Arnaiz et al. Cilia 2014, 3:9
http://www.ciliajournal.com/content/3/1/9

Cilia

METHODOLOGY Open Access

Remodeling Cildb, a popular database for cilia and links for ciliopathies

Olivier Arnaiz, Jean Cohen, Anne-Marie Tassin and France Koll*

Abstract

Background: New generation technologies in cell and molecular biology generate large amounts of data hard to exploit for individual proteins. This is particularly true for ciliary and centrosomal research. Cildb is a multi–species knowledgebase gathering high throughput studies, which allows advanced searches to identify proteins involved in centrosome, basal body or cilia biogenesis, composition and function. Combined to localization of genetic diseases on human chromosomes given by OMIM links, candidate ciliopathy proteins can be compiled through Cildb searches.

Methods: Othology between recent versions of the whole proteomes was computed using Inparanoid and ciliary high throughput studies were remapped on these recent versions.

Results: Due to constant evolution of the ciliary and centrosomal field, Cildb has been recently upgraded twice, with new species whole proteomes and new ciliary studies, and the latter version displays a novel BioMart interface, much more intuitive than the previous ones.

Conclusions: This already popular database is designed now for easier use and is up to date in regard to high throughput ciliary studies.

Background

Whatever the field studied in biology, due to the prevalence of new generation technologies, retrieving relevant information from high throughput studies represents a most important challenge. In this view, five years ago, we developed Cildb, a knowledgebase that allowed data mining concerning cilia and ciliopathies (http://cildb.cgm.cnrs-gif.fr/) [1]. Cildb progressively became a reference cilium database, with a number of users reaching now 700 per month. Since its creation and publication [1], Cildb underwent several modifications and improvements, yielding an evolution to Version 2.1 in 2010 and now to Version 3.0 in 2014. Although data in Cildb are raw data treated automatically, so that false positives and false negatives may occur, results are fully informative and make easier searches on ciliary genes.

The purpose of this note is fourfold, reminding the reader of the main uses of this database already described in more detail by Arnaiz et al. [1], providing explanation of the updates, describing the new interface and evaluating the orthology relationships as calculated in Cildb.

* Correspondence: france.koll@cgm.cnrs-gif.fr
Centre de Génétique Moléculaire, CNRS, Avenue de la Terrasse, 91198 Gif sur Yvette, France

Cildb, a database for ciliary studies… and more

In the early 2000's, high throughput studies started to appear concerning cilia, a re-emerging organelle at that time [2], and centrioles [3], precursors of basal bodies of cilia in metazoans. Such studies generated large amounts of data on cilia, basal body, centriole, and centrosome proteomes, on transcriptome analyses realized under various conditions (ciliogenesis etc.), and on computation issued from comparative genomics between centric (i.e. with cilia/flagella or at least centrioles at some stage of their life cycle) and acentric organisms. Developing a way to browse these data became essential, not only from the statistician's point of view, but also for experimental biologists who want to seek information on individual proteins from the bulk of the results.

Methods

The originality of Cildb was in its backbone that related on the one side a network of orthology between the whole proteomes, complete sets of protein sequences, of all the species taken pair-wise, calculated with the algorithm of Inparanoid version 4.1 with default parameters [4], and on the other side the detection of each protein in a set of ciliary studies [1]. Therefore, the database allows searches for possible ciliary properties on the whole

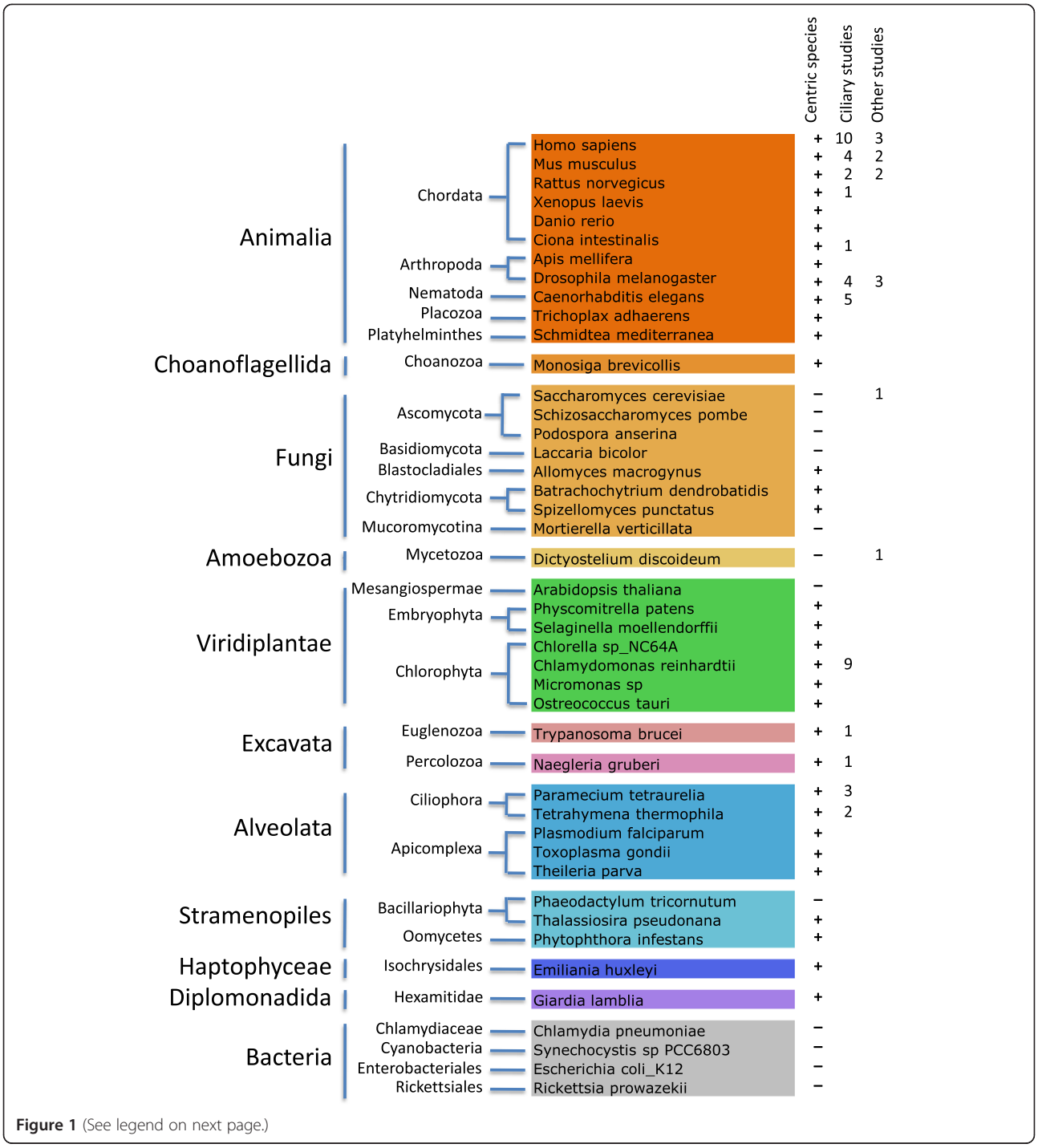

Figure 1 (See legend on next page.)

(See figure on previous page.)
Figure 1 The species whose whole proteome has been included into Cildb V3.0 are gathered by taxonomy groups, with indication whether they are centric or not and of the number of high throughput studies, ciliary or not, performed in the species. The choice of species to include into Cildb was 1) species in which high throughput ciliary studies have been performed, 2) species routinely used as models in ciliary studies in general, and 3) centric and acentric species, because the presence/absence of certain proteins may be relevant for the conservation of ciliary proteins through evolution. The case of the Bug22/GTL3/C16orf80 protein, composed of a domain called DUF667, essential for ciliary motility [6], was carefully examined for the choice of fungi to add in Cildb for comparative genomics. Bug22 is a protein highly conserved in all centric species, be they metazoans, protozoa, plants or fungi and curiously also highly conserved in the acentric land plants, but absent from the genomes of higher fungi already sequenced at the time of the publication, i.e. acentric ascomycetes [6]. Owing to constant new genome sequencing, novel fungal whole proteomes appeared and the occurrence of Bug22 was different from what was thought earlier. It is still undetectable in ascomycetes, but is found conserved in the acentric *Mortierella verticillata* (accession MVEG_01915), and a more divergent Bug22 with recognizable DUF667 domain is found in several basidiomycetes represented in Cildb by *Laccaria bicolor* (accession 598201). This property was one of the reasons to include those two fungi proteomes into Cildb V3.0. This also emphasizes that constant arrival of new knowledge as new genomes are sequenced can put into questions former assumptions such as the absence of particular proteins in some species, here Bug22 in fungi.

proteome of one species, e.g. *Homo sapiens*, based on ciliary properties established by studies conducted in another species, e.g. flagellum proteomics in *Chlamydomonas* [5]. In addition, the whole human proteome has been linked to the OMIM database (http://www.ncbi.nlm.nih.gov/omim/) that gathers all known human genetic disorders with the corresponding genes. This allows searches of proteins involved in diseases and to display the OMIM description as attribute in the output of a search. Conversely, searches in the whole proteome of any non-human species can tell if the resultant proteins are orthologous to human proteins linked to human diseases.

In addition to the ciliary properties of proteins, Cildb contains other information such as synonyms, descriptions, molecular weight, isoelectric point, probability of presence of a signal peptide, of transmembrane helices, as well as the FASTA sequence. This extra information can be searched for and displayed as properties using Cildb.

Cildb has been imagined and worked out to manipulate outputs of high throughput studies. All data coming from studies dedicated to the function of only a specific or of several proteins are not included in Cildb so that some ciliary proteins may escape from Cildb searches if they are not revealed by high throughput studies.

Results and discussion

What is new in Cildb V3.0?

Since the last version of Cildb, new high throughput ciliary studies have appeared and more model organisms have been used for ciliary studies. Thus, we remodeled Cildb to include the proteomes of altogether 44 species, among which are 41 eukaryotes and 3 bacteria (http://cildb.cgm.cnrs-gif.fr/v3/cgi/genome_versions; Figure 1) and 66 studies, among which 55 directly concern cilia, and 11 other, related studies (http://cildb.cgm.cnrs-gif.fr/v3/cgi/ciliary_studies; Table 1). BLAST server and human GBrowse facilities are maintained in the new version. In addition, a Motif Search tool has been implemented in order to search proteomes with a sequence motif using the patmatdb program from the EMBOSS package (http://bioweb2.pasteur.fr/docs/EMBOSS/patmatdb.html), based on the format of pattern used in the PROSITE database (http://prosite.expasy.org/prosuser.html). For example, an amino acid motif such as MKK[KP]K, in which either K or P can stand at the fourth position, can be queried in the proteome of any species of Cildb.

Species implemented in Cildb V3.0

Cildb V3.0 contains now whole proteomes of 41 eukaryotes among which 32 are centric species. Fifteen of these species were used for the 66 high throughput studies of Cildb. The 17 other species are good models for ciliary experiments although no high throughput study has been published as of yet. Nine eukaryotic acentric species which lack cilia and centrioles were also taken because they represent 'negative controls' in comparative genomics experiments: two species for which two analyses on spindle pole proteomes are available and seven species without high throughput relevant studies.

Since orthology relationships are a major tool in Cildb, we corrected an inconsistency in the proteome composition in various species. Indeed, species present in Cildb are not homogeneous in their whole proteome, some of them including organelle proteomes (mitochondria, chloroplasts), others not. Organelle proteomes represent a minor part of all the proteins, but since some organellar proteins can be encoded either by nuclear genes or by the organelle, according to the species, this may influence the orthology calculation in some cases. This issue has been fixed in Cildb V3.0. In addition, to study the origin of organellar proteins, we added the whole proteomes of three bacteria because they are closest to those of mitochondria (*Rickettsia prowazekii*) and chloroplasts (*Synechocystis sp PCC6803*, *Chlamydia pneumoniae*).

Since the original publication of Cildb [1], the whole proteomes of 26 novel eukaryotic species have been introduced into Cildb. A notable proportion of fungi, eight fungal whole proteomes, are incorporated in Cildb mainly because fungi represent a phylum at a hinge position in the evolution of centric and acentric species.

Table 1 High throughput studies compiled in Cildb V3.0

Reference for the study	Method	Species	Ciliary analysis
Andersen et al., 2003 [3]	Centriole proteome	*Homo sapiens*	yes
Arnaiz et al., 2009 [1]	Cilium proteome	*Paramecium tetraurelia*	yes
Arnaiz et al., 2010 [7]	Expression during ciliogenesis	*Paramecium tetraurelia*	yes
Avidor-Reiss et al., 2004 [8]	Comparative genomics	*Drosophila melanogaster*	yes
Baker et al., 2008a [9]	Spermatozoa proteome	*Mus musculus*	no
Baker et al., 2008b [10]	Spermatozoa proteome	*Rattus norvegicus*	no
Bechstedt et al., 2010 [11]	Expression in tissues containing sensory cilia	*Drosophila melanogaster*	yes
Blacque et al., 2005 [12]	Differential expression between ciliated and non ciliated cells	*Caenorhabditis elegans*	yes
Blacque et al., 2005 [12]	Genomic screening for X-boxes in promoters	*Caenorhabditis elegans*	yes
Boesger et al., 2009 [13]	Flagellum phosphoproteome	*Chlamydomonas reinhardtii*	yes
Broadhead et al., 2006 [14]	Flagellum proteome	*Trypanosoma brucei*	yes
Cachero et al., 2011 [15]	Expression in early development of future neural cells	*Drosophila melanogaster*	no
Cao et al., 2006 [16]	Sperm flagellar axonemes proteome	*Mus musculus*	yes
Chen et al., 2006 [17]	Expression in daf-19 mutant	*Caenorhabditis elegans*	yes
Datta et al., 2011 [18]	Gene expression with HIPPI expression modulation	*Homo sapiens*	no
Dorus et al., 2006 [19]	Spermatozoa proteome	*Drosophila melanogaster*	no
Efimenko et al., 2005 [20]	Genomic screening for X-boxes in promoters	*Caenorhabditis elegans*	yes
Fritz-Laylin and Cande, 2010 [21]	Flagellum proteome	*Naegleria gruberi*	yes
Geremek et al., 2011 [22]	Expression in primary ciliary dyskinesia patients	*Homo sapiens*	yes
Geremek et al., 2014 [23]	Expression in primary ciliary dyskinesia patients	*Homo sapiens*	yes
Guo et al., 2010 [24]	Proteomics associated with spermiogenesis	*Mus musculus*	no
Hodges et al., 2011 [25]	Comparative genomics	*Chlamydomonas reinhardtii*	yes
Hoh et al., 2012 [26]	Expression in multiciliated cells from trachea	*Mus musculus*	yes
Huang et al., 2008 [27]	Proteomics associated with spermiogenesis	*Mus musculus*	no
Hughes et al., 2008 [28]	Proteome of Microtubule-Associated Proteins	*Drosophila melanogaster*	no
Ishikawa et al., 2012 [29]	Primary cilium proteome	*Mus musculus*	yes
Ivliev et al., 2012 [30]	Expression profile in different tissues	*Homo sapiens*	yes
Jakobsen et al., 2011 [31]	Centrosome proteomics	*Homo sapiens*	yes
Keller et al., 2005 [32]	Expression during ciliogenesis	*Chlamydomonas reinhardtii*	yes
Keller et al., 2005 [32]	Basal body proteome	*Chlamydomonas reinhardtii*	yes
Kilburn et al., 2007 [33]	Basal body proteome	*Tetrahymena thermophila*	yes
Kim et al., 2010 [34]	Ciliogenesis modulation	*Homo sapiens*	yes
Kubo et al., 2008 [35]	Expression in ciliated tissues	*Homo sapiens*	yes
Laurençon et al., 2007 [36]	Genomic screening for X-boxes in promoters	*Drosophila melanogaster*	yes
Lauwaet et al., 2011 [37]	Homology search for basal body proteins	*Giardia lamblia*	yes
Lauwaet et al., 2011 [37]	Basal body proteome	*Giardia lamblia*	yes
Li et al., 2004 [38]	Comparative genomics	*Chlamydomonas reinhardtii*	yes
Liu et al., 2007 [39]	Cilium proteome	*Mus musculus*	yes
Martínez-Heredia et al., 2006 [40]	Spermatozoa proteome	*Homo sapiens*	no
Mayer et al., 2008 [41]	Cilium proteome	*Rattus norvegicus*	yes
Mayer et al., 2009 [42]	Cilium proteome	*Rattus norvegicus*	yes
McClintock et al., 2008 [43]	Expression in ciliated tissues	*Mus musculus*	yes
Merchant et al., 2007 [44]	Comparative genomics	*Chlamydomonas reinhardtii*	yes
Müller et al., 2010 [45]	Centrosome proteome	*Drosophila melanogaster*	yes

Table 1 High throughput studies compiled in Cildb V3.0 *(Continued)*

Nakachi et al., 2011 [46]	Sperm tail proteome	*Ciona intestinalis*	yes
Nogales-Cadenas et al., 2009 [47]	Centrosome human curation	*Homo sapiens*	yes
Ostrowski et al., 2002 [2]	Cilium proteome	*Homo sapiens*	yes
Pazour et al., 2005 [5]	Expression during ciliogenesis	*Chlamydomonas reinhardtii*	yes
Pazour et al., 2005 [5]	Flagellum proteome	*Chlamydomonas reinhardtii*	yes
Phirke et al., 2011 [48]	Down and upregulated genes in daf-19 mutant	*Caenorhabditis elegans*	yes
Reinders et al., 2006 [49]	Nuclear-associated body proteome	*Dictyostelium discoideum*	no
Ross et al., 2007 [50]	Expression during ciliogenesis	*Homo sapiens*	yes
Sakamoto et al., 2008 [51]	Proteome of Microtubule-Associated Proteins	*Rattus norvegicus*	no
Sauer et al., 2005 [52]	Mitotic spindle proteome	*Homo sapiens*	no
Smith et al., 2005 [53]	Cilium proteome	*Tetrahymena thermophila*	yes
Stolc et al., 2005 [54]	Expression during ciliogenesis	*Chlamydomonas reinhardtii*	yes
Stubbs et al., 2008 [55]	Expression Under FoxJ1 silencing	*Xenopus laevis*	yes
Wigge et al., 1998 [56]	Spindle pole body proteome	*Saccharomyces cerevisiae*	no
Yano et al., 2013 [57]	Ciliary membrane proteome	*Paramecium tetraurelia*	yes

The high throughput studies present in Cildb V3.0 are summarized in the table with indication in the second column whether it is a proteomic, gene expression, or genomic study. The species in which the studies have been performed are specified in the third column. In the fourth column is the fact whether a given study is ciliary (concerns cilia, flagella, basal bodies, centrioles, centrosomes or spindle pole bodies) or not. The table is ordered alphabetically by first author of publication of the studies present in Cildb V3.0.

Studies in Cildb V3.0

The 66 studies incorporated in Cildb V3.0 mainly consist in high throughput proteomics, differential expression, and comparative genomics studies. 53 of these studies approach ciliary and centriolar/basal body components, structure, function or biogenesis. We also integrated 13 studies concerning related topics, such as microtubule-associated proteins, spindle proteins, spindle pole bodies, nuclear-associated bodies, whole sperm proteome, and others. Compared to Cildb V1.0, 45 novel studies have been introduced in Cildb.

High throughput studies concerning cilia appear monthly in the literature, but computation in Cildb needs full recalculation of the database, so that it cannot be updated each time. However, if the output of a study not present in Cildb has to be compared to a study already present, this can be performed using the keyword box in the general properties filter by querying a list of gene or protein IDs bordered by '%', one per line. The limitation is that the query is slow, since this is not the main task designed for BioMart queries.

Simplified interface and structure for Cildb V3.0

For users trained with previous versions of Cildb, the most prominent change is the new interface. Indeed, it takes advantage of the novel environment provided by BioMart Version 9 [58] (Figure 2). In consequence, making an advanced search becomes much more intuitive than earlier, even for non-trained users, who can easily enter the functionalities of the database.

The simplification of the interface is accompanied by a simplification of the structure of the database. First of all, the orthology calculation has been exclusively centered on Inparanoid [4]. Formerly, users could choose between Inparanoid and Inparanoid plus 'in house' filtered blast hits. The most recent version of Inparanoid appears efficient enough to prevent the output of too many false negatives that occurred with the previous versions, so that the addition of 'in house' filtered blast hits was no more necessary, as detailed in the next section and in the legend of Table 2. We also simplified the way to filter ciliary studies and removed less useful other searches (operator 'OR', customized searches). However, the functions removed in the query menu compared to previous Cildb versions can be applied by another process that consists of downloading data as tables with relevant attributes and sorting these tables thereafter using a spreadsheet software.

The changes brought to Cildb may have unexpected impact and we would be grateful for any feedback by the users. In addition, since genome annotations evolve with time, proteins can be gained or lost in the deduced proteomes from a time to the next. For all these reasons, we kept the former "data freeze" versions of Cildb available through the "Version" menu for comparisons when it is necessary.

Evolutionary conservation viewed through Cildb, the example of centrosomal proteins

To evaluate the identification of orthologs by Inparanoid, called 'inparalogs', we studied centrosomal proteins

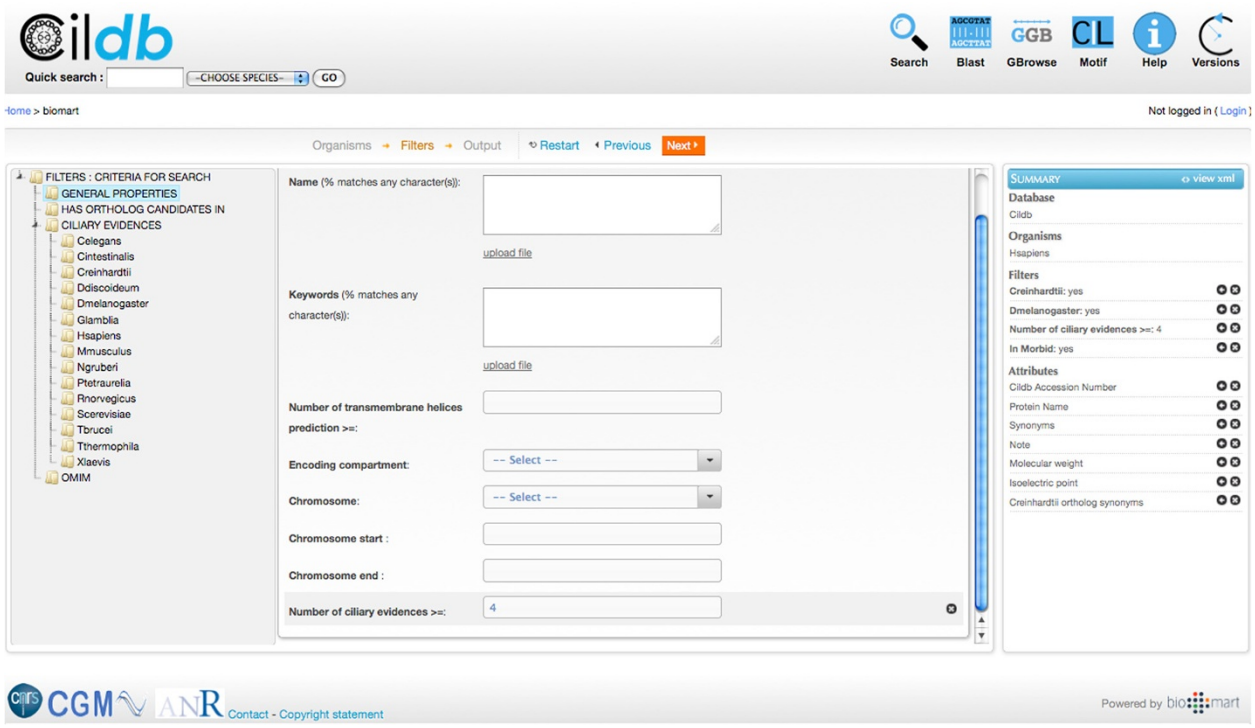

Figure 2 An advanced search on Cildb V3.0 is started by clicking on the 'Search' button on the top row on the right. Then, it is necessary to choose the species in which the proteome has to be searched for. The filter window then appears to adjust the filters in the left panel (no filter means that the full proteome will be retrieved). Similarly, the output window allows displaying particular properties (attributes) in columns for each filtered protein. A summary on the right reminds the user of all the filters and attributes currently used. This also allows direct modification of the orders of the columns in the output by moving the attributes up and down in the list. The last operation of the process is to show the results. The results are given by pages of 20 items with a maximum of 1000 items. To see all results, they have to be downloaded as a file. At any time, if the result output seems incomplete or inappropriate, the filters and attributes can be modified by using the 'Back' button (edit results) to refine the search and show the results again. The quick search allows a rapid search by keywords. The result can be processed the same way as the one described above, with the possibility to add attributes by 'Edit results' and to download the file. Note the direct access to BLAST, Human genome Gbrowse, Motif search, Help and access to older Versions of Cildb on the top row buttons to the right.

in more detail, since they are conserved proteins already pretty well known. We wondered whether centrosomal proteins identified in three studies in *Homo sapiens* would reveal the orthologs, when they exist, in other species. We used the following protocol:

- click the 'Search' button on the bar on the to right
- select 'Hsapiens' as organism in the scroll-down menu
- click 'Next' and open 'Ciliary Evidences' on the left menu
- click 'Hsapiens' and select 'yes' for the centrosomal studies [3,31] and [47]
- click 'Next' and display ortholog names, synonyms, etc. for any desired species listed in the left menu. You can select here as an output the stringency for the studies chosen in the queries, if you want to sort the output table thereafter.
- click 'Results' to visualize the output
- modification of the filters and output can be obtained by the back button 'Edit Results'
- when satisfied with the result, click 'Download data'

We chose to emphasize the orthologs in *Mus musculus*, *Rattus norvegicus*, *Danio rerio*, *Apis mellifera* and *Drosophila melanogaster* in the output to follow the evolutionary conservation, as viewed with Inparanoid. Among the 113 human proteins encoded by 77 genes found as centrosomal by this filter, inparalogs were detected for 76 genes in mouse, 75 in rat, 68 genes in fish, 37 genes in bee and 33 genes in fly (Table 2). A vast majority of these proteins were identified in mammals, as well as in fish, a vertebrate. More negative examples were found in the insects bee and fly. To check whether homologues were indeed absent when no Inparalogs were found, we performed BLAST searches on individual species proteomes using the Cildb BLAST. Except for the two cases discussed in the legend of Table 2, all the absence of Inparalogs corresponds to no or weak BLAST hit detection. In addition, none of the BLAST targets were found in the previous version of Cildb as filtered best hits, a calculation method that we suppress in the present version. Altogether, although reciprocal BLAST searches are always useful to

Table 2 Evolutionary conservation of centrosomal proteins viewed through Cildb V3.0

	Protein ID	Synonyms	Mus	Rattus	Danio	Apis	Drosophila	Class
1	ENSP00000380378	PAFAH1B1,LIS1,LIS2,MDCR	yes	yes	yes	yes	yes	1 (yyyyy)
2	ENSP00000364691	CROCC,ROLT,ROLT,rootletin	yes	yes	yes	yes	yes	1 (yyyyy)
3	ENSP00000309591	PRKACA,PKACA,PKACa	yes	yes	yes	yes	yes	1 (yyyyy)
4	ENSP00000263710	CLASP1,MAST1	yes	yes	yes	yes	yes	1 (yyyyy)
5	ENSP00000263811	DYNC1I2,DNCI2,IC2	yes	yes	yes	yes	yes	1 (yyyyy)
6	ENSP00000216911	AURKA,AIK,ARK1,AURA,AURORA2	yes	yes	yes	yes	yes	1 (yyyyy)
7	ENSP00000364721	MAPRE1,EB1,EB1	yes	yes	yes	yes	yes	1 (yyyyy)
8	ENSP00000265563	PRKAR2A,PKR2,PRKAR2	yes	yes	yes	yes	yes	1 (yyyyy)
9	ENSP00000355966	NEK2,HsPK21,NEK2A,NLK1	yes	yes	yes	yes	yes	1 (yyyyy)
10	ENSP00000261965	TUBGCP3,GCP3,SPBC98	yes	yes	yes	yes	yes	1 (yyyyy)
11	ENSP00000252936	TUBGCP2,GCP2,Grip103,h103p	yes	yes	yes	yes	yes	1 (yyyyy)
12	ENSP00000251413	TUBG1,CDCBM4,GCP-1	yes	yes	yes	yes	yes	1 (yyyyy)
13	ENSP00000456648	TUBGCP4,76P,GCP-4,GCP4	yes	yes	yes	yes	yes	1 (yyyyy)
14	ENSP00000323302	POC1B,PIX1,TUWD12,WDR51B	yes	yes	yes	yes	yes	1 (yyyyy)
15	ENSP00000324464	CSNK1D,ASPS,CKIdelta,FASPS2,HCKID	yes	yes	yes	yes	yes	1 (yyyyy)
16	ENSP00000270861	PLK4,SAK,STK18,Sak	yes	yes	yes	yes	yes	1 (yyyyy)
17	ENSP00000356785	NME7,MN23H7,NDK7	yes	yes	yes	yes	yes	1 (yyyyy)
18	ENSP00000273130	DYNC1LI1,DNCLI1,LIC1	yes	yes	yes	yes	yes	1 (yyyyy)
19	ENSP00000359300	CETN2,CALT,CEN2	yes	yes	yes	yes	yes	1 (yyyyy)
20	ENSP00000287380	TBC1D31,Gm85,WDR67	yes	yes	yes	yes	yes	1 (yyyyy)
21	ENSP00000287482	SASS6,SAS-6,SAS6	yes	yes	yes	yes	yes	1 (yyyyy)
22	ENSP00000300093	PLK1,PLK,STPK13	yes	yes	yes	yes	yes	1 (yyyyy)
23	ENSP00000257287	CEP135,CEP4,MCPH8	yes	yes	yes	yes	yes	1 (yyyyy)
24	ENSP00000439376	DCTN2,DCTN50,DYNAMITIN,RBP50	yes	yes	yes	yes	yes	1 (yyyyy)
25	ENSP00000395302	CKAP5,ch-TOG,CHTOG,MSPS	yes	yes	yes	yes	yes	1 (yyyyy)
26	ENSP00000342510	CEP97,LRRIQ2	yes	yes	yes	yes	yes	1 (yyyyy)
27	ENSP00000348965	DYNC1H1,DHC1,DHC1a	yes	yes	yes	yes	yes	1 (yyyyy)
28	ENSP00000469720	CETN2,CALT,CEN2	yes	yes	yes	yes	yes	1 (yyyyy)
29	ENSP00000317156	CEP192,PPP1R62	yes	yes	yes	yes	no	2 (yyyyn)
30	ENSP00000270708	WRAP73,WDR8	yes	yes	yes	yes	no	2 (yyyyn)
31	ENSP00000248846	TUBGCP6,GCP-6,GCP6,MCCRP,MCPHCR	yes	yes	yes	yes	no	2 (yyyyn)
32	ENSP00000393583	AZI1,AZ1,Cep131,ZA1	yes	yes	yes	yes	no	2 (yyyyn)
33	ENSP00000283645	TUBGCP5,GCP5	yes	yes	yes	yes	**no**	2 (yyyyn)
34	ENSP00000303058	CEP120,CCDC100	yes	yes	yes	yes	no	2 (yyyyn)
35	ENSP00000313752	SSNA1,N14,NA-14	yes	yes	yes	yes	no	2 (yyyyn)
36	ENSP00000355812	FGFR1OP,FOP	yes	yes	yes	yes	no	2 (yyyyn)
37	ENSP00000343818	CDK5RAP2,C48,Cep215,MCPH3	yes	yes	yes	**no**	yes	3 (yyyny)
38	ENSP00000344314	OFD1,CXorf5,JBTS10,RP23	yes	yes	yes	no	no	4 (yyynn)
39	ENSP00000317144	PIBF1,C13orf24,CEP90	yes	yes	yes	no	no	4 (yyynn)
40	ENSP00000204726	GOLGA3,GCP170,MEA-2,golgin-160	yes	yes	yes	no	no	4 (yyynn)
41	ENSP00000206474	HAUS4,C14orf94	yes	yes	yes	no	no	4 (yyynn)
42	ENSP00000281129	CEP128,C14orf145,C14orf61,LEDP/132	yes	yes	yes	no	no	4 (yyynn)
43	ENSP00000262127	CEP76,C18orf9,HsT1705	yes	yes	yes	no	no	4 (yyynn)
44	ENSP00000370803	CCP110,Cep110,CP110	yes	yes	yes	no	no	4 (yyynn)
45	ENSP00000263284	CCDC61	yes	yes	yes	no	no	4 (yyynn)
46	ENSP00000223208	CEP41,JBTS15,TSGA14	yes	yes	yes	no	no	4 (yyynn)

Table 2 Evolutionary conservation of centrosomal proteins viewed through Cildb V3.0 *(Continued)*

47	ENSP00000303769	AKNA	yes	yes	yes	no	no	4 (yyynn)
48	ENSP00000302537	MDM1	yes	yes	yes	no	no	4 (yyynn)
49	ENSP00000264935	CEP72,FLJ10565	yes	yes	yes	no	no	4 (yyynn)
50	ENSP00000419231	CEP70,BITE	yes	yes	yes	no	no	4 (yyynn)
51	ENSP00000306105	CEP89,CCDC123	yes	yes	yes	no	no	4 (yyynn)
52	ENSP00000380661	CEP250,C-NAP1,CEP2,CNAP1	yes	yes	yes	no	no	4 (yyynn)
53	ENSP00000356579	CEP350,CAP350,GM133	yes	yes	yes	no	no	4 (yyynn)
54	ENSP00000260372	HAUS2,C15orf25,CEP27,HsT17025	yes	yes	yes	no	no	4 (yyynn)
55	ENSP00000360540	CEP55,C10orf3,CT111,URCC6	yes	yes	yes	no	no	4 (yyynn)
56	ENSP00000355500	CEP170,FAM68A,KAB	yes	yes	yes	no	no	4 (yyynn)
57	ENSP00000369871	HAUS6,Dgt6,FAM29A	yes	yes	yes	no	no	4 (yyynn)
58	ENSP00000371308	CENPJ,BM032,CENP-J,CPAP,LAP,LIP1,MCPH6,Sas-4	yes	yes	yes	no	no	4 (yyynn)
59	ENSP00000282058	HAUS1,CCDC5,HEI-C,HEIC	yes	yes	yes	no	no	4 (yyynn)
60	ENSP00000283122	CETN3,CDC31,CEN3	yes	yes	yes	no	no	4 (yyynn)
61	ENSP00000352572	PCNT,KEN,MOPD2,PCN,PCNT2,PCNTB	yes	yes	yes	no	no	4 (yyynn)
62	ENSP00000295872	SPICE1,CCDC52,SPICE	yes	yes	yes	no	no	4 (yyynn)
63	ENSP00000317902	CEP57,MVA2,PIG8,TSP57	yes	yes	yes	no	no	4 (yyynn)
64	ENSP00000426129	CEP63	yes	yes	yes	no	no	4 (yyynn)
65	ENSP00000308021	CEP290,BBS14,JBTS5,LCA10,MKS4,NPHP6,POC3,rd16,SLSN6	yes	yes	yes	no	no	4 (yyynn)
66	ENSP00000439056	HAUS5,dgt5	yes	yes	yes	no	no	4 (yyynn)
67	ENSP00000462740	CEP41,JBTS15,TSGA14	yes	yes	yes	no	no	4 (yyynn)
68	ENSP00000265717	PRKAR2B,PRKAR2,RII-BETA	yes	yes	no	yes	yes	5 (yynyy)
69	ENSP00000345892	NDE1,HOM-TES-87,LIS4,NDE,NUDE	yes	yes	no	yes	yes	5 (yynyy)
70	ENSP00000358921	ACTR1A,ARP1,CTRN1	yes	yes	no	yes	yes	5 (yynyy)
71	ENSP00000447907	DYNLL1,DLC1,DLC8,DNCL1,DNCLC1,hdlc1,LC8	yes	yes	no	no	no	6 (yynnn)
72	ENSP00000278935	CEP164,NPHP15	yes	yes	no	no	no	6 (yynnn)
73	ENSP00000264448	ALMS1,ALSS	yes	yes	no	no	no	6 (yynnn)
74	ENSP00000316681	KIAA1731	yes	yes	no	no	no	6 (yynnn)
75	ENSP00000456335	CNTROB,LIP8,PP1221	yes	yes	no	no	no	6 (yynnn)
76	ENSP00000348573	AKAP9,AKAP350,AKAP450,CG-NAP,HYPERION,LQT11	yes	no	no	no	no	7 (ynnnn)
77	ENSP00000384844	DCTN1,DAP-150,P135	**no**	**no**	yes	yes	yes	8 (nnyyy)

This table presents the list of 77 human proteins obtained from a BioMart search described in the text. The output gives a total of 133 proteins encoded by 77 genes, due to the presence of splice variants. For clarity, only one protein ID per gene has been presented in the table, after verification that all the splice variants of each gene displays the same orthology relationships with the species presented here. This table illustrates evolutionary conservation where a "yes" indicates that the human protein has an Inparalog in Cildb and a "no" that no Inparanoid orthology was found. The column 'class' serves to order the output genes in the table (from 5× 'yes' at the top to much fewer 'yes' at the bottom, along criteria of certain species being closer to each other than others, whereby the order from left to right goes human-mouse-rat (mammals), then fish (vertebrate), then bee and fly (insects). All instances of lacking orthology ("no") were individually verified by BLAST searches using the Cildb BLAST. The BLAST results were consistent with the absence of orthologs in the species, and only three exceptions contradict the Inparanoid results, highlighted as bold characters in the table.

1- Human Azi1 (ENSP00000393583) has no inparalog in *Drosophila* although an ortholog called *dilatory* exists. BLAST search on the *Drosophila* genome indeed light up *dilatory*, with a score very close to the one found for the *Apis* inparalogs by BLAST. The difference between these different outputs may result from the value of default thresholds taken by the Inparanoid program and the different lengths of the proteins.

2- Human cdk5rap2 (ENSP00000343818) has no Inparalog in *Apis*, although homologs are found by BLAST. Inparanoid relationships of the top three Apis proteins in the list (XP_006563202.1, XP_006563201.1, XP_392107.3) appear to be Inparalogs of *Drosophila* centrosomin (cnn, cdk5rap2) for which 8 of 12 splice variant proteins display human Inparalogs. However, no direct Inparanoid relationships exist between the Apis proteins and any human protein.

3- Human dynactin/dctn1 (ENSP00000384844) has surprisingly no Inparalogs in mouse and rat whereas some are found in fish, bee and fly. However, mouse and rat homologs are easily found by BLAST search. After careful examination, it appears that the only ENSP00000384844 dynactin protein found common to the three human centrosomal studies, is one of the splice variants excluded from Inparalog groups. Indeed, the 16 splice variants for the human dynactin gene *ENSG00000204843* and the seven splice variants for its mouse counterpart *ENSMUSG00000031865* are related by Inparanoid orthology through three groups, *hsap_mmus.17187* (one human and one mouse gene), *hsap_mmus.1073* (four human and one mouse gene) and *hsap_mmus.977* (one human and two mouse genes). The remaining ten human protein variants (among which is ENSP00000384844) and three mouse protein variants encoded by these genes are not included in the orthology groups, probably because their exon composition was too different from the other protein variants.

These three examples represent the limits of Inparanoid orthology prediction, highlighting the fact that reciprocal BLAST searches cannot be avoided, and thus represent an important complementary approach, for the analysis of individual proteins.

study the occurrence of individual proteins in various species, the orthology calculation via Inparanoid is pretty suitable for batch identification of conserved proteins using Cildb.

Conclusion

The version V3.0 of Cildb preserves its major original principles of relating orthology to ciliary studies, but, by improving its structure and its interface, makes the database more suitable for advanced searches. Altogether, Cildb V3.0 is a particularly useful tool for unraveling ciliary and ciliopathy networks and will hopefully help in identification of new orphan diseases.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

OA made bioinformatics calculations and developed, designed the database, JC and FK brought the biological knowledge on ciliary high throughput studies and species relevant to be included in the database, AMT validated the present version of the database concerning orthology of ciliary and centrosomal conserved proteins viewed by Inparanoid, JC, FK and AMT wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Funding from the Centre National de la Recherche Scientifique (CNRS) and the Foetocilpath grant from the Agence Nationale de la Recherche (ANR), are gratefully acknowledged. We are grateful to the INRA MIGALE bioinformatics platform (http://migale.jouy.inra.fr) for providing computational resources. This work was carried out in the context of the CNRS-supported European Research Group "Paramecium Genome Dynamics and Evolution".

Received: 29 September 2014 Accepted: 30 October 2014
Published: 17 Nov 2014

References

1. Arnaiz O, Malinowska A, Klotz C, Sperling L, Dadlez M, Koll F, Cohen J (2009) Cildb: a knowledgebase for centrosomes and cilia. Database (Oxford) 2009: bap022
2. Ostrowski LE, Blackburn K, Radde KM, Moyer MB, Schlatzer DM, Moseley A, Boucher RC (2002) A proteomic analysis of human cilia: identification of novel components. Mol Cell Proteomics 1:451–465
3. Andersen JS, Wilkinson CJ, Mayor T, Mortensen P, Nigg EA, Mann M (2003) Proteomic characterization of the human centrosome by protein correlation profiling. Nature 426:570–574
4. O'Brien KP, Remm M, Sonnhammer ELL (2005) Inparanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res 33(Database issue): D476–D480
5. Pazour GJ, Agrin N, Leszyk J, Witman GB (2005) Proteomic analysis of a eukaryotic cilium. J Cell Biol 170:103–113
6. Laligné C, Klotz C, De Loubresse NG, Lemullois M, Hori M, Laurent FX, Papon JF, Louis B, Cohen J, Koll F (2010) Bug22p, a conserved centrosomal/ciliary protein also present in higher plants, is required for an effective ciliary stroke in Paramecium. Eukaryotic Cell 9:645–655
7. Arnaiz O, Goût J-F, Bétermier M, Bouhouche K, Cohen J, Duret L, Kapusta A, Meyer E, Sperling L (2010) Gene expression in a paleopolyploid: a transcriptome resource for the ciliate Paramecium tetraurelia. BMC Genomics 11:547
8. Avidor-Reiss T, Maer AM, Koundakjian E, Polyanovsky A, Keil T, Subramaniam S, Zuker CS (2004) Decoding cilia function: defining specialized genes required for compartmentalized cilia biogenesis. Cell 117:527–539
9. Baker MA, Hetherington L, Reeves GM, Aitken RJ (2008) The mouse sperm proteome characterized via IPG strip prefractionation and LC-MS/MS identification. Proteomics 8:1720–1730
10. Baker MA, Hetherington L, Reeves G, Müller J, Aitken RJ (2008) The rat sperm proteome characterized via IPG strip prefractionation and LC-MS/MS identification. Proteomics 8:2312–2321
11. Bechstedt S, Albert JT, Kreil DP, Müller-Reichert T, Göpfert MC, Howard J (2010) A doublecortin containing microtubule-associated protein is implicated in mechanotransduction in Drosophila sensory cilia. Nat Commun 1:11
12. Blacque OE, Perens EA, Boroevich KA, Inglis PN, Li C, Warner A, Khattra J, Holt RA, Ou G, Mah AK, McKay SJ, Huang P, Swoboda P, Jones SJM, Marra MA, Baillie DL, Moerman DG, Shaham S, Leroux MR (2005) Functional genomics of the cilium, a sensory organelle. Curr Biol 15:935–941
13. Boesger J, Wagner V, Weisheit W, Mittag M (2009) Analysis of flagellar phosphoproteins from Chlamydomonas reinhardtii. Eukaryotic Cell 8:922–932
14. Broadhead R, Dawe HR, Farr H, Griffiths S, Hart SR, Portman N, Shaw MK, Ginger ML, Gaskell SJ, McKean PG, Gull K (2006) Flagellar motility is required for the viability of the bloodstream trypanosome. Nature 440:224–227
15. Cachero S, Simpson TI, Zur Lage PI, Ma L, Newton FG, Holohan EE, Armstrong JD, Jarman AP (2011) The gene regulatory cascade linking proneural specification with differentiation in Drosophila sensory neurons. PLoS Biol 9:e1000568
16. Cao W, Gerton GL, Moss SB (2006) Proteomic profiling of accessory structures from the mouse sperm flagellum. Mol Cell Proteomics 5:801–810
17. Chen N, Mah A, Blacque OE, Chu J, Phgora K, Bakhoum MW, Newbury CRH, Khattra J, Chan S, Go A, Efimenko E, Johnsen R, Phirke P, Swoboda P, Marra M, Moerman DG, Leroux MR, Baillie DL, Stein LD (2006) Identification of ciliary and ciliopathy genes in Caenorhabditis elegans through comparative genomics. Genome Biol 7:R126
18. Datta M, Choudhury A, Lahiri A, Bhattacharyya NP (2011) Genome wide gene expression regulation by HIP1 protein interactor, HIPPI: prediction and validation. BMC Genomics 12:463
19. Dorus S, Busby SA, Gerike U, Shabanowitz J, Hunt DF, Karr TL (2006) Genomic and functional evolution of the Drosophila melanogaster sperm proteome. Nat Genet 38:1440–1445
20. Efimenko E, Bubb K, Mak HY, Holzman T, Leroux MR, Ruvkun G, Thomas JH, Swoboda P (2005) Analysis of xbx genes in C. elegans. Development 132:1923–1934
21. Fritz-Laylin LK, Cande WZ (2010) Ancestral centriole and flagella proteins identified by analysis of Naegleria differentiation. J Cell Sci 123(Pt 23):4024–4031
22. Geremek M, Bruinenberg M, Ziętkiewicz E, Pogorzelski A, Witt M, Wijmenga C (2011) Gene expression studies in cells from primary ciliary dyskinesia patients identify 208 potential ciliary genes. Hum Genet 129:283–293
23. Geremek M, Ziętkiewicz E, Bruinenberg M, Franke L, Pogorzelski A, Wijmenga C, Witt M (2014) Ciliary genes are down-regulated in bronchial tissue of primary ciliary dyskinesia patients. PLoS One 9:e88216
24. Guo X, Shen J, Xia Z, Zhang R, Zhang P, Zhao C, Xing J, Chen L, Chen W, Lin M, Huo R, Su B, Zhou Z, Sha J (2010) Proteomic analysis of proteins involved in spermiogenesis in mouse. J Proteome Res 9:1246–1256
25. Hodges ME, Wickstead B, Gull K, Langdale JA (2011) Conservation of ciliary proteins in plants with no cilia. BMC Plant Biol 11:185
26. Hoh RA, Stowe TR, Turk E, Stearns T (2012) Transcriptional program of ciliated epithelial cells reveals new cilium and centrosome components and links to human disease. PLoS One 7:e52166
27. Huang X-Y, Guo X-J, Shen J, Wang Y-F, Chen L, Xie J, Wang N-L, Wang F-Q, Zhao C, Huo R, Lin M, Wang X, Zhou Z-M, Sha J-H (2008) Construction of a proteome profile and functional analysis of the proteins involved in the initiation of mouse spermatogenesis. J Proteome Res 7:3435–3446
28. Hughes JR, Meireles AM, Fisher KH, Garcia A, Antrobus PR, Wainman A, Zitzmann N, Deane C, Ohkura H, Wakefield JG (2008) A microtubule interactome: complexes with roles in cell cycle and mitosis. PLoS Biol 6:e98
29. Ishikawa H, Thompson J, Yates JR, Marshall WF (2012) Proteomic analysis of mammalian primary cilia. Curr Biol 22:414–419
30. Ivliev AE, 't Hoen PAC, Van Roon-Mom WMC, Peters DJM, Sergeeva MG (2012) Exploring the transcriptome of ciliated cells using in silico dissection of human tissues. PLoS One 7:e35618
31. Jakobsen L, Vanselow K, Skogs M, Toyoda Y, Lundberg E, Poser I, Falkenby LG, Bennetzen M, Westendorf J, Nigg EA, Uhlen M, Hyman AA, Andersen JS (2011) Novel asymmetrically localizing components of human centrosomes identified by complementary proteomics methods. EMBO J 30:1520–1535
32. Keller LC, Romijn EP, Zamora I, Yates JR, Marshall WF (2005) Proteomic analysis of isolated chlamydomonas centrioles reveals orthologs of ciliary-disease genes. Curr Biol 15:1090–1098
33. Kilburn CL, Pearson CG, Romijn EP, Meehl JB, Giddings TH, Culver BP, Yates JR, Winey M (2007) New Tetrahymena basal body protein components identify basal body domain structure. J Cell Biol 178:905–912

34. Kim J, Lee JE, Heynen-Genel S, Suyama E, Ono K, Lee K, Ideker T, Aza-Blanc P, Gleeson JG (2010) Functional genomic screen for modulators of ciliogenesis and cilium length. Nature 464:1048–1051
35. Kubo A, Yuba-Kubo A, Tsukita S, Tsukita S, Amagai M (2008) Sentan: a novel specific component of the apical structure of vertebrate motile cilia. Mol Biol Cell 19:5338–5346
36. Laurençon A, Dubruille R, Efimenko E, Grenier G, Bissett R, Cortier E, Rolland V, Swoboda P, Durand B (2007) Identification of novel regulatory factor X (RFX) target genes by comparative genomics in Drosophila species. Genome Biol 8:R195
37. Lauwaet T, Smith AJ, Reiner DS, Romijn EP, Wong CCL, Davids BJ, Shah SA, Yates JR, Gillin FD (2011) Mining the Giardia genome and proteome for conserved and unique basal body proteins. Int J Parasitol 41:1079–1092
38. Li JB, Gerdes JM, Haycraft CJ, Fan Y, Teslovich TM, May-Simera H, Li H, Blacque OE, Li L, Leitch CC, Lewis RA, Green JS, Parfrey PS, Leroux MR, Davidson WS, Beales PL, Guay-Woodford LM, Yoder BK, Stormo GD, Katsanis N, Dutcher SK (2004) Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. Cell 117:541–552
39. Liu Q, Tan G, Levenkova N, Li T, Pugh EN, Rux JJ, Speicher DW, Pierce EA (2007) The proteome of the mouse photoreceptor sensory cilium complex. Mol Cell Proteomics 6:1299–1317
40. Martínez-Heredia J, Estanyol JM, Ballescà JL, Oliva R (2006) Proteomic identification of human sperm proteins. Proteomics 6:4356–4369
41. Mayer U, Ungerer N, Klimmeck D, Warnken U, Schnölzer M, Frings S, Möhrlen F (2008) Proteomic analysis of a membrane preparation from rat olfactory sensory cilia. Chem Senses 33:145–162
42. Mayer U, Küller A, Daiber PC, Neudorf I, Warnken U, Schnölzer M, Frings S, Möhrlen F (2009) The proteome of rat olfactory sensory cilia. Proteomics 9:322–334
43. McClintock TS, Glasser CE, Bose SC, Bergman DA (2008) Tissue expression patterns identify mouse cilia genes. Physiol Genomics 32:198–206
44. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Maréchal-Drouard L, Marshall WF, Qu L-H, Nelson DR, Sanderfoot AA, Spalding MH, Kapitonov VV, Ren Q, Ferris P, Lindquist E, Shapiro H, Lucas SM, Grimwood J, Schmutz J, Cardol P, Cerutti H, Chanfreau G, Chen C-L, Cognat V, Croft MT, Dent R et al (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science 318:245–250
45. Müller H, Schmidt D, Steinbrink S, Mirgorodskaya E, Lehmann V, Habermann K, Dreher F, Gustavsson N, Kessler T, Lehrach H, Herwig R, Gobom J, Ploubidou A, Boutros M, Lange BMH (2010) Proteomic and functional analysis of the mitotic Drosophila centrosome. EMBO J 29:3344–3357
46. Nakachi M, Nakajima A, Nomura M, Yonezawa K, Ueno K, Endo T, Inaba K (2011) Proteomic profiling reveals compartment-specific, novel functions of ascidian sperm proteins. Mol Reprod Dev 78:529–549
47. Nogales-Cadenas R, Abascal F, Díez-Pérez J, Carazo JM, Pascual-Montano A (2009) CentrosomeDB: a human centrosomal proteins database. Nucleic Acids Res 37(Database issue):D175–D180
48. Phirke P, Efimenko E, Mohan S, Burghoorn J, Crona F, Bakhoum MW, Trieb M, Schuske K, Jorgensen EM, Piasecki BP, Leroux MR, Swoboda P (2011) Transcriptional profiling of C. elegans DAF-19 uncovers a ciliary base-associated protein and a CDK/CCRK/LF2p-related kinase required for intraflagellar transport. Dev Biol 357:235–247
49. Reinders Y, Schulz I, Gräf R, Sickmann A (2006) Identification of novel centrosomal proteins in Dictyostelium discoideum by comparative proteomic approaches. J Proteome Res 5:589–598
50. Ross AJ, Dailey LA, Brighton LE, Devlin RB (2007) Transcriptional profiling of mucociliary differentiation in human airway epithelial cells. Am J Respir Cell Mol Biol 37:169–185
51. Sakamoto T, Uezu A, Kawauchi S, Kuramoto T, Makino K, Umeda K, Araki N, Baba H, Nakanishi H (2008) Mass spectrometric analysis of microtubule co-sedimented proteins from rat brain. Genes Cells 13:295–312
52. Sauer G, Körner R, Hanisch A, Ries A, Nigg EA, Silljé HHW (2005) Proteome analysis of the human mitotic spindle. Mol Cell Proteomics 4:35–43
53. Smith JC, Northey JGB, Garg J, Pearlman RE, Siu KWM (2005) Robust method for proteome analysis by MS/MS using an entire translated genome: demonstration on the ciliome of Tetrahymena thermophila. J Proteome Res 4:909–919
54. Stolc V, Samanta MP, Tongprasit W, Marshall WF (2005) Genome-wide transcriptional analysis of flagellar regeneration in Chlamydomonas reinhardtii identifies orthologs of ciliary disease genes. Proc Natl Acad Sci U S A 102:3703–3707
55. Stubbs JL, Oishi I, Izpisúa Belmonte JC, Kintner C (2008) The forkhead protein Foxj1 specifies node-like cilia in Xenopus and zebrafish embryos. Nat Genet 40:1454–1460
56. Wigge PA, Jensen ON, Holmes S, Souès S, Mann M, Kilmartin JV (1998) Analysis of the Saccharomyces spindle pole by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry. J Cell Biol 141:967–977
57. Yano J, Rajendran A, Valentine MS, Saha M, Ballif BA, Van Houten JL (2013) Proteomic analysis of the cilia membrane of Paramecium tetraurelia. J Proteomics 78:113–122
58. Guberman JM, Ai J, Arnaiz O, Baran J, Blake A, Baldock R, Chelala C, Croft D, Cros A, Cutts RJ, Di Genova A, Forbes S, Fujisawa T, Gadaleta E, Goodstein DM, Gundem G, Haggarty B, Haider S, Hall M, Harris T, Haw R, Hu S, Hubbard S, Hsu J, Iyer V, Jones P, Katayama T, Kinsella R, Kong L, Lawson D et al (2011) BioMart Central Portal: an open database network for the biological community. Database 2011:bar041–bar041

10.1186/2046-2530-3-9
Cite this article as: Arnaiz *et al.*: Remodeling Cildb, a popular database for cilia and links for ciliopathies. *Cilia* 2014, 3:9