The above comprehensive identification of IFT genes and RFX TFs shows clearly that all metazoans have both ciliary genes and RFX genes. Since IFT genes have been

demonstrated to be regulated by RFX TFs in *C. elegans*, *D. melanogaster*, and humans, IFT genes in all metazoans are likely regulated by RFX TFs. Our analysis strongly suggests that IFT genes and RFX TFs evolved independently. In addition to the budding yeast (*S. cerevisiae*) and
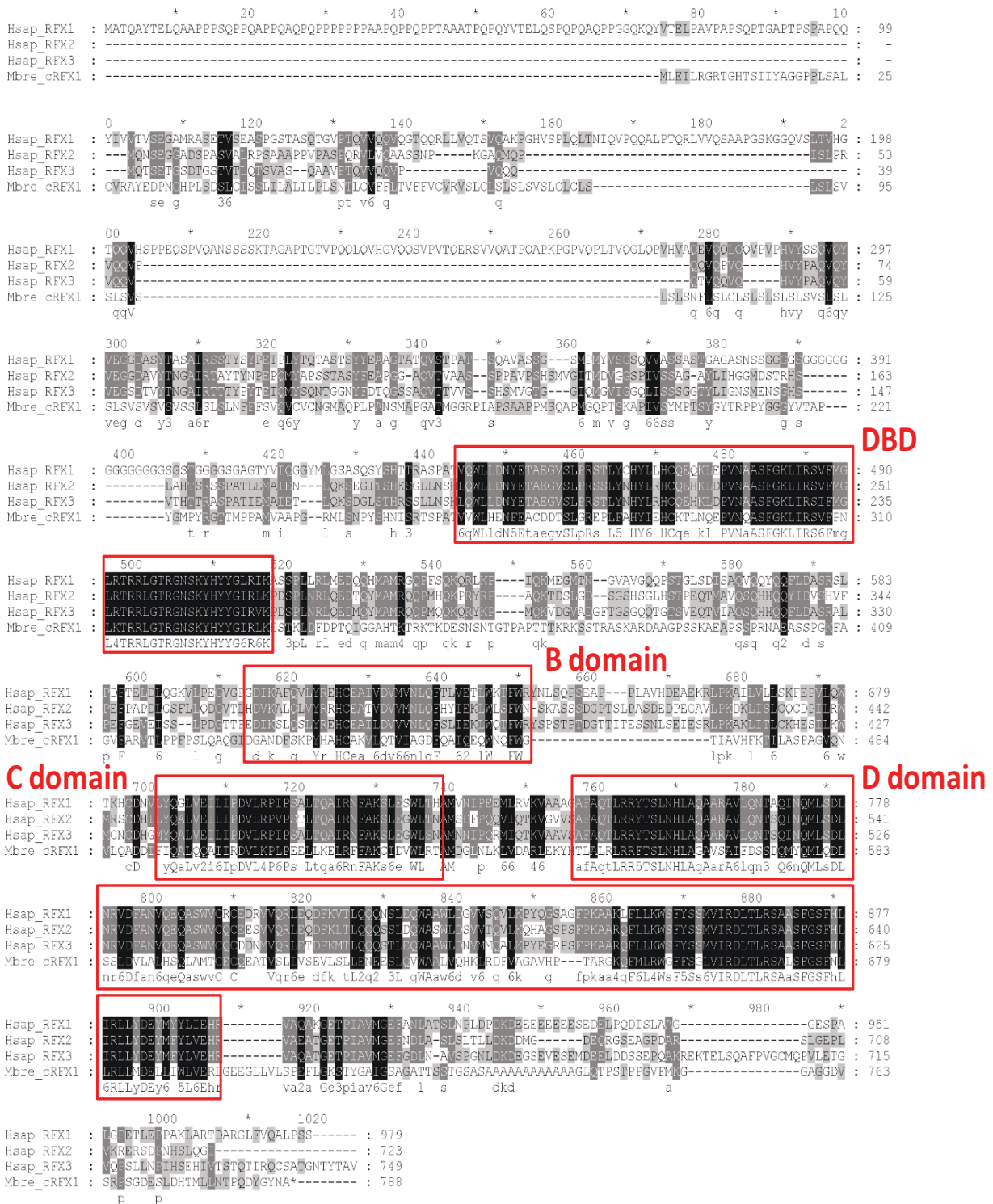
Figure 4 Sequence alignment between *M. brevicollis* Mbre_cRFX1 and Human RFX1-3 with all functional domains highlighted. Amino acid residues are color coded with darker color representing higher conservation. Putative functional domains are encircled and labeled.

Figure 5 The phylogenetic tree of all RFX DBDs found in this study. Each distinct group of RFX is labeled. Labels for each putative RFX refer to records in Additional file 1. The colored branches indicate three major groups of RFX TFs: RFX1-3 in red, RFX4-6 in green, and RFX5-7 in blue. The fish RFX8 TFs cluster with the RFX1-3 group, while the fish RFX9 TFs cluster with the RFX5-7 group. All nematodes are grouped together (labeled DAF-19) with the RFX1-3 group. Some insects RFX TFs (labeled as dRFX) group with RFX1-3, while others (labeled dRFX1) with RFX5-7. *M. brevicollis* (cRFX1 and cRFX2) are shown in open squares (□). *Drosophila* dRFX2 [14] is shown in the tree but it is not found in the sequenced *D. melanogaster* genome nor any sequenced *Drosophila* genome. It is likely located in the heterochromatic region (William Gelbart, *personal communication*). The phylogenetic tree was inferred using the Neighbor-Joining method [45]. Phylogenetic analysis was performed using MEGA4 [46].

fission yeast (*S. pombe*), we have identified 41 fungus species that have single RFX genes but no IFT genes, thus RFX genes in these species do not regulate ciliary genes expression. Indeed, Crt1/RFX in the budding yeast plays a role in DNA damage response [34]. Outside of metazoans, only two sequenced genomes have both IFT and RFX genes, the choanoflagellate *M. brevicollis* and the fungus *A. macrogynus*. Outside of metazoans, choanoflagellates,

and fungi, none of the sequenced genomes possess a single RFX gene, regardless of the possession of IFT genes.

Discussion

This is the first project to comprehensively identify and compare RFX TFs in the entire "tree of life" since Emery and colleagues described RFXs in humans (RFX1-5), mice (RFX1-3 and RFX5), *C. elegans*, and the budding

and fission yeasts domains more than a decade ago [17]. In this paper, we identified for the first time (1) nine RFX genes in all sequenced fish genomes; (2) two RFX genes in the choanoflagellate *M. brevicollis* genome; (3) single RFX genes in many fungus genomes. Additionally, we have identified RFX genes in many vertebrates. Furthermore, we have identified a third RFX (*dRFX1*) in the fruit fly *D. melanogaster*. Based on our phylogenetic analysis of all RFX TFs identified in the "tree of life", we have confirmed the hypothesis proposed by Emery and colleagues that *C. elegans* has lost RFX genes as it evolved [17].

More importantly, comparative analysis of the molecular evolution of IFT genes and RFX genes revealed a compelling converging relationship between these two gene groups, which is summarized in a model illustrated in Figure 6. We propose that the common ancestor of metazoans, choanoflagellates, and fungi was ciliated and had one RFX gene. Even though the common ancestor of all fungus species was ciliated and had one RFX gene in some fungus species, including *Batrachochytrium dendrobatidis*, inherited cilia but lost RFX, while other species, including budding yeast and fission yeast, lost their cilia but retained RFX, and some species, including *Cryptococcus neoformans grubii*, lost both cilia and RFX TFs, leaving only a few fungal species, including *Allomyces*

Figure 6 RFX TF-mediated transcription and the origin of metazoans. The common ancestor of metazoans, choanoflagellates, and fungi was likely a ciliated unicellular eukaryote with a single RFX TF. Over the course of evolution, some fungi lost RFX TFs while preserving cilia, some lost cilia but kept RFX, and some lost both. Only a few fungi kept both cilia and RFX. The last common ancestor (LCA) of *Monosiga* and metazoans preserved both cilia and RFX.

macrogynus, that retain both RFX TFs and cilia (Figure 1 and Figure 6). In ciliated fungus species, which do not have RFX genes, ciliary genes are likely regulated by factors other than RFX TFs. In contrast, the common ancestor of metazoans and choanoflagellates was ciliated and had multiple RFX genes. The plurality of RFX genes was probably due to gene duplication (Figure 6). The expansion of the RFX gene family, in the common ancestor of metazoans and choanoflagellates, might have provided a platform for the development of interactions between RFX TFs and IFT genes and the establishment of transcriptional regulatory relationships between RFX TFs and IFT genes in metazoans. The convergent molecular evolution of IFT genes and RFX TFs might have provided a pivotal driving force in the emergence and evolution of metazoans.

The evolution of multicellular metazoans from a unicellular protozoan ancestor represents a major and what we consider to be the most spectacular transition in the "history of life". This transition is demonstrated by the abrupt appearance of a huge variety of metazoans in the fossil record approximately 560 million years ago during the Cambrian explosion [35]. Many environmental, ecological, and other evolutionary factors have been proposed to have contributed to this transition [36,37]. Great efforts have been made to understand this transition by studying protein-coding regions of numerous genes and gene families that are ubiquitous in and limited to metazoans. Findings obtained in these studies showed that many genes and gene families previously found to be expressed only in metazoans are also found in choanoflagellates giving evidence that metazoans arose from choanoflagellates. For example, work by King and colleagues clearly demonstrated that choanoflagellates have a receptor tyrosine kinase that is found in metazoans but not in other eukaryotes [35]. Manning and colleagues searched the sequenced choanoflagellates *M. brevicollis* genome [20], and identified a highly elaborate tyrosine kinase signaling network [38]. Many additional genes are shared by *M. brevicollis* and metazoans, including cadherin, which are essential for metazoan development [39], and transcription factors such as P53 and Myc [20]. These findings encouraged additional large scale searches, including the UNICORN (unicellular opisthokont research initiative) project [36], for genes and gene families critical for the transition from unicellularity to multicellularity. However, accumulating evidence is showing that these genes predated the origin of metazoans and played different roles from their counterparts in metazoans. Thus these genes, even though some have been co-opted to perform novel functions in metazoans, are probably not be the main driving force underlying the transition from unicellular protozoans to multicelluar metazoans.
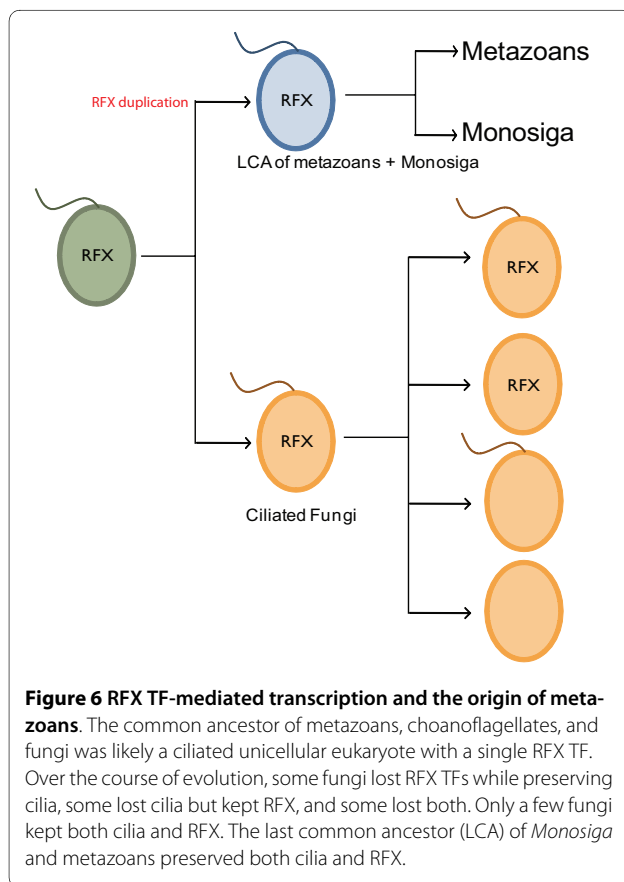
What then was the main factor driving this transition? In contrast to coding sequences of genes, which are usually under strong purifying selection, regulatory sequences show much more rapid evolution. Compelling evidence suggests that changes in *cis*-regulatory sequences and transcriptional regulation in general play a pivotal role in evolution [37,40]. Kingsley and colleagues recently identified changes in *cis*-regulatory modules that dictate dramatic changes in pigmentation in sticklebacks and humans [41]. Thus the transition from unicellular flagellates to multicellular metazoans may have been driven by innovations at the transcriptional level. The convergent evolution of RFX TFs and ciliary genes (IFT genes in particular) in the common ancestor of metazoans and choanoflagellates prompt us to propose that the acquired tight control of ciliary genes at the transcription level by RFX TFs served as one of the critical driving forces in the establishment of multicellularity and the rise of metazoans.

Conclusion
RFX TFs and IFT genes evolved independently in premetazoans and their convergence, or the acquired transcriptional regulation of IFT genes by RFX TFs, may have played a pivotal role in the establishment of metazoan.

Methods
Data sources
All sequence data (both genomic DNA sequences and gene annotation data including cDNA and protein sequences) were downloaded from public databases. The list of genomes and the data source are described in Additional file 2. The initial set of DNA binding domains that were used as queries for BLAST searches were taken from Human RFX1-7 [16], *C. elegans* DAF-19 [3], *D. melanogaster* dRFX [13], and yeast RFX1 [32].

Identification of RFX TFs
We carried out similarity searches using WU-BLAST (version 2.2.6; http://blast.wustl.edu) with e-value 0.01 and without sequence filter (option -F). The initial set of DBDs was used as query to search against all the mammalian proteomes (entire collection of protein peptides). The resulting DBDs were added to the query list and used to search against arthropods. The iteration of adding DBD and blasting continues until all species have been searched. A hit is accepted as a candidate DBD if the corrected percent identity over the entire domain length is >= 40%. The corrected percent identity was calculated as the number of identical positions divided by total length of the query. We also searched for candidate RFX TFs in genome sequences (DNA sequences) to ensure that no RFX TFs have been missed in the gene annotations.

Identification of ciliary genes
We carried out similarity searches using WU-BLAST (version 2.2.6; http://blast.wustl.edu) with e-value 0.01 and without sequence filter (without -F). Human protein sequences were taken from NCBI and used as queries (See accession number in Additional file 3). PID was calculated as the number of identical amino acids reported by WU-BLAST over the entire length of the query.

Phylogenetic analysis
Phylogenetic analysis was done using MEGA4 [42]. Multiple sequence alignment was done using CLUSTALW (included in META4) with default settings. Phylogenetic trees were inferred using the Neighbor-Joining method.

Functional domain identification and analysis
Sequences for activation, B, C, and D domains were taken from previous publications. The multiple sequence alignment was performed for each domain and used as input for hmmbuild to generate a HMM profile for each domain. hmmsearch was used to scan the proteome of selected species to find regions of similar profile. Both hmmbuild and hmmsearch are part of the HMMER suite [43]http://hmmer.janelia.org.

Additional material

Additional file 1 List of all putative RFX genes. All RFX TF genes identified and shown in Figure 5 are included in the table contained in this file. For each RFX gene, the table provides an ID (shown in the first and second columns) that is used for display in Figure 5, species name, common name of the species, phylum, proposed gene name, chromosomal coordinates, as well as the sequence of the RFX DBD identified in the gene.

Additional file 2 List of genomes processed in this project. This file lists genomes that are found to contain RFX TF genes. For each genome, the file provides the species name, database from which the genome is downloaded and the version of the database.

Additional file 3 Database information of the IFT genes used as queries. List of IFT genes and their corresponding NCBI accession numbers.

Authors' contributions
NC and DLB conceived the study. JSCC and NC conducted the experiments and wrote the manuscript. All authors have read and approved the final manuscript.

Author Details
Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada

References

1. Satir P, Mitchell DR, Jekely G: **How did the cilium evolve?** *Curr Top Dev Biol* 2008, **85**:63-82.
2. Mitchell DR: **The evolution of eukaryotic cilia and flagella as motile and sensory organelles.** *Adv Exp Med Biol* 2007, **607**:130-140.
3. Swoboda P, Adler HT, Thomas JH: **The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in C. elegans.** *Mol Cell* 2000, **5**:411-421.
4. Chen N, Mah A, Blacque OE, Chu J, Phgora K, Bakhoum MW, Newbury CR, Khattra J, Chan S, Go A, *et al.*: **Identification of ciliary and ciliopathy genes in Caenorhabditis elegans through comparative genomics.** *Genome Biol* 2006, **7**:R126.
5. Badano JL, Mitsuma N, Beales PL, Katsanis N: **The ciliopathies: an emerging class of human genetic disorders.** *Annu Rev Genomics Hum Genet* 2006, **7**:125-148.
6. Wheatley DN, Wang AM, Strugnell GE: **Expression of primary cilia in mammalian cells.** *Cell Biol Int* 1996, **20**:73-81.
7. Ou G, Koga M, Blacque OE, Murayama T, Ohshima Y, Schafer JC, Li C, Yoder BK, Leroux MR, Scholey JM: **Sensory ciliogenesis in Caenorhabditis elegans: assignment of IFT components into distinct modules based on transport and phenotypic profiles.** *Mol Biol Cell* 2007, **18**:1554-1569.
8. Pedersen LB, Rosenbaum JL: **Intraflagellar transport (IFT) role in ciliary assembly, resorption and signalling.** *Curr Top Dev Biol* 2008, **85**:23-61.
9. Dorn A, Durand B, Marfing C, Le Meur M, Benoist C, Mathis D: **Conserved major histocompatibility complex class II boxes--X and Y--are transcriptional control elements and specifically bind nuclear proteins.** *Proc Natl Acad Sci USA* 1987, **84**:6249-6253.
10. Efimenko E, Bubb K, Mak HY, Holzman T, Leroux MR, Ruvkun G, Thomas JH, Swoboda P: **Analysis of xbx genes in C. elegans.** *Development* 2005, **132**:1923-1934.
11. Blacque OE, Perens EA, Boroevich KA, Inglis PN, Li C, Warner A, Khattra J, Holt RA, Ou G, Mah AK, *et al.*: **Functional genomics of the cilium, a sensory organelle.** *Curr Biol* 2005, **15**:935-941.
12. Laurencon A, Dubruille R, Efimenko E, Grenier G, Bissett R, Cortier E, Rolland V, Swoboda P, Durand B: **Identification of novel regulatory factor X (RFX) target genes by comparative genomics in Drosophila species.** *Genome Biol* 2007, **8**:R195.
13. Durand B, Vandaele C, Spencer D, Pantalacci S, Couble P: **Cloning and characterization of dRFX, the Drosophila member of the RFX family of transcription factors.** *Gene* 2000, **246**:285-293.
14. Otsuki K, Hayashi Y, Kato M, Yoshida H, Yamaguchi M: **Characterization of dRFX2, a novel RFX family protein in Drosophila.** *Nucleic Acids Res* 2004, **32**:5636-5648.
15. Reith W, Satola S, Sanchez CH, Amaldi I, Lisowska-Grospierre B, Griscelli C, Hadam MR, Mach B: **Congenital immunodeficiency with a regulatory defect in MHC class II gene expression lacks a specific HLA-DR promoter binding protein, RF-X.** *Cell* 1988, **53**:897-906.
16. Aftab S, Semenec L, Chu JS, Chen N: **Identification and characterization of novel human tissue-specific RFX transcription factors.** *BMC Evol Biol* 2008, **8**:226.
17. Emery P, Durand B, Mach B, Reith W: **RFX proteins, a novel family of DNA binding proteins conserved in the eukaryotic kingdom.** *Nucleic Acids Res* 1996, **24**:803-807.
18. Bonnafe E, Touka M, AitLounis A, Baas D, Barras E, Ucla C, Moreau A, Flamant F, Dubruille R, Couble P, *et al.*: **The transcription factor RFX3 directs nodal cilium development and left-right asymmetry specification.** *Mol Cell Biol* 2004, **24**:4417-4427.
19. Ait-Lounis A, Baas D, Barras E, Benadiba C, Charollais A, Nlend Nlend R, Liegeois D, Meda P, Durand B, Reith W: **Novel function of the ciliogenic transcription factor RFX3 in development of the endocrine pancreas.** *Diabetes* 2007, **56**:950-959.
20. King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, *et al.*: **The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans.** *Nature* 2008, **451**:783-788.
21. Wickstead B, Gull K: **Dyneins across eukaryotes: a comparative genomic analysis.** *Traffic* 2007, **8**:1708-1721.
22. Pedersen LB, Veland IR, Schroder JM, Christensen ST: **Assembly of primary cilia.** *Dev Dyn* 2008, **237**:1993-2006.
23. Pan J: **Cilia and ciliopathies: from Chlamydomonas and beyond.** *Sci China C Life Sci* 2008, **51**:479-486.
24. Dawson SC, Sagolla MS, Mancuso JJ, Woessner DJ, House SA, Fritz-Laylin L, Cande WZ: **Kinesin-13 regulates flagellar, interphase, and mitotic microtubule dynamics in Giardia intestinalis.** *Eukaryot Cell* 2007, **6**:2354-2364.
25. Wakasugi M, Ohta J: **Studies on the amoebo-flagellate transformation in Physarum polycephalum.** *Journal of Plant Research* 1973, **86**:299-308.
26. Wright M, Moisand A, Mir L: **The structure of the flagellar apparatus of the swarm cells ofPhysarum polycephalum.** *Protoplasma* 1979, **100**:231-250.
27. Aldrich HC: **The development of flagella in swarm cells of the myxomycete Physarum flavicomum.** *J Gen Microbiol* 1968, **50**:217-222.
28. Ginger ML, Portman N, McKean PG: **Swimming with protists: perception, motility and flagellum assembly.** *Nat Rev Microbiol* 2008, **6**:838-850.
29. Briggs LJ, Davidge JA, Wickstead B, Ginger ML, Gull K: **More than one way to build a flagellum: comparative genomics of parasitic protozoa.** *Curr Biol* 2004, **14**:R611-612.
30. Cracraft J, Donoghue MJ: *Assembling the tree of life* New York: Oxford University Press; 2004.
31. James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J, *et al.*: **Reconstructing the early evolution of Fungi using a six-gene phylogeny.** *Nature* 2006, **443**:818-822.
32. Huang M, Zhou Z, Elledge SJ: **The DNA replication and damage checkpoint pathways induce transcription by inhibition of the Crt1 repressor.** *Cell* 1998, **94**:595-605.
33. Gajiwala KS, Chen H, Cornille F, Roques BP, Reith W, Mach B, Burley SK: **Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding.** *Nature* 2000, **403**:916-921.
34. Lubelsky Y, Reuven N, Shaul Y: **Autorepression of rfx1 gene expression: functional conservation from yeast to humans in response to DNA replication arrest.** *Mol Cell Biol* 2005, **25**:10665-10673.
35. Conway-Morris S: **The Cambrian "explosion" of metazoans and molecular biology: would Darwin be satisfied?** *Int J Dev Biol* 2003, **47**:505-515.
36. Ruiz-Trillo I, Burger G, Holland PW, King N, Lang BF, Roger AJ, Gray MW: **The origins of multicellularity: a multi-taxon genome initiative.** *Trends Genet* 2007, **23**:113-118.
37. King N: **The unicellular ancestry of animal development.** *Dev Cell* 2004, **7**:313-325.
38. Manning G, Young SL, Miller WT, Zhai Y: **The protist, Monosiga brevicollis, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan.** *Proc Natl Acad Sci USA* 2008, **105**:9674-9679.
39. Abedin M, King N: **The premetazoan ancestry of cadherins.** *Science* 2008, **319**:946-948.
40. Wray GA: **The evolutionary significance of cis-regulatory mutations.** *Nat Rev Genet* 2007, **8**:206-216.
41. Miller CT, Beleza S, Pollen AA, Schluter D, Kittles RA, Shriver MD, Kingsley DM: **cis-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans.** *Cell* 2007, **131**:1179-1189.
42. Kumar S, Nei M, Dudley J, Tamura K: **MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences.** *Brief Bioinform* 2008, **9**:299-306.
43. Durbin R, Eddy S, Krogh A, mitchison G: *Biological sequence analysis: probabilistic models of proteins and nucleic acids* Cambridge University Press; 1998.
44. The Tree of Life Web Project   [http://tolweb.org]
45. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4**:406-425.
46. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24**:1596-1599.