Comparative Genomics Identifies a Flagellar and Basal Body Proteome that Includes the BBS5 Human Disease Gene

Jin Billy Li,¹ Jantje M. Gerdes,² Courtney J. Haycraft,³ Yanli Fan,⁵ Tanya M. Teslovich,⁴ Helen May-Simera,⁶ Haitao Li,¹ Oliver E. Blacque,⁶ Linya Li,¹ Carmen C. Leitch,² Richard Allan Lewis,⁵ Jane S. Green,⁷ Patrick S. Partrey,⁵ Michel R. Leroux,² William S. Davidson,⁵ Philip L. Beales,⁷ Lisa M. Guay-Woodford,⁷ Bradley K. Yoder,⁴ Gary D. Stormo,¹ Nicholas Katsanis,³,² and Susan K. Dutcher¹,⁎

¹Department of Genetics
Washington University School of Medicine
St. Louis, Missouri 63110

²McKusick-Nathans Institute of Genetic Medicine
Johns Hopkins University
Baltimore, Maryland 21205

³Department of Cell Biology
University of Alabama at Birmingham
Birmingham, Alabama 35294

⁴Department of Molecular Biology and Biochemistry
Simon Fraser University
Burnaby, British Columbia
Canada V5A 1S6

⁵Department of Molecular and Translational Medicine
Departments of Medicine, Pediatrics, and Genetics
University of Alabama at Birmingham
Birmingham, Alabama 35294

⁶Departments of Molecular and Human Genetics, Ophthalmology, Pediatrics, and Medicine
Baylor College of Medicine
Houston, Texas 77030

⁷Departments of Medical Genetics and Clinical Epidemiology
Memorial University
St. John’s, Newfoundland
Canada

*Correspondence: dutcher@genetics.wustl.edu

Introduction

Cilia and flagella are central to many developmental processes in a diverse range of organisms, and disruption of their structure or function can have profound phenotypic consequences (Pazour and Witman, 2003). Among vertebrates, the primary cilium is virtually ubiquitous, comprising a “9 + 0” axoneme that consists of nine outer doublet microtubules. Centrioles can interconvert to basal bodies by assembling a transition zone, which is essential for the assembly of cilia or flagella. The transition zone serves as a docking site for intraflagellar transport (IFT) proteins and their motors, kinesin II (Cole et al., 1998; Piperno and Mead, 1997). Basal bodies can also convert to centrioles that serve as the microtubule-organizing center (MTOC) in animal and algal cells. However, land plants, fungi, and diatoms do not contain centrioles, which suggests that these organisms may have evolved other structures to organize microtubules (Pickett-Heaps, 1971). In organisms that retain centrioles, these structures are essential and play key roles (Bobinnec et al., 1999). Their removal suggests that centrioles are important to complete cytokinesis and are essential for the transition from G1 to S phase of the cell cycle (Hinchcliffe et al., 2001; Khodjakov and Rieder, 2001). Missegregation or overduplication of centrioles is associated with genomic instability (Meraldi and Nigg, 2001). Notably, neoplastic cells have increased numbers of centrioles and abnormal centriole architecture, which is similar to the morphologies observed in mutant Chlamydomonas strains (Dutcher, 2003a; Lingle and Salisbury, 1999).

Cilia and basal bodies have been implicated directly in several developmental processes including left-right asymmetry, heart development, maintenance of the renal epithelium, respiratory function, electrolyte balance in the cerebrospinal fluid, and reproductive fecundity. Nodal cilia in early development are fundamental to initiate the molecular cascade that ultimately leads to the visceral asymmetry in vertebrates (Essner et al., 2002; Tabin and Vogan, 2003; Takeda et al., 1999). A lack or dysfunction of sensory cilia in the epithelium of renal tubules underlies polycystic kidney disease, although the exact process of cyst formation is not understood (Haycraft et al., 2001; Pazour et al., 2000). In the retina, successful translocation of several molecules from the cell body to the outer segment of the photoreceptor depends on transport along a modified cilium; defective transportation of certain proteins results in the retinal...
degeneration associated with retinitis pigmentosa (Marszalek and Goldstein, 2000). Although the exact role has not been elucidated, sensory cilia present on neuronal cells have somatostatin receptors embedded in their ciliary membrane (Handel et al., 1999). Defective basal bodies have been implicated in the pathogenesis of the pleiotropic Bardet-Biedl syndrome (BBS). This multisystemic disorder is characterized primarily by retinal dystrophy, obesity, polydactyly, renal and genital malformations, and learning disabilities. Six genes associated with BBS (BBS1, BBS2, BBS4, BBS6/MKKS, BBS7, and BBS8) have been identified (Ansley et al., 2003; Badano et al., 2003a; Katsanis et al., 2000; Mykytyn et al., 2001, 2002; Nishimura et al., 2001; Slavotinek et al., 2000).

Centrioles and flagella are biochemically complex organelles. The flagellar axoneme of Chlamydomonas contains over 250 polypeptides (Dutcher, 1995a), whereas the surrounding flagellar membrane and matrix together contain over 50 polypeptides (Witman et al., 1972; our unpublished data). Preparations enriched in Chlamydomonas basal bodies show over 150 polypeptides (Dutcher, 1999b).

We undertook a comparative genomics approach to identify the eukaryotic flagellar/ciliary and basal body proteome. Our strategy rests on two observations. First, Chlamydomonas flagellar and basal body proteins show high similarity to human orthologs. Second, organisms that have lost flagella and basal bodies through evolution have also lost most of the genes needed for building and regulating these organelles (Cole, 2003; Dutcher, 2004; Perrone et al., 2003). Cumulatively, our data provide a comprehensive description of the conserved flagellar and basal body proteome and a rich platform to understand the function of these cellular structures.

Results

The recently released genome sequence of Chlamydomonas reinhardtii has approximately 95 Mb of sequence and encodes 19,832 predicted proteins (http://shake.jgi-psf.org/chlre2/chlre2.home.html). These predicted proteins were compared to the 32,035 predicted proteins of the human genome using WU-BLASTP (Lopez et al., 2003) with a cutoff E value of 10⁻10. These 4,348 matches were then compared to the proteome of Arabidopsis (blue) by WU-BLASTP to find all matches with a cutoff E value of 10⁻10. These 4,348 matches were then compared to the proteome of Chlamydomonas (yellow) to identify the eukaryotic flagellar/ciliary and basal body proteome (FABB) Proteome containing over 50 polypeptides (Witman et al., 1972; our unpublished data). Preparations enriched in the FABB proteome were then compared to the best matches from human, mouse, or Arabidopsis proteomes both present in organisms with flagella and basal bodies and absent from an organism that lacks flagella and basal bodies. This procedure was repeated using the genomes of Caenorhabditis elegans, Ciona intestinalis, Drosophila melanogaster, and Mus musculus. The comparisons with each organism yield between 400 and 688 proteins (Table 1). Using human, mouse, and Ciona in combination yielded 504 proteins, which suggests that many of the same genes are found in all of the organisms tested (Table 1). The datasets generated with Caenorhabditis elegans and Drosophila melanogaster are smaller (400 and 481, respectively) than the human, mouse, or Ciona sets. We will refer to the Chlamydomonas + Human – Arabidopsis collection of proteins as the flagellar apparatus-basal body (FABB) proteome.

Validation with Known Chlamydomonas Flagellar and Basal Body Genes

We used 58 well-studied genes that encode flagellar proteins or components of the IFT complex in Chlamydomonas (Supplemental Table S2 on the Cell website). This set of test genes excludes tubulins, kinesins, and calcium binding proteins because these proteins have required roles in all eukaryotes. Thus, any gene retained in Arabidopsis for another function should not be found in the FABB proteome. Fifty-two of the fifty-eight known flagellar and IFT genes (90%) are present in the FABB proteome. Six genes that localize or are required for the assembly of the basal body of Chlamydomonas were examined, and four are in the FABB proteome (67%).

The similarity and identity of these 64 Chlamydomonas proteins were compared to the best matches from human, mouse, Ciona, Drosophila, C. elegans, and Arabidopsis (Supplemental Table S2 online). Overall, human and mouse matches show more significant E values with Chlamydomonas proteins than with Drosophila and...
C. elegans. Importantly, C. elegans lacks motile cilia and has simpler basal bodies with singlet microtubules. Proteins that are present in *Chlamydomonas* and *Drosophila* but absent from *C. elegans* may be proteins needed for motile flagella and more complex basal bodies. There are 326 proteins of the FABB proteome that are not found in *C. elegans* (Supplemental Table S1). Consistent with this prediction, this set includes the known radial spoke, central pair, and dynein arm proteins, which are needed for motility.

In addition to established *Chlamydomonas* basal body and flagellar genes, the FABB proteome contains several known mammalian disease genes (Supplemental Table S1). Five of the six genes that result in Bardet-Biedl syndrome in humans appear in the dataset (Ansley et al., 2003; Badano et al., 2003; Mykytyn et al., 2001, 2002; Nishimura et al., 2001). Two genes that cause kidney disease in humans, NPHP4 (Mollet et al., 2002) and fibrocystin (Ward et al., 2002), are also present in the FABB proteome. Fibrocystin localizes to primary cilia of the kidney (Wang et al., 2004). Hydin, which causes hydrocephalus in mice and localizes to cilia, is present in the FABB proteome (Davy and Robinson, 2003). IFT72 and IFT88 are in the FABB proteome. Mice with mutations in these proteins show neural tube patterning defects that act through the hedgehog signaling pathway (Huangfu et al., 2003). IFT88 is also needed for proper kidney function (Pazour et al., 2002; Taulman et al., 2001). A tubby superfamily member (TUSP) is also present (Li et al., 2001).

**Validation with Transcriptional Regulation of Flagellar Genes**

Amputation of the flagella in *Chlamydomonas* is followed by transcription of flagellar genes. Consequently, a 3- to 12-fold upregulation of the mRNAs for structural components of the flagella, the IFT complex, and its molecular motors has been observed 30 min after deflagellation (Lefebvre et al., 1980). Therefore, the genes in the FABB proteome that are transcriptionally upregulated following deflagellation are likely to encode flagellar proteins. RSP3, α-tubulin, IFT57, and D2LIC serve as positive controls and β-tubulin and ε-tubulin are genes that show no upregulation (Supplemental Table S4 online). We found 38% (n = 103) of the genes tested to be upregulated 3- to 13-fold (Figure 3; Supplemental Table S4) and to be likely to play roles in flagellar assembly and function. Twenty-two genes were also tested at 10 min and at 50 min after deflagellation. Fifteen of the genes that were not upregulated at 30 min also showed no change at either 10 or 50 min. For the seven genes that were upregulated, the peak levels were at 30 min and not at 10 or 50 min.

To determine what fraction of genes not in the FABB proteome is upregulated by deflagellation, ten genes that are present in *Chlamydomonas*, human, and Arabidopsis were analyzed. This set should not be enriched for flagellar and basal body genes. Only one of the ten, 167087, is upregulated. 167087 is the homolog of SRP54 (signal recognition particle); this upregulation may reflect a need for increased protein sorting upon deflagellation.
Validation of Basal Body Genes by RNA Interference

RNA interference (RNAi) has been used successfully in *Chlamydomonas* to lower the levels of several mRNAs and thereby to produce mutant phenotypes (Huang and Beck, 2003; Pfannenschmidt et al., 2003; Sineshchekov et al., 2002; Wu-Scharf et al., 2000). Constructs expressing short (150–200 bp) hairpins loops were transformed into wild-type *Chlamydomonas*. Multiple transformants were examined to generate an allelic series.

To determine if the genes in the FABB proteome that were not upregulated after deflagellation represented false positives or bona fide basal body candidates, genes for six predicted proteins (171684, 171242, 158402, 162018, 165684, and 169453) that do not show upregulation were tested. The phenotypes of four to nine transformants for each gene were examined. We observed flagellar phenotypes in the RNAi strains for five of the six genes (Table 2). These phenotypes ranged from slowly swimming, which resembled the defect observed in inner dynein arm mutant strains, to paralyzed flagella, to lack of flagella altogether. Previously studied *Chlamydomonas* strains with defects in basal body assembly show a constellation of phenotypes (Dutcher, 2003a). Many of these mutant strains lack flagella or have abnormal numbers of flagella. Other phenotypes include incomplete basal body microtubules, disorganized cytoplasmic microtubules known as rootlet microtubules, and disorganized centrin fibers. Disorganized rootlet microtubules and centrin fibers result in misplaced cleavage furrows (Ehler et al., 1995). Three strains (171242, 158402, and 162018) show obvious defects in cleavage furrow placement (Table 2). The level of reduction in the specific mRNAs was quantified by real-time RT-PCR in duplicate experiments. Interestingly, differences in phenotypes among transformants are reflected by differences in the degree of mRNA reduction. For example, three independent 171242 transformants have different phenotypes. 171242-T1 cells completely lack flagella and have a defect in cleavage furrow placement; the mRNA is reduced 7.8-fold. 171242-T2 cells show a weaker phenotype in that they are only 59% aflagellate and do not have a cleavage furrow defect; the mRNA is reduced only 3.4-fold. Finally, 171242-T3 cells show the weakest phenotype and the mRNA is reduced only 1.4-fold. In general, the more severe phenotypes are observed in cells with greater reduction in mRNA for most of the different sets of transformants that show variable phenotypes (Table 2).

Only one of six proteins chosen for analysis has a clear ortholog. 171684, which shows no phenotype, is the homolog of echinoderm microtubule-associated protein. Perhaps, greater reductions in level of message result in lethality so that no transformants with a strong phenotype were identified. Protein 171242 has a covalent adenosyltransferase domain whose function in flagellar and basal bodies is unknown. Protein 162018 has a NIPSNAP1 domain, and its function is unclear. The remaining three proteins do not contain recognizable domains or motifs (Bateman et al., 2004).

Identifying Human Disease Genes with the FABB Proteome

Recent data suggest that several genes that cause Bardet-Biedl syndrome encode proteins involved in...
Table 2. Phenotypes Observed with RNA Interference for Potential Basal Body Genes

<table>
<thead>
<tr>
<th>Chlamydomonas Protein</th>
<th>Phenotypes of Representative Transformants N = 200 Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Two Flagella (%)</td>
</tr>
<tr>
<td>171684 Echinoderm microtubule-associated protein</td>
<td>T1 95</td>
</tr>
<tr>
<td>171242 Cobalamin adenosyltransferase domain</td>
<td>T1 0</td>
</tr>
<tr>
<td>158402 No known Pfam domain</td>
<td>T1 95</td>
</tr>
<tr>
<td>162018 NIPSNAP1 domain</td>
<td>T1 95</td>
</tr>
<tr>
<td>165684 No known Pfam domain</td>
<td>T1 95</td>
</tr>
</tbody>
</table>

basal body/flagellar assembly and function (Ansley et al., 2003). The FABB proteome contains orthologs of all known BBS genes except for BBS6, which appears to have evolved specifically in the mammalian lineage. Therefore, we hypothesized that genes in our dataset might offer useful candidates for identifying novel BBS genes and would further validate our approach.

BBS5 maps to human chromosome 2q31 between D2S156 and D2S1238; the region spans 14 Mb and contains approximately 230 predicted genes (Young et al., 1998). Two genes in this interval are present in the FABB proteome. They are FLJ11457 (NM_024753), which encodes IFT139 (Cole, 2003), and DKFZp762194 (NM_152384), which encodes a well-conserved, novel protein of unknown function. The complete open reading frame and intron-exon junctions of FLJ11457 in a patient from Newfoundland family NFB9 that defined the BBS5 locus (Young et al., 1998) was sequenced and yielded several polymorphisms, but no potentially pathogenic mutations. We then sequenced each of the 12 coding exons of DKFZp762194 and identified a homozygous A→G transition at the +3 position of the exon 6 splice donor that segregates with the BBS phenotype in the extended family. The two available affected individuals are homozygous G/G, both parents are heterozygous A/G, and four unaffected sibs are either homozygous A/A or heterozygous A/G (Figure 4A). Because this allele segregates with the disease and it is absent from 100 Newfoundland control chromosomes, its effect in a lymphoblastoid cell line from one of the NFB9 patients was investigated. Amplification of exons 4 to 9 of DKFZp762194 in a control lymphoblastoid cell line and subsequent cloning and sequencing of the PCR products revealed two RNA splice isoforms; one lacks the 63 bp exon 8 (Figure 4B). In contrast, patient NFB9 produces three RNA species. One species lacks exon 6, one species lacks exons 6 and 8, and the third uses a cryptic splice site (/GTA) in intron 5 (Figure 4B). All three messages result in a frameshift and premature termination in exon 7. Therefore, we conclude that this intronic mutation compromises the splicing of DKFZp762194 and its protein product. Thus, DKFZp762194 is a strong candidate for BBS5.

To investigate further the candidacy of this gene, we genotyped 24 Saudi Arabian consanguineous families with markers encompassing each of the known BBS loci. We determined that in one of these families, KK63, both affected sibs, but none of the unaffected individuals, were homozygous by descent (HBD) across the BBS5 critical interval. The other seven BBS loci were excluded by haplotype analysis. Sequencing DKFZp762194 in this family revealed a premature termination mutation L→X in exon 6 that is not present in 166 ethnically matched control chromosomes (Figure 4C). This mutation may lead to nonsense-mediated decay of the message. The combined data from these two families support the hypothesis that the candidate gene DKFZp762194 is BBS5.

We next ascertained the contribution of BBS5 mutations to the pool of BBS patients by screening a combined cohort of 259 independent BBS families from various ethnic backgrounds. We identified likely pathogenic mutations in two families. In consanguineous family PB108 of Turkish descent, we identified an insertion-deletion mutation, 283–271 indelGCTCTTA, which has the net effect of removing a single base and resulting in a premature termination codon (Figure 4D). In a second family, PB127 of Kurdish descent, the patient harbors a homozygous nonsense mutation W→X, which by conceptual translation will lead to premature termination (Figure 4E). These data suggest that BBS5 contributes approximately 2% to the BBS mutation pool; this estimate is consistent with the contribution of most other BBS loci (Ansley et al., 2003; Badano et al., 2003b; Beales et al., 2001; Katsanis et al., 2002).

BBS5 May Interact Genetically with BBS1

We have shown previously that some families segregate three mutations in two BBS genes and that the presence of the third mutation may be a modifier of either pene-
Figure 4. DKFZp762I194 Is BBS5

(A) Newfoundland family NFB9 segregates an A→G transition at the +3 position of the exon 6 splice donor; genotypes are shown below each individual.

(B) Diagram of exons 4 to 9 that were amplified by RT-PCR and sequenced. Unaffected (A/A) individuals have an mRNA with exons 4–9 and an alternatively spliced mRNA that lacks exon 8. Affected members (G/G) of family NFB9 have three predominant mRNA species. One lacks exon 6, one lacks exons 6 and 8, and one starts at a cryptic splice site in intron 5.

(C, D, and E) Segregation of mutations in three families: each mutation is predicted to result in premature termination. Families KK63 (Saudi) and PB108 (Turkish) are consanguineous; consanguinity is also suspected for the Kurdish family PB127.

The Search for X Boxes

In the nematode C. elegans, 60 of the 302 neurons in hermaphrodites have cilia and multiple ciliated neurons are located in the male tail (Ward et al., 1975; White et al., 1976; Wicks et al., 2000). Several genes needed for functional cilia have a 14 nucleotide sequence known as the X box, which is found in the first 150 bp upstream of the ATG initiation codon. This motif is thought to be the binding site for the RFX transcription factor DAF-19 (Schafer et al., 2003; Swoboda et al., 2000). We searched the C. elegans genes in the FABB proteome for this motif with the algorithm PATSER (Hertz and Stormo, 1999). Twenty-five of the 400 C. elegans genes contain the X box motif (data not shown). Thus, the FABB proteome is enriched for C. elegans genes containing the X box motif by about 6-fold.

Expression and Localization of BBS5

The presence of an X box in the promoter of bbs-5 suggests that it will be expressed specifically in ciliated cells and that it will be regulated by DAF-19. To test this possibility, a transcriptional reporter construct that consists of 500 bp of upstream sequence and the first two exons of bbs-5 was fused in-frame to the gfp coding region and used to generate transgenic worms. In agreement with a conserved role for BBS-5 in basal bodies at the base of cilia, GFP fluorescence was detected in all ciliated sensory neurons including the amphids, labial neurons, phasmids, and the sensory rays of the male tail (Figures 6A–6C and data not shown). No GFP was evident in other cell types, suggesting that bbs-5 expression is restricted to ciliated cells.
Figure 5. BBS5 May Interact Genetically with BBS1

(A) Both patients from North American family AR291 are homozygous for the M390R mutation in BBS1 and are heterozygous for the N184S mutation in BBS5.

(B) Schematic of the positions of the DM16 repeat domains of BBS5 and alignment of both DM16 domains contained in each BBS5 ortholog across ten species. Red depicts similarity of at least 90% in both domains, green depicts similarity of at least 80% in domain 1, and blue depicts similarity of at least 80% in domain 2. The N184 residue involved in family AR291 is conserved (asterisk). Cr = Chlamydomonas reinhardtii; Hs = Homo sapiens; Mm = Mus musculus; Rn = Rattus norvegicus; Dr = Danio rerio; Cl = Ciona intestinalis; Ce = Caenorhabditis elegans; Dm = Drosophila melanogaster; Tb = Trypanosoma brucei; Tc = Trypanosoma cruzi. Amino acid positions within each sequence are given in parentheses.

To confirm a role for DAF-19 in regulating bbs-5 transcription, the expression of bbs-5 in wild-type and daf-19 mutant worms was analyzed in vivo and by Northern blot analysis. daf-19(m86) transgenic lines carrying the bbs-5::gfp construct as an extrachromosomal array were mated to daf-19 worms. The resulting progeny were assayed for DAF-19 function based on the inability of daf-19(m86) worms to stain with DiO, a lipophilic fluorescent dye normally absorbed by a subset of ciliated sensory neurons (Figures 6F and 6G) (Fujiwara et al., 1999). Similar to the results reported for osm-5, a known DAF-19 transcriptional target, expression of the bbs-5::gfp reporter was markedly reduced but not abolished in the absence of DAF-19 (compare Figures 6D and 6E) (Haycraft et al., 2001). Northern blot analysis detected a 1.2 kb transcript that agrees with the predicted mRNA in WormBase (http://www.wormbase.org). The amount of the 1.2 kb transcript was reduced in daf-
19 mutant worms, which confirms the in vivo expression analysis (data not shown).

Analysis of a *C. elegans* strain expressing a translational *bbs-5::GFP* transgene revealed that the GFP-tagged BBS-5 protein localizes specifically to the base of cilia in the ciliated head and tail neurons. BBS-5::GFP staining pattern is observed at the base of the cilia in the amphid neurons in the worm head and at the base of the left and right phasmid neurons in the tail (Figures 6H and 6I).

To explore the function of BBS5 further, a rabbit anti-peptide antibody was generated to the predicted mouse BBS5 protein. On immunoblots, the antibody recognizes two proteins with apparent molecular weights of 37 and 40 kDa in extracts from cultured IMCD cells and whole mouse (E10.5) embryos (Supplemental Figure S1). These sizes are consistent with the proteins predicted by the conceptual translation of the mouse cDNA and alternatively spliced isoform (Figure 4B). Importantly, neither of these products is detected on immunoblots probed with the preimmune sera.

In agreement with the localization of the BBS-5::GFP in *C. elegans*, prominent immunofluorescent staining was detected at the apical surface of the multiciliated ependymal cells lining the ventricles of the brain (Figures 6L–6T). The localization to basal bodies was confirmed by colocalization of BBS5 with γ-tubulin, a centrosomal and basal body marker, and the localization of BBS5 just beneath the cilia as detected by antibodies to acetylated α-tubulin. The domain of BBS5 staining was also found to extend beneath that of γ-tubulin signal, indicating that BBS5 is also localized to regions surrounding the basal bodies. Faint staining was detected in the cilium, although it was difficult to confirm that this signal was significantly different from that of the background.

**RNA Interference of the BBS5 Homolog in *Chlamydomonas***

Multiple *Chlamydomonas* transformants were obtained with a hairpin construct for *BBS5* (AY581824). Among nine transformants, four completely lack flagella and show a weak cleavage furrow defect. The remaining five

(B and C) Confocal images showing bbs-5::gfp expression in the amphid and labial neurons and phasmids of the adult hermaphrodite. Anterior is toward the left in (A)–(K).

(D–G) Confocal images of bbs-5::gfp expression (D and E) and DiO uptake (F and G) in wild-type (D and F) and mutant (E and G) worms. Expression of bbs-5::gfp is markedly reduced in *daf-19(m86)* mutant worms (compare D and E). Similar results were observed with four independent transgenic strains.

(H–K) Images showing the GFP fluorescence patterns from strains carrying a translational (*BBS-5::GFP*) transgene; shown are the GFP staining patterns in the amphid/labial neuron (H) and phasmid neuron (I) basal bodies, along with DIC-merged images (J and K).

(L–T) Confocal images of the ependymal cells of the lining of the mouse ventricles. (L, N, and O) Antibodies to γ-tubulin are in green, preimmune serum is red, and DNA is blue. (M, N, and Q) Antibodies to γ-tubulin are in green, antibodies to mouse BBS5 are in red, and DNA is in blue. The boxes in (L) and (M) indicate the regions enlarged in (N), (O), (P), and (Q). The box in (R) indicates the region enlarged in (S) and (T). Antibodies to BBS5 are in red, antibodies to acetylated α-tubulin are in green, and DNA is in blue (R, S, and T). BBS5 is found at the base of the cilia.
transforms are partially aflagellate. The severity of phenotype correlates with the fold-reduction in mRNA levels. In the aflagellate strains with the cleavage furrow defect, the mRNA was reduced 2.2- to 3.1-fold. In contrast, the cells that show a partial flagellar assembly defect have a 1.5- to 1.7-fold reduction. These results strongly support a role for Bbs5 in flagellar and basal body function and assembly.

Discussion

The study of model organisms has elucidated many critically important conserved processes that include the cell division cycle, apoptosis, and development. We have recently recognized the importance of cilia in diverse human diseases. To understand and to treat cilia-based diseases, we must understand the assembly and the function of flagella/cilia and basal bodies (Pazour and Rosenbaum, 2002). Chlamydomonas and C. elegans have been opportune model organisms to study flagella/cilia and centrioles/basal bodies. In this work, we have exploited the observation that Chlamydomonas and many metazoans have flagella/cilia and basal bodies while Arabidopsis lacks these structures to identify by comparative genomics a set of 688 proteins that is enriched for components of the flagella and basal body. Several methods have demonstrated that the dataset is highly enriched in components of the flagella and centriole/basal body. Our FABB proteome contains 88% of established Chlamydomonas flagellar and basal body genes. Of the eight known genes not in the FABB proteome, five are absent from the human proteome. One was present (IFT20) but, perhaps due to its small size, its E value was not significant. Two of the missing proteins (Oda11 and Pf16) have significant matches to human homologs, but missed the imposed cutoff E value of 10^-10 (Supplemental Table S2). This reflects a potential caveat of our methodology; these are two proteins that could be included if the criteria were adjusted to reflect the difference rather than the absolute E values in the flagellated and nonflagellated comparisons. Pf16 is the sperm-associated antigen 8 (SPAG6) in mammals and is clearly a flagellar component. Including these two proteins would raise our identification rate to 91%.

Amputation of Chlamydomonas flagella induces transcription of flagellar genes (Lefebvre et al., 1980). Thirty-eight percent of the 103 genes tested show at least a 3-fold increase in mRNA levels following deflagellation and about two-thirds of these genes encode unknown proteins. We suggest that these genes are components of the flagellar axoneme, membrane, or IFT particles. Among the upregulated genes that were characterized in mammals are hydin, NYD-SP28, which is involved in testis development, a leucine-rich testis-specific protein, a testis-specific AMY-1 binding protein, and KPL2, which is induced during ciliogenesis in the trachea.

Nineteen of the genes tested showed an increase in mRNA levels between 1.5- and 3-fold. These include BBS1, 2, 4, 5, 7, and 8 as well as tubulin tyrosine ligase and a testis-expressed gene. This intermediate induction level would be consistent with these genes encoding components needed in the transition zone of Chlamydomonas. The distal portions of the transition zone are lost upon deflagellation, and modest increases in transcription may be necessary for its reassembly (Dutcher, 2004). This interpretation agrees well with the localization of BBS-5 in C. elegans and in BBS5 mouse (Figure 6). However, without immunoelectron microscopy, it is difficult to determine if the proteins are localized to the basal body or to the transition zone.

The remaining 45 tested genes do not show significant changes in mRNA levels following deflagellation. These may represent basal body genes. Alternatively, they may be noise in our dataset. RNAi treatment of five of the six non-upregulated strains show flagellar defects; several show cleavage furrow defects. The severity of the phenotype correlates with the reduction in mRNA observed after transformation with RNAi constructs. These results suggest that this class is enriched for basal body genes.

One potential drawback to our dataset is the low abundance of coiled-coil proteins. These proteins have low information content since the structure may be more important than the exact sequence of the protein. Other proteins possibly missing from the FABB proteome include proteins that have functions in processes other than the flagella and basal bodies. Although the newer members of the tubulin superfamily, α-, β-, and γ-tubulin clearly play a role in basal body function and assembly (Dutcher, 2003b), they are subtracted from our dataset by α-, β-, and γ-tubulin just as are kinesin- and calmodulin-like proteins. We may miss other proteins that are members of families with diverse roles. Finally, we will miss any proteins whose genes are evolving rapidly, which may include BBS6 or cystin.

Taken together, our comparative genomics approach has produced a robust protein dataset that represents a significant portion of the Chlamydomonas proteome necessary for flagellar and basal body biogenesis and function. The core observation was the strong conservation of flagellar and basal body proteins. It is possible that a similar approach could be applied to other organisms and processes conserved in some phyla, but lost in others, to compartmentalize proteomes into functional subsets. The success of such approaches will depend on the extent of functional conservation of the process under study, the fidelity of the subtraction, particularly with respect to the choice of target genomes, and the ability to validate the subtracted protein fractions.

Notably, nearly half the proteins in our FABB set have no predicted function, and over a quarter of those lack recognizable motifs. This highlights the ability of proteome subtractions to assign initial function at a genome-wide level. Given the disparities between the rate of genomic data production and functional characterization of gene products, these types of approaches might provide the means to expedite annotation and accelerate the acquisition of biological tools. Finally, our FABB proteome will potentially serve as a useful template for the identification of genes involved in human ciliation disorders. The BBS critical region spanned a 14 Mb interval containing in excess of 200 genes; the FABB proteome enabled us to clone BBS5 by interrogating only two transcripts. Given that the involvement of cilia in a range of disorders is becoming more evident (Pazour and Rosenbaum, 2002), we predict that several genes involved in retinal degeneration, nephropathies, and pleiotropic syndromes with phenotypic overlap to BBS...
will be included in the FABB proteome. Our collection will not only facilitate the genetic characterization of human disease phenotypes, but also provide initial information into the cellular basis of dysfunction.

Experimental Procedures

Searching the Proteomes

We used WU-BLASTP (Lopez et al., 2003) with E ≤ 10^-10 to align all of the proteins predicted by GreenGenie (Li et al., 2003) in *Chlamydomonas* to all predicted proteins in several organisms with and without cilia.

*Chlamydomonas* Growth Conditions and Strains

Strain CC-124 was used in all transformation experiments. For RNA isolation and transformation, cells were grown in Sager and Granick II medium with stirring at 25°C to a density of 8 × 10^6 cells/ml. Transformation was performed as described previously (Perrone et al., 2003).

Quantitative Real-Time RT-PCR

We employed real-time RT-PCR to quantify the change in gene expression level after *Chlamydomonas* cells were deflagellated by pH shock (Piperno et al., 1977). Seventy-six genes had an EST to confirm the exon-intron junction; 27 genes had a well-conserved human homolog. Real-time RT-PCR was performed in 40 cycles of 95°C for 15 s and 60°C for 1 min using the ABI PRISM 7000 Sequence Detection System. The set of primers for each gene preferentially span an intron and sizes range from 50 to 150 bp. Oligonucleotides were obtained from Integrated DNA Technologies, Inc. (Corvalville, Iowa). The SYBR Green PCR master mix (Applied Biosystems, Foster City, California) and uracil DNA glycosidase (Invitrogen, Carlsbad, California) were used according to the manufacturer’s guidelines.

RNA Interference

Short hairpin constructs were constructed from a gene fragment containing exon and intron sequence and a fragment containing the same exonic sequence with added restriction enzyme sites. PCR fragments from genomic DNA were cloned into the TOPO TA-vector (Invitrogen) and confirmed by restriction enzyme digestion. The fragments were ligated together with T4 DNA ligase according to the manufacturer’s guidelines (New England Biolabs, Boston, Massachusetts). The construct was cloned into the plasmid STL2, which is a derivative of pHSGrGFP (Fuhrmann et al., 1999). The resulting plasmid was cotransfected with pHS103, a plasmid carrying the aphII gene that confers resistance to paromomycin (Sizova et al., 2001). The presence of RNAi plasmid in transformants was verified by colony PCR with primers to the sequence adjacent to the cloning site on pSTL2. The flagellar and cleavage furrow phenotypes of the transformants were examined by phase microscopy. Duplicate quantitative real-time RT-PCR was used to measure changes in the mRNA levels in transformants.

*C. elegans* Analysis

Growth and culture of all *C. elegans* strains was carried out using standard techniques (Brenner, 1974). RNA was isolated from mixed stage daf-19(+) and daf-19(m86) strains by standard protocols (Haya-craft et al., 2001). Both daf-19(+) and daf-19(m86) strains also carry the daf-12(sa204) allele to suppress the Daf-c phenotype of daf-19(m86), daf-12(sa204), and daf-19(m86) adult hermaphrodites were transformed using standard techniques (Mello et al., 1991). Four independent transgenic lines were generated and analyzed. To obtain daf-19(-) worms for expression analysis, transgenic daf-12(sa204); daf-19(m86) males were mated to daf-12(sa204) hermaphrodites. Cross progeny were identified based on the presence of the extrachromosomal array and allowed to self-fertilize. The subsequent generation was screened for rescue of the daf-19(m86) dpy-filling defective phenotype by incubation in the hydrophobic dye DiO (Molecular Probes, Inc., Eugene, Oregon) as previously described (Fujiwara et al., 1999).

A promoter fusion was constructed with 500 bp upstream of the predicted translational start site of R01H10.5 through the second coding exon and was cloned into the GFP expression vector pPD95.75 (gift of A. Fire). The resulting construct was sequenced to ensure the correct translational reading frame and X box sequences. The entire bbs-5 gene, including 496 bp of upstream promoter sequence, was fused in-frame with the coding region of gfp and the unc-54 3′ UTR. Transgenic lines expressing the bbs-5::gfp transgene were generated as described previously (Ansley et al., 2003).

Bardet-Biedl Syndrome Patients and Analysis

Individuals with Bardet-Biedl syndrome were ascertained by RAL, PBL, JSG, or PSP. Diagnosis was based on previously established criteria. For all kindreds, interviews were conducted, and genealogical records were constructed. With informed consent, all available medical records were obtained and reviewed. Appropriate individuals were genotyped for published markers from each BBS locus, and haplotypes constructed as described (Ansley et al., 2003). Primer sequences and amplification conditions are available upon request for the exons sequenced.

Generation of Antibodies to Mouse BBS5 Protein

Antisera against mammalian BBS5 were generated by injection of a peptide (DSIEDTKGNNGDRGC) corresponding to amino acids 35 to 48 of the mouse homolog into New Zealand White rabbits according to the standard protocols of Sigma-Genosys (Woodlands, Texas). Immunoblot analysis of mouse embryos (embryonic day 10.5) or cultured renal inner medullary collecting duct (IMCD; Stanford, 1997) cells were performed as described with a 1:500 dilution (Yoder et al., 2002). The horseradish peroxidase signal was detected using SuperSignal West Dura Extended Duration Substrate (Pierce, Rockford, Illinois) according to the manufacturer’s instructions. For immunofluorescence, brain tissue was embedded in O.C.T. freezing compound (Sakura Finetek U.A., Inc., Torrence, California), cut into 4–6 μm sections, and fixed in 4% paraformaldehyde. Antibodies to acetylated α-tubulin (diluted 1:3000; Sigma, St. Louis, Missouri) or γ-tubulin (diluted 1:1000; Sigma) were used. Secondary antibodies were obtained from Jackson ImmunoResearch (West Grove, Pennsylvania) and were stained using Hoechst No. 33342 (Sigma). All images were captured under identical settings using a Nikon TE2000 eclipse equipped with a Roper Scientific CoolSnap HQ CCD camera and MetaMorph imaging software (Universal Imaging, Downingtown, Pennsylvania).

Acknowledgments

Diego Martinez and Dan Rokhsar of DOE/JGI made much of this work possible and facilitated obtaining the *Chlamydomonas* genomic sequence. We thank the BBS families for their continued participation and the Research Department of the King Khaled Eye Specialist Hospital, Riyadh, KSA, and particularly Dr. John Cavender, former Director of Research. We thank N. Morrisssette, J. Esperanza, J. Keller, and A. Lippa for useful comments on the comparative genomics as well as S. Ansley, J. Badano, J. Kloss, C. Li, and A. Mah for technical assistance. J.B.L. thanks Dr. Robert Thach for his enthusiastic encouragement. S.K.D. is supported by NIH (GM-32843). J.B.L. was supported by the Monsanto Fellowship at Washington University. N.K. is supported by NIH (HD04250) and the March of Dimes. J.M.G. is supported by a fellowship from the German Academic Exchange Service. G.D.S. is supported by NIH (GM28755). B.K.Y. is supported by NIH (DK065655 and DK-62758). L.M.G.-W. is supported by NIH (DK055534). M.R.L. is supported by the HSFBC&Y and holds scholar- ship awards from the CIHR and MSFHR.

Received: March 26, 2004
Revised: April 22, 2004
Accepted: April 23, 2004
Published: May 13, 2004

References


Ansley, S.J., Badano, J.L., Blacque, O.E., Hill, J., Hoskins, B.E.,


