

Evolution of modular intraflagellar transport from a coatomer-like progenitor

Teunis J. P. van Dam^a, Matthew J. Townsend^b, Martin Turk^{c,1}, Avner Schlessinger^{c,2}, Andrej Sali^{c,d,e}, Mark C. Field^{b,3}, and Martijn A. Huynen^{a,3,4}

^aCentre for Molecular and Biomolecular Informatics, Radboud University Medical Centre, 6500 HB, Nijmegen, The Netherlands; ^bDepartment of Pathology, University of Cambridge, Cambridge CB2 1QP, United Kingdom; and ^cDepartment of Bioengineering and Therapeutic Sciences, ^dDepartment of Pharmaceutical Chemistry, and ^eCalifornia Institute for Quantitative Biosciences, University of California, San Francisco, CA 94158

Edited by Russell F. Doolittle, University of California at San Diego, La Jolla, CA, and approved March 15, 2013 (received for review December 4, 2012)

The intraflagellar transport (IFT) complex is an integral component of the cilium, a quintessential organelle of the eukaryotic cell. The IFT system consists of three subcomplexes [i.e., intraflagellar transport (IFT)-A, IFT-B, and the BBSome], which together transport proteins and other molecules along the cilium. IFT dysfunction results in diseases collectively called ciliopathies. It has been proposed that the IFT complexes originated from vesicle coats similar to coat protein complex (COP) I, COPII, and clathrin. Here we provide phylogenetic evidence for common ancestry of IFT subunits and α , β' , and ϵ subunits of COPI, and trace the origins of the IFT-A, IFT-B, and the BBSome subcomplexes. We find that IFT-A and the BBSome likely arose from an IFT-B-like complex by intracomplex subunit duplication. The distribution of IFT proteins across eukaryotes identifies the BBSome as a frequently lost, modular component of the IFT. Significantly, loss of the BBSome from a taxon is a frequent precursor to complete cilium loss in related taxa. Given the inferred late origin of the BBSome in cilium evolution and its frequent loss, the IFT complex behaves as a “last-in, first-out” system. The protocoatomer origin of the IFT complex corroborates involvement of IFT components in vesicle transport. Expansion of IFT subunits by duplication and their subsequent independent loss supports the idea of modularity and structural independence of the IFT subcomplexes.

complex modularity | molecular evolution

The eukaryotic cilium or flagellum is a structure protruding from the cell into the environment. The cilium provides motility by a controlled whip-like or rotational beating. Construction and maintenance of the cilium, together with additional signaling functions, depend on the process of intraflagellar transport (IFT). IFT provides active, bidirectional transport of proteins and other molecules along the length of the cilium, delivering structural components and other factors in the organelle. IFT dysfunction results in the inability of the cilium to maintain a normal structure and failure of signaling and sensory pathways, causing complex system-wide disorders and syndromes (1).

IFT is mediated by a large cohort of evolutionarily conserved subunits, which can be grouped by biochemical and genetic criteria into three subcomplexes: IFT-A, IFT-B, and BBSome. Broadly, mutations in any subunit of each of these complexes phenocopy each other, indicating close cooperativity and a requirement for complete holocomplexes for functional IFT. Significantly, six IFT complex subunits (WDR19, WDR35, IFT140, IFT122, IFT172, and IFT80) have predicted secondary structure elements and folds similar to those present in multiple subunits of vesicle coat complexes and the nuclear pore complex (NPC) (2–4). Their N-terminal region contains WD40 repeats, likely forming two β -propeller folds, whereas their C-terminal region contains tetratricopeptide repeats (TPR), likely forming an α -solenoid-like fold.

The IFT system has been shown to be homologous to the protocoatomer family of complexes, which includes coat protein complex (COP) I, COPII, clathrin/adaptin complex, and the

NPC scaffold (2–4). This classification was based on sequence similarity of IFT subunits to the COPI- α and - β' subunits, further supported by secondary structure predictions. However, a full phylogenetic reconstruction and structural analysis of the IFT complex has not been performed. Such an analysis is necessary because the abundance of the WD40 and TPR domains in non-coatomer subunit proteins requires more than sequence similarity to establish a close phylogenetic relationship. Here, we have reconstructed the evolution of the IFT complex in detail, and provide phylogenetic evidence that the IFT complex is indeed a sister structure to COPI. Analysis of the presence of the individual subcomplexes in currently living eukaryotes shows that the presence and inferred order of the loss of subcomplexes mirrors their origin—the IFT subcomplex that was added latest in evolution is the first to be lost.

Results

The known IFT system consists of three subcomplexes, IFT-A, IFT-B, and BBSome, together comprising 33 subunits in *Homo sapiens* ($n = 7$, $n = 17$, and $n = 10$, respectively). Twenty-one of these subunits can be divided into four groups based on homology relationships and predicted structures (Fig. 1A). The first group (Fig. 1A, blue) comprises WDR19, WDR35, IFT140, IFT122, IFT172, and IFT80, whose domain structure resembles COP- α and - β subunits (2–4) (as detailed later). For brevity, we will henceforth refer to these proteins as the $\alpha\beta$ -IFT subunits. The second group (Fig. 1A, yellow) comprises TTC21, IFT88, TTC26, TTC30A/B, BBS4, and BBS8, whose domain structure resembles the COP- ϵ subunit and are henceforth referred to as ϵ -IFT subunits. The third group (Fig. 1A, red) comprises the small GTPases IFT22, IFT27, and BBS3. Finally, the fourth group (Fig. 1A, green) comprises BBS1, BBS2, BBS7, and BBS9, and represents four homologous subunits in the BBSome. The remaining IFT subunits (Fig. 1A, white) do not share any detectable sequence relationships with each other, or with any other proteins. Hence, as they do not contain any phylogenetic information on the origin of the IFT complex, they will not be further discussed. Interestingly, members of the four homologous groups

Author contributions: T.J.P.v.D., M.C.F., and M.A.H. designed research; T.J.P.v.D., M.J.T., M.T., and A. Schlessinger performed research; T.J.P.v.D., M.J.T., M.T., A. Schlessinger, A. Sali, and M.C.F. analyzed data; and T.J.P.v.D., M.T., A. Schlessinger, A. Sali, M.C.F., and M.A.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹Present address: Gene Center and Center for Integrated Protein Science Munich and Department of Biochemistry, University of Munich, 81377 Munich, Germany.

²Present address: Department of Pharmacology and Systems Therapeutics and Tisch Cancer Institute, Mount Sinai School of Medicine, New York, NY 10029.

³M.C.F. and M.A.H. contributed equally to this work.

⁴To whom correspondence should be addressed. E-mail: huynen@cmbi.ru.nl.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1221011110/-DCSupplemental.

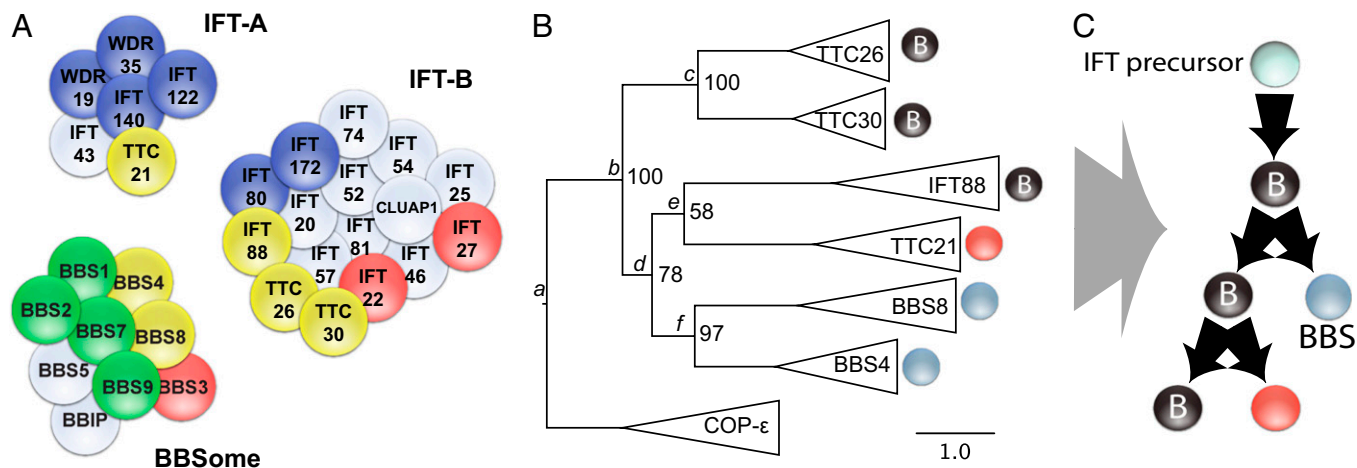


Fig. 1. Phylogenetic analyses of the ϵ -IFT subunits and IFT complex composition. (A) Composition of the IFT subcomplexes. Blue, $\alpha\beta$ -IFT subunits with domain structures similar to COPI- α and - β' ; yellow, ϵ -IFT subunits with domain structures similar to COPI- ϵ ; red, small GTPases; green, putative β -propeller BBS subunits; white, subunits that are not homologous to other subunits. Positions of the subunits do not reflect their actual positions within the IFT complex. (B) Phylogenetic tree of ϵ -IFT subunits. (C) Evolutionary scenario for the origin of IFT-A, IFT-B, and BBSome subcomplexes, based on B.

are not confined to a specific subcomplex, indicating a convoluted origin of the three subcomplexes. Here we discuss two of these groups, the $\alpha\beta$ -IFT and ϵ -IFT subunits, and report an evolutionary reconstruction of their origin. Discussions of the other two groups are provided in *SI Discussion*.

Common Descent of IFT and COPI- α , - β' , and - ϵ Subunits. Sensitive sequence similarity searches [i.e., hidden Markov models (HMMs) and PSI-BLAST] using the sequences of $\alpha\beta$ -IFT subunits (Fig. 1A, blue) as queries retrieved many TPR- and WD40-containing protein sequences, including the α and β' subunits of the COPI complex. However, none of these were retrieved consistently for all $\alpha\beta$ -IFT subunits. This lack of consistency in detection of proteins that are most similar to the $\alpha\beta$ -IFT subunits argues for a phylogenetic approach to identify the origin of the $\alpha\beta$ -IFT subunits. Unfortunately, variability in the number and length of the WD40 and TPR domains within the $\alpha\beta$ -IFT subunits prevents unambiguous alignment of these sequences. To overcome this, we searched for a common region of sequence similarity among the $\alpha\beta$ -IFT subunits. We detected a region of ~ 150 aa residues that lies between the β -propeller and α -solenoid-like segments in all $\alpha\beta$ -IFT subunits, and that aligned consistently without the need to insert long gaps into the alignment (Fig. 2 and Fig. S1). Iterative similarity searches with the use of an HMM for this region resulted in the retrieval of all $\alpha\beta$ -IFT subunits as well as, importantly, the α and β' subunits of the COPI complex. No other significant hits were retrieved, strongly suggesting a common origin for the $\alpha\beta$ -IFT and COPI- α and COP- β' subunits.

For the ϵ -IFT subunits (Fig. 1A, yellow), we used an HMM vs. HMM search to determine whether the ϵ -IFT subunits are each other's closest paralogs and whether the COP- ϵ subunit indeed represents the closest non-IFT subunit relative. HMMs for each of the ϵ -IFT subunits and COP- ϵ were constructed and added to the complete set of HMMs of protein families in the Panther database (5). In most comparisons (Table S1), COP- ϵ and the ϵ -IFT subunits represent reciprocal best hits, suggesting that COP- ϵ is indeed the closest non-IFT paralog of the ϵ -IFT subunits.

We subsequently constructed multiple sequence alignments and phylogenetic trees for the $\alpha\beta$ - and ϵ -IFT subunits. Importantly, inclusion of COP- α , COP- β' , and COP- ϵ sequences allowed us to root the phylogenetic trees and infer the order in which the individual $\alpha\beta$ - and ϵ -IFT subunits originated. The topology of the ϵ -IFT phylogeny (Fig. 1B) suggests that the proto-IFT complex

was IFT-B-like (the IFT-B subunits can be found in both clades originating in node *b*, whereas the BBSome and IFT-A subunits emerge later). BBSome subunits BBS4 and BBS8 originate from a duplication at node *d* followed by a duplication in node *f*, suggesting that the BBSome subcomplex emerged later in the proto-IFT complex. Duplication of the ancestral ϵ -IFT subunit at node *e* gave rise to IFT88 (IFT-B) and TTC21 (IFT-A), suggesting that the IFT-A subcomplex is the latest addition to the proto-IFT complex and completes the extant IFT system. Fig. 1C shows a cartoon representation of the sequence of subcomplex emergence.

The $\alpha\beta$ -IFT phylogenetic tree is not fully resolved and supports two distinct evolutionary scenarios with respect to the order in which the subcomplexes originated (*SI Discussion* provides more details), one of which is congruent with the scenario for the ϵ -IFT subunits.

Loss of IFT Subcomplexes Reflects Modularity Within IFT Complex.

The full IFT system is not retained in all eukaryotic species. In fact, cilium loss has occurred in multiple taxa, including Apicomplexa, most fungi, and seed plants (6, 7). To obtain a high-resolution picture of IFT presence and loss, we searched for orthologues of known IFT subunits in a selected set of 52 sequenced genomes of divergent eukaryotic species by using sensitive homology detection methods, including PSI-BLAST and HMMs (Fig. 3). We included ciliate and nonciliate species to determine the exclusiveness of IFT subunits to ciliated species. All the subunits reported for the human IFT complexes are conserved throughout the eukaryotic lineage. Therefore, IFT-A, IFT-B, and BBSome were likely present in the last eukaryotic common ancestor (LECA) and comprised all currently known IFT subunits from human and *Chlamydomonas reinhardtii*, in agreement with earlier observations (4, 7).

Despite the correlation between the presence of IFT subunits and the cilium, IFT subunits are not universally conserved in all ciliated species (Fig. 3). Most interesting is the loss of the BBSome in *Batrachomyxium dendrobatidis*, *Selaginella moellendorffii*, *Physcomitrella patens*, *Thalassiosira pseudonana*, and *Toxoplasma gondii* (*SI Discussion* provides a detailed description of these species). These species represent four independent lineages in eukaryotic evolution, and hence the losses represent separate events. Interestingly, all the species lacking the BBSome are closely related to species that have also lost the entire cilium

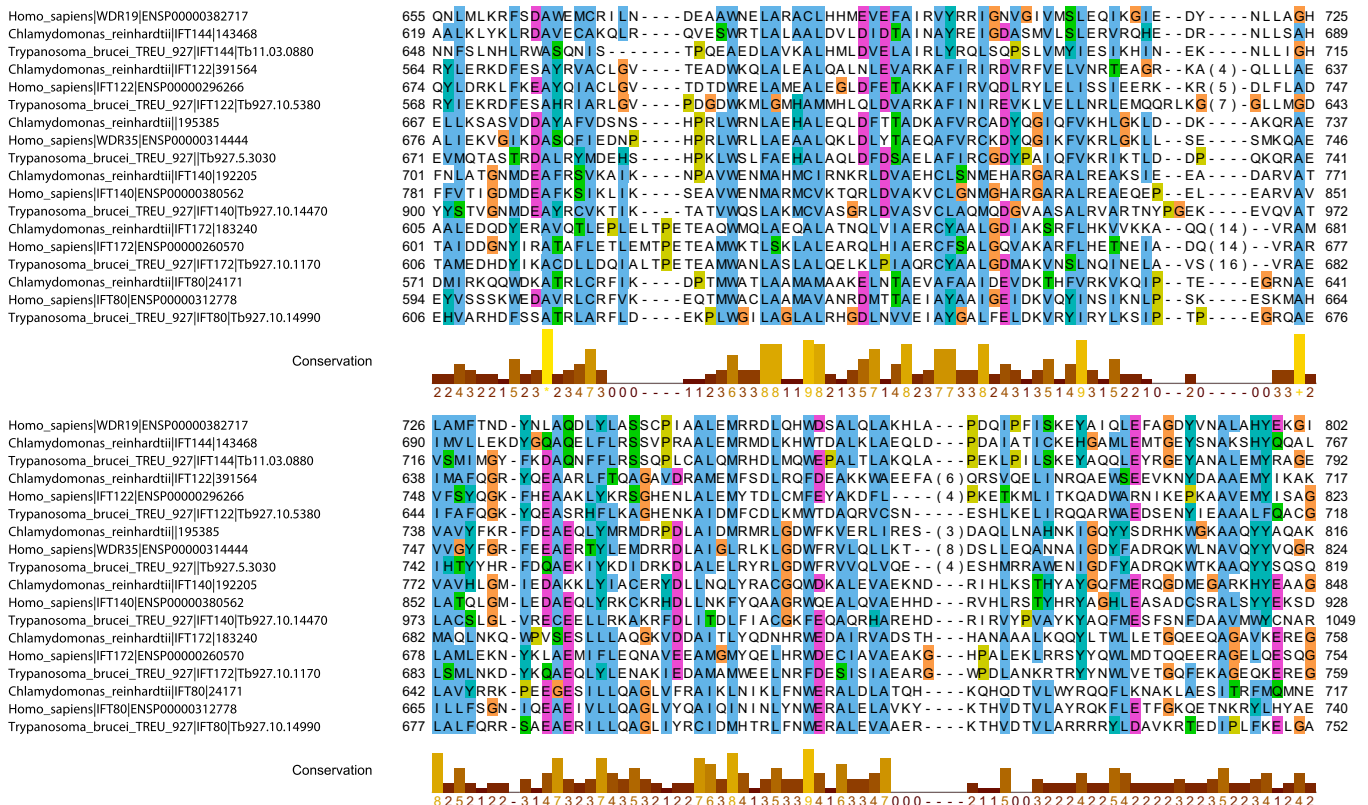


Fig. 2. Multiple sequence alignment of the $\alpha\beta$ -IFT conserved region extracted from the full alignment. The full alignment contains 52 sequences. Larger inserts have been removed and are represented by number of residues removed between parentheses. An overview of the whole alignment is shown in Fig. S1.

(fungi, seed plants, *Phaeodactylum tricornutum*, and *Cryptosporidium parvum*, respectively).

This pattern of BBSome loss thus appears to precede the loss of the cilium, and may indicate a reduced role for cilia in BBSome-negative lineages before the cilium is lost entirely. The existence of multiple species with functional flagella, but lacking the BBSome, suggests that the BBSome is a nonessential component of IFT. The “moderate” importance of the BBSome is reflected in the viable, albeit sometimes severe, phenotypes associated with Bardet–Biedl syndrome in humans (8).

Disrupting the expression of BBSome subunits has profound effects on the other IFT complexes. BBSome dysfunction results in instability and incorrect assembly of the IFT complex, resulting in dissociated IFT-A and IFT-B complexes (9). This suggests that there is functional interaction between the BBSome and IFT-A and B. However, from our analysis, it appears that the removal of the BBSome can be tolerated in some species, indicating that this functional interaction must be nonessential. It will be interesting to understand how species compensate for loss of the BBSome, and what evolutionary steps are required to facilitate that loss. Consequently, this may provide insights into possible treatments for patients with Bardet–Biedl syndrome.

Further, the secondary loss of individual subunits observed in each subcomplex indicates that there may be a tolerance within the system for subunit loss. The phylogenetic patterns are potentially correlated with severity of mutant phenotype or the inner structure of the IFT complexes. Indeed, the pattern appears to be nonrandom: subunits of the IFT-B subcomplex that suffer the most losses (IFT74, IFT27, IFT22, and IFT25) are part of the salt-stable core (10–15) ($P = 0.029$, Fisher exact test; Table S2). Nevertheless, we did not observe a correlation between the number of losses of a subunit and the severity of its phenotype as measured by the severity of cilia length reduction ($P = 0.34$,

Fisher exact test for the IFT-B complex; Table S2). This suggests that the conservation of an IFT subunit may depend more upon the structure of the IFT complex rather than the severity of the phenotype alteration in the mutant. Further biochemical research into the structures and mechanisms of the IFT may provide an explanation for this counterintuitive observation.

Orthologous IFT and BBSome Subunits from *Trypanosoma brucei* and *H. sapiens* Are Generally Conserved in Sequence and Structure.

The variability in protein length and domain composition between homologous IFT subunits requires us to determine to which extent protein structure is conserved between subunits, as well as for orthologous subunits between eukaryotic species. We compared the sequences of the IFT and BBSome subunits in *T. brucei* and *H. sapiens* as well as their various predicted structural features, including secondary structure segments, disordered regions, coiled-coil regions, TPR repeats, and folds (Fig. 4 and Fig. S2). The orthologous IFT and BBSome subunit sequences are well conserved, despite the large evolutionary distances between them, and despite the variable presence of the subunits per species. For example, human IFT172 (1,749 residues) and *T. brucei* IFT172 (1,747 residues) have sequence identity of 41%, even though *T. brucei* belongs to the Excavata, an arguably early branching supergroup of eukaryotes (16, 17). The high similarity between the predicted secondary structure elements suggests that the orthologous proteins in the IFT complex are structurally conserved to a remarkable level.

We further explored structural similarities by comparing predicted coiled coils, disordered regions, TPR repeats, and fold types. Such an approach was instrumental in proposing a common ancestral protocoatomer for coated vesicle and nuclear pore subcomplexes, despite weak sequence similarities among the constituent subunits (2, 18). As for the sequences and secondary

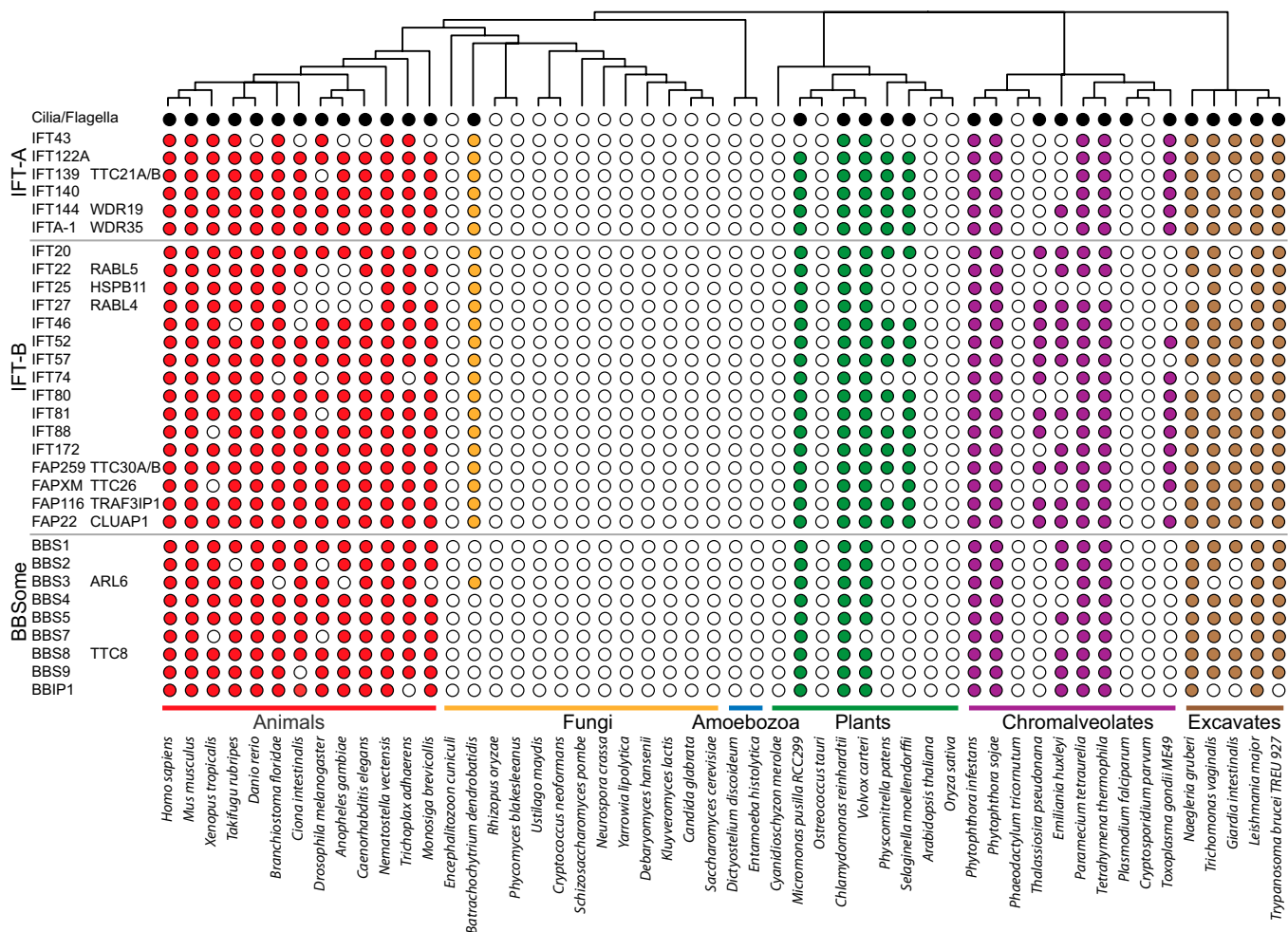


Fig. 3. Coulson plot demonstrating presence and absence (or loss) of IFT subunits in 52 eukaryotic genomes. Complexes are divided into IFT-A and -B and BBSome (rows) and taxa are displayed as columns. Super groups are color-coded for clarity, and phylogenetic relationships are shown at the top schematically. The presence of a cilium is also shown in the top row (black).

structure segments, the fold types are also conserved between orthologues from human and *T. brucei*. Moreover, fold types are likewise conserved among subunits within each of the three subcomplexes. Thus, the IFT and BBSome subunits can be organized into three structural classes as follows (Fig. 4). First, the ϵ -IFT subunits are all- α proteins with α -solenoid-like/TPR repeats, some of which contain disordered and/or coiled-coil regions (e.g., IFT88). Second, the $\alpha\beta$ -IFT subunits consist of at least one β -propeller fold, followed by an α -helical region that may contain TPR repeats (e.g., WDR35). The exception is IFT80, which is predicted to contain only a short α -helical region unlikely to be a TPR repeat. The helical C terminus of the *T. brucei* IFT80 is longer than that of *H. sapiens*, perhaps indicating a significant structural difference within this one orthologous pair. Third, the BBSome subunits contain a β -propeller fold, followed by a short coiled-coil region and a C terminus that is rich in β -strands and α -helices. The folds of the C-terminal β -strand and α -helix-rich regions cannot be assigned reliably, although the α -helix-rich region in BBS9/PTHB1 exhibits distant sequence similarity to tropomyosin (Fig. S2).

Discussion

Our findings on the origin of the IFT subcomplexes and their subsequent loss in various lineages have implications for IFT evolution and structure. The phylogenetic reconstruction of the origins of the IFT subunits and observed modularity in the

presence/absence profiles indicates that gain and loss of IFT components most likely occurred in distinct modular steps (Fig. 5). With respect to origins and acquisition of the IFT system, our results suggest that the BBSome and IFT-A emerged from an IFT-B-like complex by intracomplex duplications. Whether the IFT-A or the BBSome was the first additional subcomplex to emerge is unresolved at this time.

With respect to subsequent secondary losses, the apparent modularity of the IFT subcomplexes implies a distinct order to the loss of these factors, which, in some lineages, progressed to the loss of the cilium. We identified at least four independent loss events for the BBSome in *B. dendrobatidis*, *P. patens*, *S. moellendorffii*, *T. gondii* and *T. pseudonana*. All these taxa are closely related to species that have lost the cilium and IFT genes altogether (fungi, spermatophytes, *C. parvum*, and *P. tricornutum*, respectively; Fig. 3). Additionally, *T. pseudonana* has lost the IFT-A subcomplex. These observations suggest that the BBSome may be dispensed with while maintaining a level of cilium function. Subsequently the IFT-A subcomplex can be lost (*T. pseudonana*) before complete loss of IFT (*Plasmodium falciparum*) and of the cilium (*C. parvum*). These observations indicate that IFT-B could be viewed as the most critical subcomplex, as it is the last to be retained, and hence its presence essentially dictates if a cilium is present (Fig. 5). Stepwise emergence and loss suggests that IFT is an example of a “last-in,

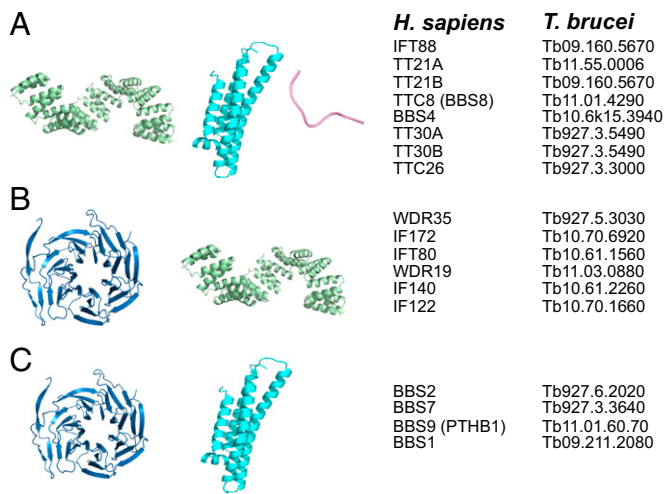


Fig. 4. Predicted fold types present within the IFT and BBSome subunits in *H. sapiens* and *T. brucei*. (A) ϵ -IFT subunits (yellow in Fig. 1A), (B) $\alpha\beta$ -IFT subunits (blue in Fig. 1A), and (C) BBSome subunits (green in Fig. 1A). The β -propeller fold is indicated in blue, the TPR/ α -solenoid-like fold in green, a disordered region in pink, and a coiled coil in cyan.

first-out” evolutionary system, i.e., whereby modules added last are also the first ones to be lost.

A protocoatomer origin for IFT provides a rationale for involvement of IFT subunits in coated vesicle transport to the ciliary base (19). Subunits of the IFT and BBSome are implicated in various transport pathways, and, therefore, with respect to function, IFT has not fallen far from the coatomer tree. Whether IFT can function as a separate coatomer-like structure and whether individual IFT subunits associate with other coatomer complexes remains to be resolved; significantly, the BBSome has been suggested to function as a coatomer complex, as it has been shown to localize on to membranes and to assemble an electron dense coat (20).

The phylogenies of homologous IFT subunits provide a framework for elucidating IFT subunit assembly within the cilium.

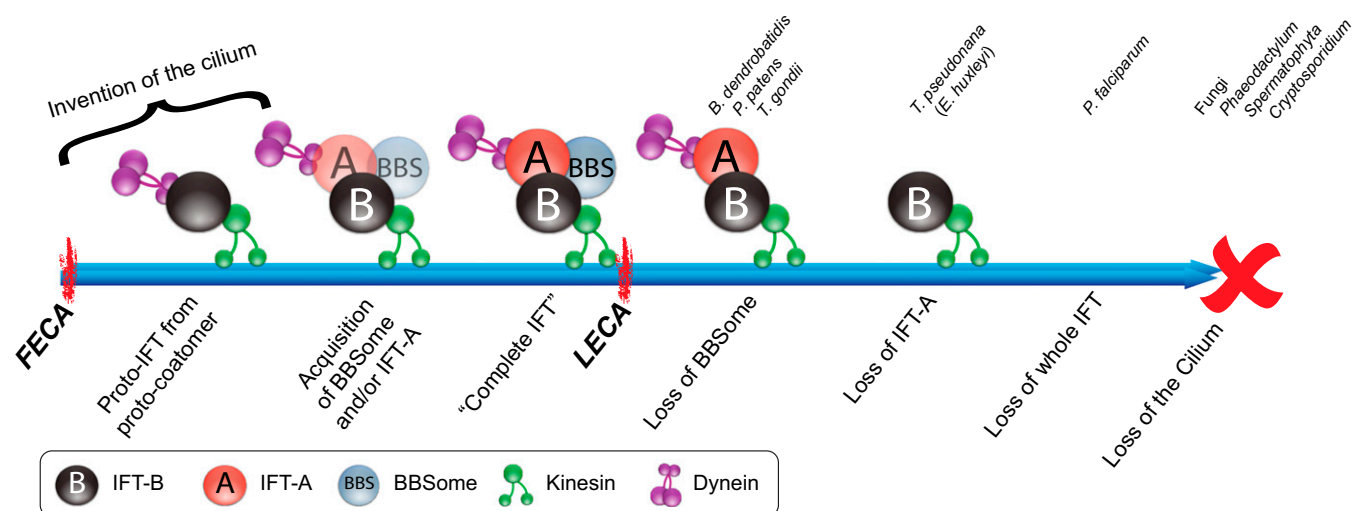


Fig. 5. Predicted origin and subsequent loss of each IFT subcomplex based on the phylogenetic trees for the COP- α -, COP- β -, and COP- ϵ -like IFT subunits, as well as the presence/absence profiles of individual subunits. As a result of the absence of BBSome subunits in the $\alpha\beta$ -like IFT phylogenetic tree, it is uncertain whether IFT-A or BBSome emerged as the second IFT subcomplex. The last common eukaryotic ancestor (LECA) already contained the complete set of IFT subunits observed in *C. reinhardtii*, *Leishmania major*, and humans, and this must have evolved in the transition between the last common eukaryotic ancestor and the first common eukaryotic ancestor (FECA). Species related to noniliated species have lost the BBSome and, in one case, have also lost IFT-A.

Mimicking the coatomer complexes, the $\alpha\beta$ - and ϵ -IFT subunits likely have structural roles in IFT. By analogy to their homologous COPI subunits, they are likely to bind in a head-to-head and tail-to-tail configuration. Close paralogs could be expected to bind each other directly or to produce module variants by mutually exclusive binding with identical partners (21). The phylogenies therefore may assist by constraining the number of potential subunit arrangements. Although module variants remain to be described for IFT, there is a possibility that recently duplicated IFT subunits, such as TTC30A, TTC30B, TTC21A, and TTC21B, interact with the IFT in a mutually exclusive manner, further diversifying IFT function.

In conclusion, our results formally unite IFT with the coatomer protein complexes and the NPC, folding them into the protocoatomer family, as well as demonstrating that IFT is closely related to the COP I complex. Our phylogenetic reconstruction provides compelling evidence for functional as well as structural modularity within the IFT complex. Furthermore, the complex evolution of the IFT and its origin from a protocoatomer complex provides a keystone for understanding how the eukaryotic cell was able, by repurposing existing pathways and complexes, to evolve such a complex and highly organized organelle as the cilium.

Methods

Sequence Searches and Phylogenetic Analysis. We gathered protein sequences of 52 genomes of ciliated and noniliated eukaryotes (Table S3). Orthologous IFT subunits were identified by OrthoMCL (version 2.0), followed by manual refinement based in part on HMMER (22) and PSI-BLAST (23) searches to find additional orthologues. Absences of subunits were checked against the respective genome with TBLASTN and available EST databases. The orthologous sequences were pruned to a limited set of diverse species to exclude problematic sequences but retain a wide phylogenetic coverage. Initial alignments were made with MAFFT LINSI (24). In the initial $\alpha\beta$ -like IFT subunit alignment, we observed a conserved region among all IFT subunits (SI Discussion). A custom HMM was made and was used to search against our 52 genomes. The final alignments were made by first aligning orthologous sequences for each IFT and COPI subunit and then progressively aligned with each other by OPAL (25) by using a neighbor-joining tree to guide the order of adding alignments. The neighbor-joining tree was constructed by using QuickTree (26) and the initial MAFFT-based alignment. The resulting alignments were then analyzed by using PhyML (27), RAxML (28), QuickTree, and PhyloBayes (29). The appropriate model of evolution (LG) was determined by PROTTEST (30). Proportion of invariable sites and

γ -distribution shape parameters were estimated. The alternative topologies for the $\alpha\beta'$ -like IFT subunits were created manually, and branch lengths were recalculated by using RAxML. Site log-likelihoods were calculated by using RAxML. Consel (31) was used to calculate the probabilities as shown in Table S4. Sequence similarity between the IFT and COP- ϵ subunits was detected by using HHsearch (32). Custom HMM models were constructed by using HMMER (v2). HMM models of the Panther database (5) were used as background to which the custom HMM models were compared.

Structure Analysis. IFT sequences were analyzed by using sequence-based methods for predicting disordered regions [IUPred; with default parameters (33)], coiled coils [MARCOIL; at threshold 90% (34)], and secondary structure elements [PSIPRED; with five PSI-BLAST iterations (35)]. In addition, the folds of full-length IFT sequences and their domains were predicted by the fold recognition servers pDomTHREADER/mGenTHREADER (36) and Phyre (37) (using the default parameters), as well as the comparative modeling server ModWeb (<http://salilab.org/modweb>); template selection performed by using sequence-sequence, sequence-profile, and profile-profile methods,

with an E-value threshold of 1.0 (38) and TPRpred, a profile-sequence comparison tool trained for TPR repeats (39). Based on an initial inspection of the disorder, secondary structure, and fold predictions for full-length sequences, we estimated the domain boundaries for select sequences and resubmitted domain sequences to the fold assignment servers. High confidence fold predictions [Phyre (estimated precision $\geq 75\%$); pDomTHREADER (certainty of certain or score ≥ 6.2); MODWEB (Z-DOPE < 0 or sequence identity $\geq 30\%$)] from individual servers were confirmed with the Pfam database entries. The final folds were assigned if more than one server predicted the same fold for a particular domain.

ACKNOWLEDGMENTS. We thank Isabel Duarte for comments on the phylogenetic analysis, Mike Rout for discussions on the evolution of the nuclear pore complex and related systems, and the research groups of Roepman and Kremer for valuable input on IFT function. This study was supported by the European Community's Seventh Framework Programme FP7/2009 under Grant 241955, SYSCILIA, Wellcome Trust Grant 082813 (to M.C.F.), and National Institutes of Health/National Institute of General Medical Sciences Grant U54 RR022220 (to A.S.).

- Duldulua NA, Li J, Sun Z (2010) Cilia in cell signaling and human disorders. *Protein Cell* 1:726–736.
- Devos D, et al. (2004) Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS Biol* 2(12):e380.
- Avidor-Reiss T, et al. (2004) Decoding cilia function: Defining specialized genes required for compartmentalized cilia biogenesis. *Cell* 117(4):527–539.
- Jékely G, Arendt D (2006) Evolution of intraflagellar transport from coated vesicles and autogenous origin of the eukaryotic cilium. *Bioessays* 28(2):191–198.
- Thomas PD, et al. (2003) PANTHER: A browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res* 31(1):334–341.
- Hodges ME, Wickstead B, Gull K, Langdale JA (2012) The evolution of land plant cilia. *New Phytol* 195(3):526–540.
- Briggs LJ, Davidge JA, Wickstead B, Ginger ML, Gull K (2004) More than one way to build a flagellum: Comparative genomics of parasitic protozoa. *Curr Biol* 14(15):R611–R612.
- Forsythe E, Beales PL (2013) Bardet-Biedl syndrome. *Eur J Hum Genet* 21:8–13.
- Wei Q, et al. (2012) The BBSome controls IFT assembly and turnaround in cilia. *Nat Cell Biol* 14(9):950–957.
- Lucker BF, Miller MS, Dziedzic SA, Blackmarr PT, Cole DG (2010) Direct interactions of intraflagellar transport complex B proteins IFT88, IFT52, and IFT46. *J Biol Chem* 285(28):21508–21518.
- Taschner M, Bhogaraju S, Lorentzen E (2012) Architecture and function of IFT complex proteins in ciliogenesis. *Differentiation* 83(2):S12–S22.
- Wang Z, Fan Z-C, Williamson SM, Qin H (2009) Intraflagellar transport (IFT) protein IFT25 is a phosphoprotein component of IFT complex B and physically interacts with IFT27 in *Chlamydomonas*. *PLoS ONE* 4(5):e5384.
- Bhogaraju S, Taschner M, Morawetz M, Basquin C, Lorentzen E (2011) Crystal structure of the intraflagellar transport complex 25/27. *EMBO J* 30(10):1907–1918.
- Fan Z-C, et al. (2010) *Chlamydomonas* IFT70/CrDYF-1 is a core component of IFT particle complex B and is required for flagellar assembly. *Mol Biol Cell* 21(15):2696–2706.
- Richey EA, Qin H (2012) Dissecting the sequential assembly and localization of intraflagellar transport particle complex B in *Chlamydomonas*. *PLoS ONE* 7(8):e43118.
- Simpson AGB, Roger AJ (2004) The real 'kingdoms' of eukaryotes. *Curr Biol* 14(17):R693–R696.
- Roger AJ, Simpson AGB (2009) Evolution: Revisiting the root of the eukaryote tree. *Curr Biol* 19(4):R165–R167.
- Devos D, et al. (2006) Simple fold composition and modular architecture of the nuclear pore complex. *Proc Natl Acad Sci USA* 103(7):2172–2177.
- Baldari CT, Rosenbaum J (2010) Intraflagellar transport: It's not just for cilia anymore. *Curr Opin Cell Biol* 22(1):75–80.
- Jin H, et al. (2010) The conserved Bardet-Biedl syndrome proteins assemble a coat that traffics membrane proteins to cilia. *Cell* 141(7):1208–1219.
- Szklarczyk R, Huynen MA, Snel B (2008) Complex fate of paralogs. *BMC Evol Biol* 8:337.
- Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14(9):755–763.
- Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–3402.
- Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30(14):3059–3066.
- Wheeler TJ, Kececioglu JD (2007) Multiple alignment by aligning alignments. *Bioinformatics* 23(13):i559–i568.
- Howe K, Bateman A, Durbin R (2002) QuickTree: Building huge Neighbour-Joining trees of protein sequences. *Bioinformatics* 18(11):1546–1547.
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52(5):696–704.
- Stamatakis A, Ludwig T, Meier H (2005) RAxML-III: A fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21(4):456–463.
- Lartillot N, Lepage T, Blanquart S (2009) PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25(17):2286–2288.
- Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* 27(8):1164–1165.
- Shimodaira H, Hasegawa M (2001) CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17(12):1246–1247.
- Söding J (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21(7):951–960.
- Dosztányi Z, Csizmek V, Tompa P, Simon I (2005) IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21(16):3433–3434.
- Delorenzi M, Speed T (2002) An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* 18(4):617–625.
- McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16(4):404–405.
- Lobley A, Sadowski MI, Jones DT (2009) pGenTHREADER and pDomTHREADER: New methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics* 25(14):1761–1767.
- Kelley LA, Sternberg MJE (2009) Protein structure prediction on the Web: A case study using the Phyre server. *Nat Protoc* 4(3):363–371.
- Pieper U, et al. (2009) MODBASE, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res* 37(database issue):D347–D354.
- Karpenahalli MR, Lupas AN, Söding J (2007) TPRpred: A tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. *BMC Bioinformatics* 8:2.