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: Orthology 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 cillum 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.

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.)
The 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 remedated 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 MK[K/P]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 acetic 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 organelar 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 organelar 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 acetic species.
| 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 |
| Iviev 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 |
| Laumençon 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 |
| Martinez-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)

| Study Authors          | Study Title                                      | Species            | Study Type                  | Identification   |
|------------------------|--------------------------------------------------|--------------------|-----------------------------|------------------|
| Nakachi et al., 2011   | Sperm tail proteome                              | Homo sapiens       | yes                         |                  |
| Nogales-Cadenas et al., 2009 | Centrosome human curation                        | Homo sapiens       | yes                         |                  |
| Ostrowski et al., 2002 | Cilium proteome                                  | Homo sapiens       | yes                         |                  |
| Pazour et al., 2005    | Expression during ciliogenesis                   | Caenorhabditis elegans | yes                      |                  |
| Pazour et al., 2005    | Flagellum proteome                               | Homo sapiens       | yes                         |                  |
| Phirke et al., 2011    | Down and upregulated genes in daf-19 mutant       | Caenorhabditis elegans | yes                      |                  |
| Reinders et al., 2006  | Nuclear-associated body proteome                 | Dictyostelium discoideum | no                      |                  |
| Ross et al., 2007      | Expression during ciliogenesis                   | Homo sapiens       | yes                         |                  |
| Sakamoto et al., 2008  | Proteome of Microtubule-Associated Proteins      | Rattus norvegicus  | no                          |                  |
| Sauer et al., 2005     | Mitotic spindle proteome                         | Homo sapiens       | no                          |                  |
| Smith et al., 2005     | Cilium proteome                                  | Tetrahymena thermophila | yes                      |                  |
| Stolc et al., 2005     | Expression during ciliogenesis                   | Chlamydomonas reinhardtii | yes                  |                  |
| Stubbs et al., 2008    | Expression Under Fox1 silencing                 | Xenopus laevis     | yes                         |                  |
| Wigge et al., 1998     | Spindle pole body proteome                       | Saccharomyces cerevisiae | no                      |                  |
| Yano et al., 2013      | Ciliary membrane proteome                        | Paramecium tetracella | 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...
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
| Protein ID         | Synonyms                       | Mus | Rattus | Danio | Apis  | Drosophila | Class |
|-------------------|--------------------------------|-----|--------|-------|-------|------------|-------|
| ENSP00000383078   | PAFAH1B1,LS1,LS2,MDCR           | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000364691   | CROCC,ROLT,Rotletin             | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000309591   | PRKACA,PKACA                    | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000263710   | CLASP1,MAS1                     | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000263811   | DYN12,DNC2,J2C                  | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000216911   | AURKA,AIKARK1,AURA,AURORA2      | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000364721   | MAPRE1,EB1,EB                    | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000265563   | PRKAR2A,PRK2,PRK2               | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000359966   | NEK2,HPCT1,NEK2A,NLK1           | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000261965   | TUBGCP3,GCP3,GPCB9              | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000252936   | TUBGCP2,GPCP2,Grip103,h103p     | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000251413   | TUBG1,DCBM4,GCP-1               | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000456648   | TUBGCP4,76,F,GCP-4,GCP4         | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000323302   | POC18,PX1,PUW12,DWR51B          | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000324464   | CSNK1D,ASPS,CSKdelta,FASP52,HCXID | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000270861   | PLK4,SAK,STK18,Sak              | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000356785   | NME7,MZ23H7,NFD7                | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000273130   | DYN1C1U1,DNCU1,JC1              | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000359300   | CETN2,ALT,CE2W                  | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000287380   | TBC1D31,Gm85,DWR67              | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000287482   | SASS6,SA5,SA5                   | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000300093   | PLK1,PLKSTK13                   | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000257287   | CEP135,CEP4,MC1H8               | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000439376   | DCTN2,DCTN50,YAMNITIN,RBPS50    | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000395302   | OKAP5,ToG,TG1,HTO1,MOFS         | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000342510   | CEP97,LRP9Q2                    | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000348965   | DYN1CH1,DHC1,DHC1a              | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000496972   | CETN2,ALT,CE2W                  | yes | yes    | yes   | yes   | yes        | 1 (yyyyy) |
| ENSP00000317156   | CEP192,PPP1,PB2                 | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
| ENSP00000270708   | WRAP73,DWR9                    | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
| ENSP00000248846   | TUBGCP6,GCP6,GCP5,MCCRP,MC1HR   | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
| ENSP00000393583   | AZ1,AZ1,AzeC131,ZA1             | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
| ENSP00000283645   | TUBGCP5,GCP5                   | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
| ENSP00000303058   | CEP120,CCDC100                 | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
| ENSP00000313752   | SSNA1,N14,NA-14                | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
| ENSP00000355812   | FGR10P1,FOP                    | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
| ENSP00000343818   | CDK5RAP2,C48,CEP215,MCP3       | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
| ENSP00000344314   | OFD1,CXorf5,97TS,97P2R         | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
| ENSP000003171744  | PIBF1,C13orf24,CEP90           | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
| ENSP00000204726   | GOLGA3,GCP170,MEA-2,golgin-160 | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
| ENSP00000206474   | HAUS4,C14orf94                 | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
| ENSP00000229757   | CEP128,C14orf145,14orf61,LED4P132 | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
| ENSP00000262127   | CEP76,C18orf9,67T1705          | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
| ENSP00000370803   | CCP10,Cep110,CP10              | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
| ENSP00000263284   | CCDC61                         | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
| ENSP00000223208   | CEP41,9TS15,TSG14              | yes | yes    | yes   | yes   | no         | 2 (yyyny) |
## Table 2 Evolutionary conservation of centrosomal proteins viewed through Cildb V3.0 (Continued)

| Gene ID          | Gene Name(s)          | Human Orthology | Mouse Orthology | Rat Orthology | Fish Orthology | Insect Orthology | Human Inparalogy | Mouse Inparalogy | Rat Inparalogy | Fish Inparalogy | Insect Inparalogy | Conserved Proteins |
|------------------|-----------------------|-----------------|-----------------|---------------|---------------|------------------|-----------------|-----------------|----------------|----------------|------------------|--------------------|
| ENSP00000383769  | AKNA                  | yes             | yes             | yes           | no            | no               | 4 (ynynn)       |
| ENSP00000302537  | MDM1                  | yes             | yes             | yes           | no            | no               | 4 (ynynn)       |
| ENSP00000264935  | CEP72, FL/O565        | yes             | yes             | yes           | no            | no               | 4 (ynynn)       |
| ENSP00000419231  | CEP70, BITE           | yes             | yes             | yes           | no            | no               | 4 (ynynn)       |
| ENSP00000306105  | CEP89, CDC123         | yes             | yes             | yes           | no            | no               | 4 (ynynn)       |
| ENSP00000380661  | CEP250, C-NAP1, CEP2, C-NAP1 | yes | yes | yes | no | no | 4 (ynynn) |
| ENSP00000356579  | CEP350, CAP350, GM133 | yes             | yes             | yes           | no            | no               | 4 (ynynn)       |
| ENSP00000260372  | HAUS2C15orf25, CEP27HsT17025 | yes | yes | yes | no | no | 4 (ynynn) |
| ENSP00000363540  | CEP55, C1orf3, CT111, LRRCC6 | yes | yes | yes | no | no | 4 (ynynn) |
| ENSP00000355500  | CEP70, FAM68A, KAB    | yes             | yes             | yes           | no            | no               | 4 (ynynn)       |
| ENSP00000369871  | HAUS6, Dgt6, FAM29A   | yes             | yes             | yes           | no            | no               | 4 (ynynn)       |
| ENSP00000371308  | CENP1, BMM32, CENP-J, CPAP, LAPI, LIP1, MCPH6, Sars-4 | yes | yes | yes | no | no | 4 (ynynn) |
| ENSP00000282058  | HAUS1, CDC5, HEI-C, HEIC | yes | yes | yes | no | no | 4 (ynynn) |
| ENSP00000281122  | CETN1, CDC31, CEN3    | yes             | yes             | yes           | no            | no               | 4 (ynynn)       |
| ENSP00000352572  | PCNT, XEN, MOPD2, PCN, PCNT2, PCNTB | yes | yes | yes | no | no | 4 (ynynn) |
| ENSP00000295872  | SPICE1, CDC52, SPICE  | yes             | yes             | yes           | no            | no               | 4 (ynynn)       |
| ENSP00000317790  | CEP57, MAV2, PIGTSP57 | yes             | yes             | yes           | no            | no               | 4 (ynynn)       |
| ENSP00000426129  | CEP63                 | yes             | yes             | yes           | no            | no               | 4 (ynynn)       |
| ENSP00000308021  | CEP290, BBS14, JBTSS, LCA10, MKS4, NPHP6, PCOC3, rd16, SLSN6 | yes | yes | yes | no | no | 4 (ynynn) |
| ENSP00000493956  | HAUS5, Dgt5           | yes             | yes             | yes           | no            | no               | 4 (ynynn)       |
| ENSP00000462740  | CEP41, JBT51, TSGA14  | yes             | yes             | yes           | no            | no               | 4 (ynynn)       |
| ENSP00000265717  | PRKAR2B, PRKAR2R, R-8TA | yes | no | yes | yes | yes | 5 (ynyn) |
| ENSP00000345892  | NDE1, HOM-TEF715, SNE1, NUDE | yes | no | yes | yes | yes | 5 (ynyn) |
| ENSP00000383921  | ACT1, ARA1, AP1, CTR1N1 | yes | no | yes | yes | yes | 5 (ynyn) |
| ENSP00000447907  | DYN1L1, DLCHR1, DLCHR2, DNCL1, DNCL2, DCL1, DCL8 | yes | no | no | no | 6 (ynyn) |
| ENSP00000279195  | CEP164, NPHP15        | yes             | yes             | yes           | no            | no               | 6 (ynyn)       |
| ENSP00000264448  | ALMS1, ALSS           | yes             | yes             | yes           | no            | no               | 6 (ynyn)       |
| ENSP00000316681  | KIAA1731              | yes             | yes             | yes           | no            | no               | 6 (ynyn)       |
| ENSP00000456335  | CNTO8BLP18, PIP1221   | yes             | yes             | yes           | no            | no               | 6 (ynyn)       |
| ENSP00000348563  | AKAP9, AKAP550, AKAP450, CG-NAP, HYPERON, LCT11 | yes | no | no | no | 7 (ynyn) |
| ENSP00000384844  | DCTN1, DAP-150, P135  | no              | no              | no            | yes           | yes              | 8 (ynyn)       |

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 inparalog 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 Drosophila, 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/ctn1 (ENSP0000034844) 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 ENSMUSG0000031865 are related by Inparanoid orthology through three groups, hspa_mmu.17178 (one human and one mouse gene), hspa_mmu.1073 (four human and one mouse gene) and hspa_mmu.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 Foetootpath 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 contest 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, Sperlzig L, Dadielz M, Kolf F, Cohen J (2009) Cildb: a knowledgebase for centosomes and cilia. Database (Oxford) 2009: bap022
2. Ostrowski LE, Blackburn K, Radde KW, Moyer MB, Schlatter DM, Moseley A, Boucher RC (2002) A proteomic analysis of human cilium: 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 EL (2005) Inparanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res 33(Database issue): D476–D480
5. Pazour GJ, Grin L, Irwin GM (2005) Proteomic analysis of a eukaryotic cilium. J Cell Biol 170:103–113
6. Laigle C, Klotz C, De Loubresse NG, Lemut TC, Hori M, Laurent FX, Papon JP, Louis B, Cohen J, Kolf F (2008) 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, Grot JF, Bertémier M, Bouhouche K, Cohen J, Duret L, Kapusta A, Meyer E, Sperlzig L (2010) Gene expression in a paleopolyploid: a transcriptome profile of ciliated epithelial cells reveals new cilium and centrosome components and links to human disease. PLoS One 7:e52166
8. Huang X-Y, Guo X-J, Shen J, Wang Y-F, Chen X, Xie J, Wang N-L, Wang F-Q, Zhao C, Hua R, Su B, Zhou S, Sha J (2010) Proteomic analysis of proteins involved in spermiogenesis in mouse. J Proteome Res 9:1246–1256
9. Hedges ME, Wickstead B, Gull K, Langdale JA (2011) Conservation of ciliary proteins in plants with no cilia. BMC Plant Biol 11:68
10. Hoh RF, 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
11. Bechsteini S, Albert JT, Kreip DF, 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, Boroveich KA, Inglis PN, Li C, Warner A, Khetta J, Holt RA, Ou G, Mah AK, McKay SJ, Huang P, Swoboda P, Jones SJM, Mann MA, Baille HL, Moerman DG, Shaham S, Leroux MR (2005) Functional genomics of the ciliome, a sensory organ. Curr Biol 15:935–941
13. Boesiger J, Wagner V, Weisheit W, Mittag M (2009) Analysis of flagellar phosphoproteins from Chlamydomonas reinhardtii. Eur J Biochem 269:362–373
14. Broadhead R, Dawe HR, Farr H, Griffiths S, Hart SR, Portman N, Shaw MK, Ginger ML, Gaskell SJ, McLean PG, Gull K (2006) Flagellar motility is required for the viability of the bloodstream trypanosome. Nature 440:224–227
15. Gacheru S, Simpson TL, Zirage PL, Ma L, Newton FG, Holbahn 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, Kahloum MW, Newbury CRH, Khatra J, Chen J, Go A, Blacque OE, Johnsen R, Phitke P, Swoboda P, Mann MA, Moerman DG, Leroux MR, Baille HL, Steen LD (2006) Identification of ciliary and ciliprotein genes in Caenorhabditis elegans through comparative genomics. Genome Biol 7:R26
18. Datta M, Choudhury A, Lahiri A, Bhattacharyya NP (2011) Genome wide gene expression regulation by HIP1, a protein interactor, HIPPI: prediction and validation. BMC Genomics 12:463
19. Dorus S, Busby SA, Gerike U, Shabanowitz J, Hunt DF, Katt NL (2006) Genomic and functional evolution of the Drosophila melanogaster sperm proteome. Nat Genet 38:1440–1445
20. Elfenenko E, Bubb B, Mah H, Holman T, Leroux MR, Ruskun G, Thomas JH, Swoboda P (2005) Analysis of dxe genes in C. elegans. Development 132:1923–1934
21. Fritz-Laylin LK, Cande WZ (2010) Ancestral cilicate and flagella genes identified by analysis of Naegleria differentiation. J Cell Sci 123(Pt 23):4024–4031
22. Geremek M, Brünenberg 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, Brünenberg 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, Hua R, Su B, Zhou S, Sha J (2010) Proteomic analysis of proteins involved in spermiogenesis in mouse. J Proteome Res 9:1246–1256
25. Hedges ME, Wickstead B, Gull K, Langdale JA (2011) Conservation of ciliary proteins in plants with no cilia. BMC Plant Biol 11:185
26. Hoh RF, 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, Hui R, Lin M, Wang X, Zhou Z-M, Sha J-H (2008) Construction of a ciliary proteome 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. Ishiwa H, Thompson J, Yates JR, Marshall WF (2012) Proteomic analysis of mammalian primary cilia. Curr Biol 22:414–419
30. Ivey AE, ’t Hoorn PAC, Van Room-Wiric, Peters DM, 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 cilia reveals orthology of cilary-disease genes. Curr Biol 15:1090–1098
33. Kilburn CL, Pearson CG, Romijn EP, Meehl JB, Giddings TH, Culver BP, Yates JR, Wines 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, Gieseon JS (2010) Functional genomic screen for modulators of ciliogenesis and cilium length. Nature 464:1048–1051
35. Kubo A, Yuda-Kubo A, Tsukita S, Tsukita S, Armagai M (2008) Sentan: a novel specific component of the apical structure of vertebrate motile cilia. Mol Biol Cell 19:5338–5346
36. Laurentenç A, Duburulle R, Efimienko E, Grenier G, Bissert R, Cortier E, Rolland V, Svoboda P, Durand B (2007) Identification of novel regulatory factor X (RFX) target genes by comparative genomics in Drosophila species. Genome Biol 8:R191
37. Lauwarte 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 OJ, 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 JL, Speicher DW, Fierce EA (2007) The proteome of the mouse photoreceptor 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, Klimmek D, Warnken U, Schnölzer M, Frings S, Möhrlen F (2009) The proteome of the mouse photoreceptor cilium complex. Mol Cell Proteomics 8:334–334
42. Mayer U, Küller A, Daiber PC, Neudorf I, 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
43. McClintock TS, Glasser CE, Bose SC, Bergman DA (2008) Tissue expression patterns identify mouse cilia genes. Physiol Genomics 32:198–205
44. Merchant SS, Pochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fitz-Laylin LK, Marcé-Chouza L, Marshall WF, Qu L-H, Nelson DR, Sanderfoot AA, Spalding MH, Kapitonov VV, Ferris P, Lindquist E, Shapiro H, Lucas SM, Grimwood J, Schmutz J, Cardol P, Cerutti V, Swoboda P, Durand B (2007) Proteomic profiling reveals compartment-specific, novel functions of mouse cilia genes. Physiol Genomics 32:198–205
45. Müller H, Schmidt D, Steinbrink S, Mirogersdkaia E, Lehmann V, Habermann K, Diehre F, Gustavsson N, Kesler T, Lehbrach H, Herwig R, Gobom J, Plouibidou 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 I, Carazo JM, Pascual-Montano A (2009) CentrosomeDB: a human centrosomal proteins database. Nucleic Acids Res 37(Database issue):D175–D180
48. Phirke P, Efimienko E, Mohan S, Bursthorn J, Croma F, Bakhoun MW, Trieb M, Schuske K, Jorgensen EM, Piascik BP, Leroux MR, Svoboda 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 Dictyostellium discoideum by comparative proteomic approaches. J Proteome Res 5:589–598
50. Ross AJ, Daley LA, Brighton LE, Devlin RB (2007) Transcriptional profiling of mucus cell 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, Nakarnishi 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, Res A, Nigg EA, Slijé 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 XKM (2005) Robust method for proteome analysis by MS/MS using an entire translated genome: demonstration on the ciliate 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, Okshiz Belmonte JC, Kintner C (2008) The forkhead protein Fox J1 specifies node-like cilia in Xenopus and zebrafish embryos. Nat Genet 40:1454–1460
56. Wigge PA, Jensen ON, Helms 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 Paramaecium tetraurelia. J Proteomics 78:113–122
58. Guberman JM, Al J, Arnaiz O, Baran J, Blake A, Baldock R, Chelba C, Croft D, Clos A, Cutt RJ, Di Genova A, Forbes S, Fujisawa T, Gadaleta E, Goodstein DM, Gundem G, Haggarty B, Haider S, Hall M, Harris T, Hu S, Hubbard S, Hsu J, Iyer V, Jones P, Katayama T, Kinosh R, Kong L, Lawson D et al (2011) BioMart Central Portal: an open database network for the biological community. Database 2011:bba041–bba041

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at www.biomedcentral.com/submit