Cyclophilin nomenclature problems, or, 'a visit from the sequence police'

Daniel W. Nebert,[1] Nickolas A. Sophos,[2] Vasilis Vasiliou[2] and David R. Nelson[3]*

[1]Department of Environmental Health and Center for Environmental Genetics (CEG), University of Cincinnati Medical Center, Cincinnati, OH 45267-0056, USA
[2]Molecular Toxicology and Environmental Health Sciences Program, Department of Pharmaceutical Sciences, University of Colorado Health Sciences Center, Denver, CO 80262, USA
[3]Department of Molecular Sciences and The UT Center of Excellence in Genomics and Bioinformatics, University of Tennessee, Memphis, TN 38163, USA
*Correspondence to: Tel: +1 513 558 4347; Fax: +1 513 558 3562; E-mail: dan.nebert@uc.edu

Date received (in revised form): 4th May 2004

Abstract

Why is agreement on one particular name for each gene important? As one genome after another becomes sequenced, it is imperative to consider the complexity of genes, genetic architecture, gene expression, gene–gene and gene–product interactions and evolutionary relatedness across species. To agree on a particular gene name not only makes one's own research easier, it aids automated text-mining algorithms and search engines, which are increasingly employed to find relationships in the millions of abstracts in the medical research literature and sequence databases. A common nomenclature system will also be helpful to the present generation, as well as future generations, of graduate students and postdoctoral fellows who are about to enter genomics research. In this paper, the authors present some problems that arose when two separate research communities decided to choose the same root, *CYP*, for naming their gene families. They then offer a logical solution, by renaming the cyclophilin genes with a common root, such as *cyn-* in *Caenorhabditis* and *CYN-* in mammals (*Cyn* in mouse), and using evolutionary divergence to cluster genes of the highest level of relatedness.

Keywords: *human genome, mouse genome,* Caenorhabditis elegans *genome, cytochrome P450 (CYP) gene superfamily, cyclophilin gene family, immunophilins, peptidylprolyl* cis-trans *isomerases, FK506-binding proteins, tacrolimus, parvulin*

Introduction

A previous paper in this series[1] summarised the steps that one is strongly encouraged to follow in order to ensure proper nomenclature of any gene. Three examples were given to illustrate how and why one should strive for a standardised gene nomenclature system. In these examples, the focus of the paper was on using the gene names as search terms, rather than comparing a DNA or protein sequence that has just been determined by searching via BLAST.[2] The three examples included: *PTGS1* and *PTGS2* as the correct gene names for prostaglandin G/H synthase-1 and -2, also known as cyclooxygenase-1 and -2 and commonly erroneously nicknamed 'COX-1' and 'COX-2' in many journals; the short- and long-chain fatty acid synthase gene families, for which there is currently no official agreed-upon nomenclature (although *FASN* on human chromosome 17q25 is the official symbol for the fatty acid synthase gene); and *POR* as the correct name for the NADPH-P450 oxidoreductase gene.[1] Before deciding upon a new gene symbol, the reader is encouraged to visit the website describing this topic.[3]

This theme is extended in the current paper, which shows how two completely separate research communities adopted the same gene root name, while not realising that the other group had done the same thing.

Cyclophilins as '*cyp-*' in a *Caenorhabditis elegans* database

As Head of the Cytochrome P450 (*CYP*) Superfamily Gene Nomenclature Committee, David Nelson maintains a website dedicated to cytochrome P450 gene nomenclature.[4] The *C. elegans* genome has 76 full-length P450 genes and nine pseudogenes, which have been assembled by Nelson during the past several years and were nearly completed after the

Table 1. List of cyclophilin and P450 genes in *C. elegans*.

Cyclophilin genes	Alternative name		WormPep accession #
cyp-1Y49A3A.5		CGC approved	CE22213
cyp-2ZK520.5		CGC approved	CE16730
cyp-3Y75B12B.5		CGC approved	CE20374
cyp-4F59E10.2	mog-6	CGC approved	CE01596
cyp-5F31C3.1		CGC approved	CE17730
cyp-6F42G9.2		CGC approved	CE01301
cyp-7Y75B12B.2		CGC approved	CE20371
cyp-8D1009.2a	D1009.2b	CGC approved	CE04286
cyp-9T27D1.1		CGC approved	CE03745
cyp-10B0252.4a	B0252.4b	CGC approved	CE02420
cyp-11T01B7.4		CGC approved	CE03588
cyp-12C34D4.12		CGC approved	CE17506
cyp-13Y116A8C.34		CGC approved	CE24152
cyp-14F39H2.2a	F39H2.2b	CGC approved	CE32410
cyp-15Y87G2A.6		CGC approved	CE24686
cyp-16Y17G7B.9		CGC approved	CE19042
cyp-17ZC250.1		CGC approved	CE28157
P450 genes			
ccp-13A7T10B9.10		CGC approved	CE01655
ccp-14A5F08F3.7		CGC approved	CE09262
ccp-31A1	C01F6.3	CGC approved	unavailable
ccp-44ZK177.5	cyp-44	CGC approved	CE25682

Data taken from Jonathan Hodgkin, CGC Genetic Map and Nomenclature Curator (**C***aenorhabditis* **G**enetics **C**enter), Genetics Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK.

genome sequence had been published. Recently, Dan Lawson of WormPep[5] asked Nelson to review assemblies of these genes, following the recent revision of the worm's genome. While carrying out this request, Nelson discovered that Dan Lawson had referred to several P450 genes as '*ccp-xx*'. Nelson then explored WormPep further and confirmed that *ccp* was being used as the root for cytochrome P450 genes. Although the usual root for P450 gene names is *CYP*, this term (*cyp*) was being used in the *C. elegans* protein database for the cyclophilins (Table 1). Can this be a problem — i.e. the same gene root being used for different gene families by colleagues in two separate, very distant research fields?

CYP for cytochrome P450 genes in all species

The mammalian cytochrome P450 (*CYP*) superfamily encodes enzymes involved in: the metabolism of pharmaceuticals, foreign chemicals and pollutants; arachidonic acid metabolism and eicosanoid biosynthesis; cholesterol, sterol and bile acid biosynthesis; steroid synthesis and catabolism; vitamin D_3 synthesis and catabolism; retinoic acid hydroxylation; biogenic amine and neuroamine metabolism; and several orphan CYP genes still of unknown function.[6] There are 102 and 57 putatively functional *CYP* genes in

the mouse and human, respectively.[7] To date, more than 3,400 P450 sequences have been named with the three-letter root of *CYP*. This nomenclature has been in place[8,9] since 1987, and is growing every day.[4] The official root names for mouse and human P450s are *Cyp* and *CYP*, respectively. The *Drosophila* nomenclature[10] also uses *Cyp*. There are now 727 genes in rice and *Arabidopsis* that have been named *CYP*.[4] It is anticipated that the number of named P450 genes will exceed 4,000 by the end of 2004.

Whereas continuing to use the *CYP* root for cyclophilin genes will be a nightmare for cyclophilin researchers, P450 researchers might find this an annoyance but not really much of a problem. To prevent conflicts over nomenclature, it becomes increasingly urgent to rename the cyclophilin genes. What is the best root name for these genes?

Finding the best root for the cyclophilin genes

The three families of immunophilins, known as peptidylprolyl *cis-trans* isomerases (PPIases), include the cyclophilins, the FK506-binding proteins (FKBPs) and parvulin.[11–13] All three gene families are found in animals, plants and eubacteria. While two cyclophilins and two types of FKBPs exist in archaebacteria, no parvulin homologue has been found. Parvulin is unique among the immunophilins. A search of the LocusLink,[14] HUGO Gene Nomenclature Committee,[15] and the National Center for Biotechnology Information (NCBI) UniGene[16] websites using 'parvulin', shows a single gene; *Pin4* and *PIN4* are the approved mouse and human gene names, respectively. 'PIN' is an abbreviation for peptidylprolyl *cis-trans* isomerase NIMA-interacting-4. 'NIMA' stands for 'never-in-mitosis-gene-a', which was first isolated as a series of conditional cell cycle mutants that failed to enter mitosis in *Aspergillus nidulans*.[17,18] There are 11 genes (*NEK1, NEK2, ... NEK11*) in the human genome that encode NIMA-related mitotic kinases and are involved in DNA replication and genotoxic stress responses.[19,20] Although parvulin has peptidylprolyl *cis-trans* isomerase activity, it shares no evolutionary homology with the FKBPs or cyclophilins.

Immunophilins are defined as receptors for immunosuppressive drugs including cyclosporin-A, FK506 and rapamycin. FK506 is also called tacrolimus, a macrolide of fungal origin (produced by *Streptomyces tsukubaensis*) and having strong immunosuppressive actions. FK506- and rapamycin-binding proteins are abbreviated as FKBPs and share no evolutionary homology with the cyclophilins or parvulin. A search of the LocusLink, HUGO Gene Nomenclature Committee and the NCBI UniGene websites using 'fkbp', shows more than 80 *FKBP* genes in the human and mouse (*FKBP1, FKBP2, ... FKBP82*). These gene products have many unique features, such as targeting BCL2 to the mitochondria and inhibiting apoptosis.[21]

Cyclophilins, the third and last class of the PPIases, comprise cyclosporin-A-binding proteins[22] ranging in size from 17 kDa to 324 kDa.[12] This class of immunophilins carries out a wide range of functions — including acting as a chaperone to facilitate the nuclear transport of the somatolactogenic hormones,[23] facilitating the calcium-regulated mitochondrial permeability transition pore which precedes apoptosis[24] and participating in the pre-mRNA splicing machinery.[25] Cyclophilin-binding drugs are emerging as potential leads to novel targets for interference with interleukin-12 production[26] and, therefore, to the possibility of treating conditions such as multiple sclerosis and rheumatoid arthritis. Cyclosporin-A also has activity against helminth and protozoan parasites.[27]

A search of the LocusLink, HUGO Gene Nomenclature Committee and the NCBI UniGene websites using 'cyclophilin', shows 15 putatively functional genes and 22 pseudogenes. The 15 putatively functional gene names (Table 2) include *PPIA* through to *PPIH* (for peptidylprolyl isomerase-A, -B, ... -H; cyclophilin A-, B-, ... H-related), one PPIA-like (*PPIAL3*), and six cyclophilin-like (*PPIL1, PPIL2, ... PPIL6*).

PPID has the synonym 'CYP-40', but this is no longer the official name. Unfortunately, the mouse RIKEN full-length cDNAs that match this sequence are being called CYP40, not PPID, so the name is propagating itself in the literature and into the databases in an uncontrollable way. The cloning and naming of 11 cyclophilin genes from *C. elegans* (*Cyp-1* to *Cyp-11*)[28] was reported in 1996. A search of GenBank for CYP20 finds AY568517, an *Arabidopsis* thylakoid lumen cyclophilin,[29] named CYP20-2. (CYP20A1 is a chordate cytochrome P450 of unknown function, possibly involved in development.)[4] The date on this *Arabidopsis* CYP20 GenBank entry is 15th April, 2004, showing that the problem is not going away. In fact, the PubMed link from the GenBank entry leads to a publication[30] in which a nomenclature system for the 29 cyclophilin genes in the *Arabidopsis thaliana* genome is presented using *CYP* as the root.

What is the solution?

Solutions — like politics — are local. We have contacted the *C. elegans* community and alerted them to this nomenclature conflict. They are responding and will select a new root for cyclophilins and change their P450 gene names to *cyp-*, from the current *ccp-* root. This will go into the official WormPep and WormBase nomenclature and will eventually prevent use of the *cyp-* root in *C. elegans* (and, hopefully, *C. briggsae*) for cyclophilins. Additional effort will be needed for the *Arabidopsis* community, as well as for the human and mouse gene databases.

Table 2. List of putatively functional human cyclophilin genes.

Approved gene symbol	Approved gene name	Chromosomal location
PPIA	Peptidylprolyl isomerase A (cyclophilin A)	7p13-p11.2
PPIAL3	Peptidylprolyl isomerase A (cyclophilin A)-like-3	21
PPIB	Peptidylprolyl isomerase B (cyclophilin B)	15
PPIC	Peptidylprolyl isomerase C (cyclophilin C)	[reserved]
PPID	Peptidylprolyl isomerase D (cyclophilin D)	4
PPIE	Peptidylprolyl isomerase E (cyclophilin E)	1p32
PPIF	Peptidylprolyl isomerase F (cyclophilin F)	10q22-q23
PPIG	Peptidylprolyl isomerase G (cyclophilin G)	2q31.1
PPIH	Peptidylprolyl isomerase H (cyclophilin H)	1p34.1
PPIL1	Peptidylprolyl isomerase (cyclophilin)-like 1	6p21.1
PPIL2	Peptidylprolyl isomerase (cyclophilin)-like 2	22
PPIL3	Peptidylprolyl isomerase (cyclophilin)-like 3	2
PPIL4	Peptidylprolyl isomerase (cyclophilin)-like 4	6q24-25
PPIL5	Peptidylprolyl isomerase (cyclophilin)-like 5	14q21.3
PPIL6	Peptidylprolyl isomerase (cyclophilin)-like 6	6q21

Not included here are *PPIAL*, *PPIAL2*, *PPIAP*, *PPIAP2*, *PPIAP3*, *PPIAP4*, *PPIAP5*, *PPIAP6*, *PPIHP1*, *PPIHP2*, *PPIL1P1*, *PPIP1*, *PPIP2*, *PPIP3*, *PPIP4*, *PPIP5*, *PPIP6*, *PPIP7*, *PPIP8*, *PPIP9*, *PPIP10* and *PPIP11*, which represent the 22 cyclophilin pseudogenes in the Human Genome Project (HGP) database.

What might be the best root for the cyclophilin gene family? *Cyn* has been used for cyclone, a mouse gene in LocusLink; *CPN1* and *CPN2* are being used for carboxypeptidase N-1 and -2; *Cph* was considered, but *CPH1* has been used to refer to a cryptochrome or phytochrome (light-sensing protein).[31] Because of the sharing of this paper (while still being written) with Lois Maltais of Mouse Genome Informatics (MGI), she consulted with the authors of the mouse cyclone gene paper and they have now agreed to use *Cycn*, in order to free up *Cyn* and *CYN* for the mouse and human cyclophilin, respectively. After searching databases and search engines for conflicts, the present authors suggest that *Cyn-* might be the most suitable root for *C. elegans* cyclophilins, but this needs to be decided among members of the worm community. It is unfortunate that some databases (eg worm, yeast and bacteria) are mandating that gene names be limited to three letters. The authors suspect that three-letter root names for the ~19,000 *C. elegans* genes may not be enough. For example, 10,000 families will require the same number of roots. 26 cubed is only 17,576; this will require the use of odd letter combinations that have no symbolic meaning, such as *xyz1*, *cxq*, *rzx*, etc. Also, the nature of language is to use some letters more often than others, which will put great pressure on naming the genes that begin with the most often-used letters. *CYN* has now been officially approved as the root to unify all mammalian cyclophilins.

Using evolutionary trees to assign names to genes in the P450 superfamily,[4,8,9] in the authors' experiences, has been very positive. This can work in general for any other homologous group of genes and, in fact, has been used for at least 124 families and/or superfamilies to date.[32] To illustrate this point, a simple sequence alignment (Figure 1) and tree (Figure 2) are presented for the *C. elegans* cyclophilins.

The vertical lines in Figure 2 are suggested break-points for family and subfamily designations. Branches on the tree intersected by the lines would define family and subfamily clusters. The lines could be moved to modify the number of families and subfamilies. As drawn, there are six subfamilies in family 1, and one each in families 2 and 3. Moving the subfamily line to the left could reduce the number of subfamilies in family 1 from six to three. If *cyn* were used, CE28157 (at the top of Figure 2) would be named *cyn3a1* and CE20374 (at the bottom of Figure 2) would be named *cyn1a1*, and so on.

A method for creating a network of 'gene co-occurrences' from the literature, and portioning it into communities of related genes, has recently been presented.[33] In that paper, a program is described (but not named) which searches all Medline titles and abstracts and OMIM entries for occurrences and co-occurrences of gene symbols, gene names and diseases; the databases contain more than 12 million abstracts. Relationships are identified by automated

```
CE20374   MSRSKVFFDITIGGKASGRIVMELYDDVVPKTAGNFRALCTGENGIGKSGKPLHFKGSKF
CE20371   MSRPRVFFDITIAGKPTGRIVMELYNDIVPKTAENFRALCTGEKGVGKSGKPLHFKGSKF
CE16730   MPRVKVFFDITIGGKKGGRIVMELYNDIVPKTAENFRALCTGEKGKGKSGKKLHFKGSKF
CE22213   ----------------------------------------EKGVGEQGVALHFKGSKF
CE24152   ---PRVYLGVKIGIRYIGRIVIELRTDVTPKTAENFRCLCTGERGFG-------YEGSIF
CE17730   ---------------------------------FIELAKKPKGEG-------YPGSKF
CE01301   ----KVFFDMEIGGRPVGKIVIGLFGEVVPKTVKNFVELAQRAEGEG-------YVGSKF
CE03588   ----------------------------------------GVPNG-------YKNCTF
CE04286   ----RAFFDISINGEPAGRIVFSLWNHCCPRTVENFRAFCTGELGKMN-GHYASYQGSVF
CE03745   ------------------------------------------------------EF
CE01596   ----------------------------PKACENFITHCSNGYYNN----------TKF
CE02420   ---------------SGDIKIELYVDDAPKACENFLALCASDYYNG----------CIF
CE17506   ------QAPYVILDTTMGKIALELYWNHAPRTCQNFSQLAKRNYYNG----------TIF
CE19042   ------------------------------------------------------------
CE24686   ------------------------------------------------------------
CE32410   ------------------------------------------------------------
CE28157   --QPNRWFDSIHYVTNTGKVIIQMNEALKKNLTELFVKTARGEFVHPSCGKKIEYTGTIL

CE20374   HRIIP-NFMIQGGDFTRGNGTGGES-IYGEKFPDENFK-EKHTGPGVLSMANAGP--NTN
CE20371   HRIIP-EFMIQGGDFTRGNGTGGES-IYGEKFPDENFK-EKHTGPGVLSMANAGP--NTN
CE16730   HRIIP-EFMIQGGDFTEGNGTGGES-IHGEKFDDENFK-EKHTGPGVLSMANCGA--NTN
CE22213   HRIIP-EFMIQGGDFTRHNGTGGES-IYGNKFKDENFD-LKHTGPGCLSMANAGP--NTN
CE24152   HRIIP-KFMLQGGDFTKGDGTGGKS-IYGTKFDDENFT-LRHTMPGTVSMANCGA--NTN
CE17730   HRVIA-DFMIQGGDFTRGDGTGGRS-IYGEKFADENFK-LKHYGAGWLSMANAGA--DTN
CE01301   HRVIE-NFMIQGGDFTRGDGTGGRS-IYGERFEDENFK-LQHYGPGWLSMANAGE--DTN
CE03588   HRVIK-DFMIQGGDFCNGDGTGLMS-IYGSKFRDENFE-LKHIGPGMLSMANAGS--DTN
CE04286   HRVIK-GFMIQGGDITHGNGTGGYS-IYGRTFDDENLA-LKHKKPYLLSMANRGP--DTN
CE03745   HRIVK-KFMIQGGDITEGDGRGGFS-IYGRYFDDEKFK-LKHSRPYLLSMANKGP--NSN
CE01596   HRLIK-NFMLQGGDPT-GTGHGGES-IWDKPFSDEFISGFSHDARGVLSMANKGS--NTN
CE02420   HRNIK-DFMVQTGDPT-HSGKGGES-IWGGPFEDEFVSALKHDSRGCVSMANNGP--DSN
CE17506   HRIIA-DFMIQGGDPT-GTGRGGAS-IYGDKFSDEIDERLKHTGAGILSMANAGP--NTN
CE19042   -------FILQGGDPT-ATGTGGES-IYGKPFKDEIHQRLKFNRRGIVGMANAGR--DDN
CE24686   ---------------------------GEDFEDEFHPRLRHDKPFKVSMANAGGG-NTN
CE32410   ----------KGGESVYSDMYGEQ----GRYFEREDLPKMRHTRMGIVSFVNNGD--NML
CE28157   HQISTSKNMIMGGDVLNGNGCGRCAPVSRKLFQENNFSSTVQNTRGKVILLPSDTNPTVF

CE20374   GSQFFLCTVK--TEWLDGKHVVFGRVVEGLDVVKAVESNGSQSG-KPVKDCMIADCGQLK
CE20371   GSQFFLCTVK--TAWLDGKHVVFGRVVEGLDIVSKVEGNGSSSG-TPKSECLIADCGQL-
CE16730   GSQFFLCTVK--TTWLDGKHVVFGKVIEGMDVVKAIESKGSEDG-APSAPCVIADCGEMK
CE22213   GSQFFICTVD--TPWLDGGHVVFGQVTDGMSVVKKIEKMGSRSG-APAKTVTIADCGELK
CE24152   GSQFFICTEK--TDWLDGKHVVFGHVVEGMNIVRQVEQQGTPSG-KPQMVVKIVESGEIE
CE17730   GSQFFITTVK--TPWLDGRHVVFGKILEGMDVVRKIEQTEKLPGDRPKQDVIIAASG---
CE01301   GSQFFITTAK--TSWLDGKHVVFGKILEGMDVVREIEATPKGAGDRPIEDVVIANAG---
CE03588   GCQFFITCAK--TDFLDNKHVVFGRVLDGMLTVRKIENVPTGANNKPKLPIVVVQCGQL-
CE04286   GSQFFITSEE--VPHLDGKHCVFGEVIKGVEVVKAIENLETGNEDKPVCKVEITHCGEM-
CE03745   SSQFFITTAA--APHCNGKHVVFGEVVKGQNVVDYIDNLAVDDKSKPLAKVLISNCGEL-
CE01596   GSQFFITFRP--CKYLDRKHTIFGRLVGGQDTLTTIEKL----------------------
CE02420   RSQFFITYAK--QAHLDMKYTLFGKVIDGFDTLEEIE-----------------------
CE17506   GSQFFITLAP--TQHLDGKHTIFGRVAAGMKV----------------------------
CE19042   GSQFFFTIGDRGAPELDKKHTIFGKVTGPT--LFNMLKITEVETEGDR------------
CE24686   GSQFFITVCP--ADWLDGKNTLFGEVTAGMSVVQRINQVSTFERSGRPRESIQI------
CE32410   GSQFFITLGEN-LDYLDDQHTIFGQVTEGLETLEKLNEQLADTNNRPFKDIRIS------
CE28157   SSLFYVLLDKSGPSVVDG--CPIGDVIEGIEILETIVKEYGSENGRPGKKLIVHNCGHL-
```

Figure 1. Sequence alignment of the conserved regions of 17 *Caenorhabditis elegans* cyclophilin proteins.

bioinformatics methods between genes, and between genes and diseases, that might not be detected by less computationally intense methods. Such methods must rely on consistent names, or they have to deal with a list of synonyms.

Conclusions

The cyclophilin gene nomenclature has several problems. First, many in the cyclophilin field continue to use *CYP*, which has been the gene root for the large cytochrome

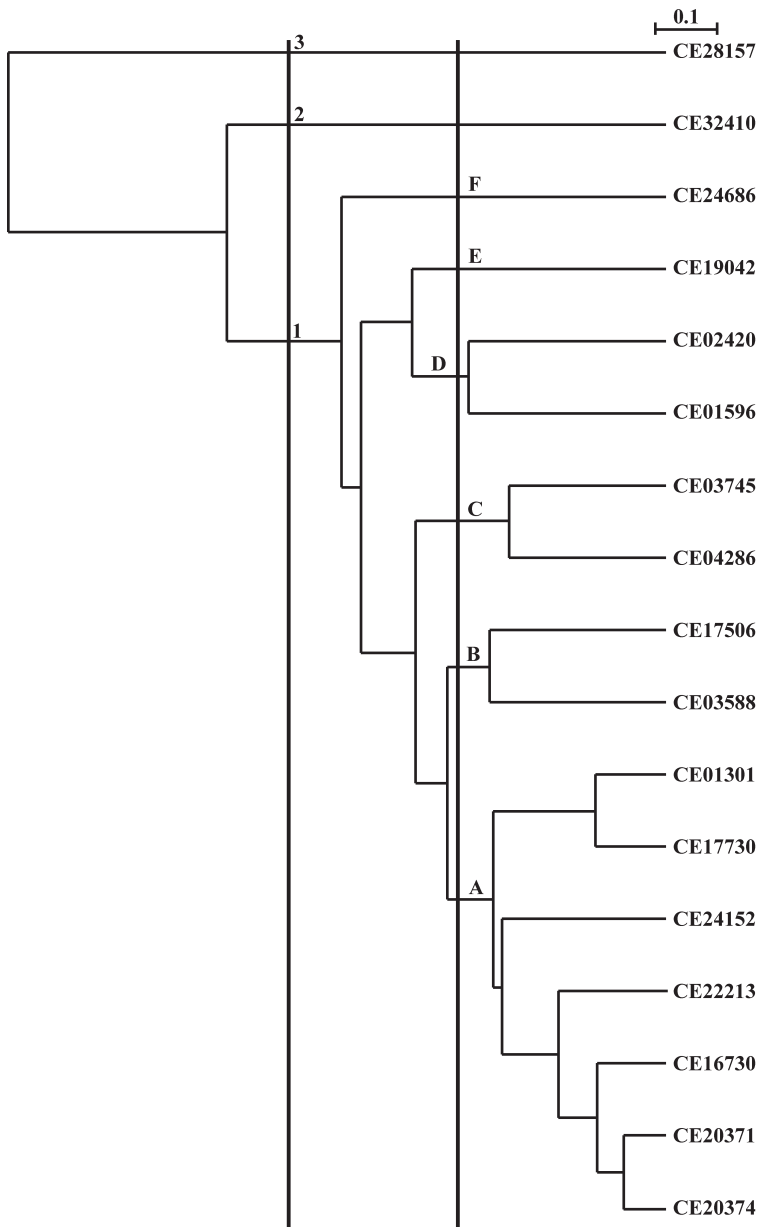

Figure 2. UPGMA tree and possible family and subfamily divisions of the *Caenorhabditis elegans* cyclophilin gene family, based on evolutionary divergence. The root of *cyn-*, is acceptable because it has not yet been used by any other gene database, except 'cyclone' in mouse. Families are designated by Arabic numerals and represent amino acid identity of 40 per cent or greater. Subfamilies are designated by letters and represent amino acid identity of 54 per cent or greater. Individual genes within subfamilies are then given Arabic numbers. Identifiers are the WormPep accession numbers.

P450 gene superfamily since 1987. Secondly, the gene root chosen by the HUGO Human/Mouse Gene Nomenclature Committees had been *PPI* for peptidylprolyl *cis-trans* isomerase, although — as detailed above — the cyclophilins represent just one of three classes of the PPIases that are perhaps functionally related but evolutionarily unrelated. Thirdly, the authors suggest the root *Cyn* for the *C. elegans* cyclophilin genes. Fourthly, eight of the 15 putatively functional human cyclophilin genes end in the letters 'A' through to 'H', while the others end in two groups of numbers (one PPIA-like and six PPI-like). It is strongly recommended that these genes be named by families and subfamilies, according to evolutionary divergence, as shown in Figure 3. Because of discussions

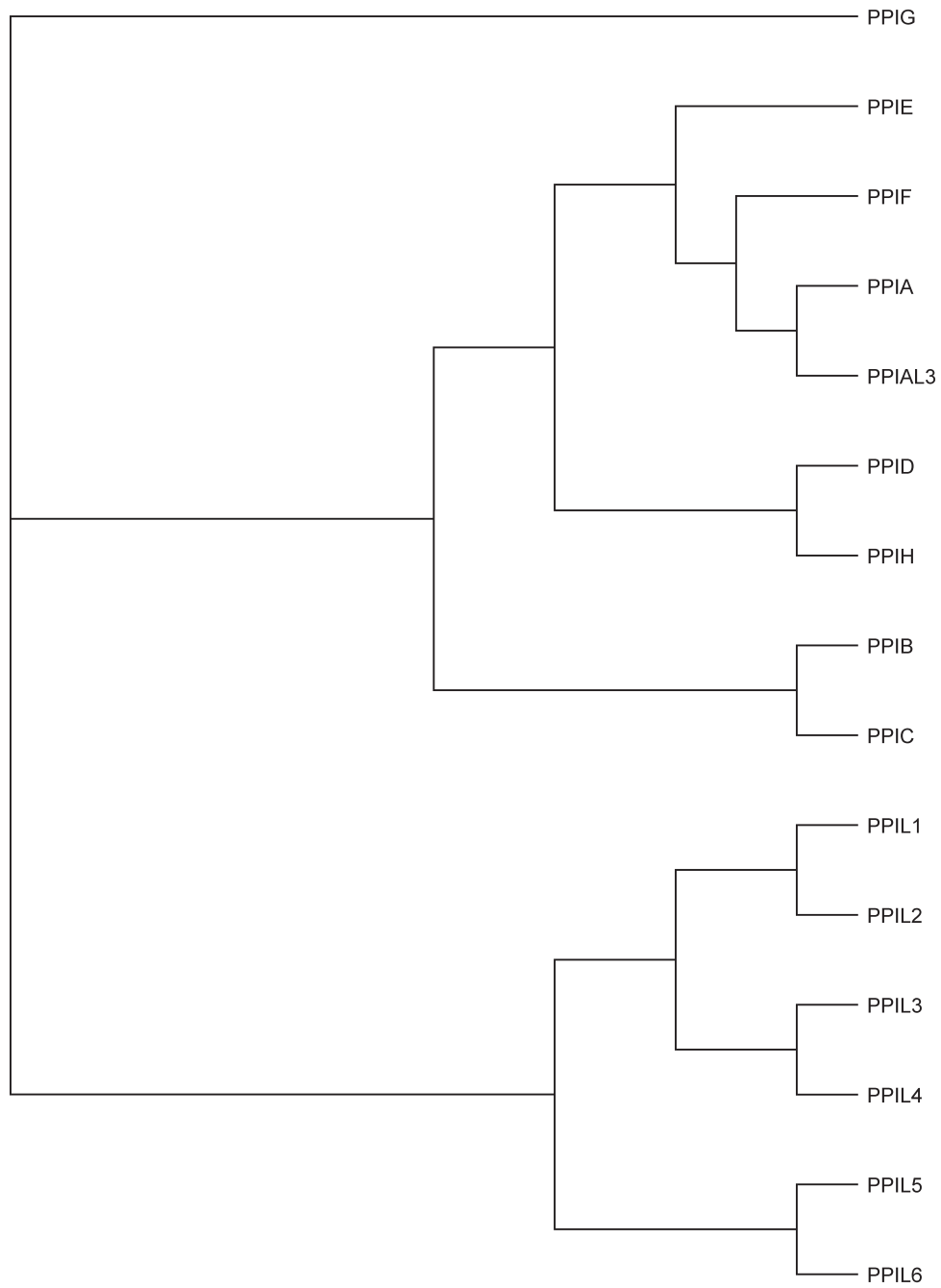

Figure 3. Rectangular cladogram of the 15 human cyclophilin proteins aligned. If one were to drop vertical lines, as in Figure 2, one might name: PPIL1 through to PPIL6 as *CYN1A1* through to *CYN1A4*, and *CYN1B1* and *CYN1B2*, respectively; PPIB and PPIC as *CYN2A1* and *CYN2A2*, respectively; PPIE, PPIF, PPIA and PPIAL3 as *CYN2C1*, *CYN2C2*, *CYN2C3* and *CYN2C4*, respectively; PPIG as *CYN3*; and so on.

related to the writing of this paper, the *CYN* root has now been officially approved for mammals. It would be desirable to incorporate as many species as possible into the naming scheme. One additional source of nomenclature friction is the strict use of three letter roots for gene names in *C. elegans*, yeast and bacteria; this automatically creates conflicts when human and mouse root names can be much longer than three letters, as in *PPIAL3* or *NIPSNAP1*; however, that is a battle for another day.

Acknowledgments

The writing of this article was funded, in part, by NIH grant P30 ES06096 (D.W.N.).

References

1. Nebert, D.W. and Wain, H.M. (2003), 'Update on human genome completion and annotations: Gene nomenclature', *Hum. Genomics* Vol. 1, pp. 66–71.
2. http://www.ncbi.nlm.nih.gov/BLAST/.
3. http://www.gene.ucl.ac.uk/nomenclature/information/check.shtml.
4. http://drnelson.utmem.edu/cytochromeP450.html.
5. The *C. elegans* protein database. http://www.sanger.ac.uk/Projects/C_elegans/wormpep/.
6. Nebert, D.W. and Russell, D.W. (2002), 'Clinical importance of the cytochromes P450', *Lancet* Vol. 360, pp. 1155–1162.
7. Nelson, D.R., Zeldin, D., Hoffman, S. *et al*. (2004), 'Comparison of cytochrome P450 (*CYP*) genes from the mouse and human genomes including nomenclature recommendations for genes, pseudogenes, and alternative-splice variants', *Pharmacogenetics* Vol. 14, pp. 1–18.
8. Nebert, D.W., Adesnik, M., Coon, M.J. *et al*. (1987), 'The P450 gene superfamily. Recommended nomenclature', *DNA* Vol. 6, pp. 1–11.
9. Nelson, D.R., Koymans, L. and Kamataki, T. *et al*. (1996), 'Cytochrome P450 superfamily: Update on new sequences, gene mapping, accession numbers, and nomenclature', *Pharmacogenetics* Vol. 6, pp. 1–42.
10. http://flybase.bio.indiana.edu/.
11. Maruyama, T. and Furutani, M. (2000), 'Archaeal peptidyl prolyl *cis-trans* isomerases (PPIases)', *Front. Biosci.* Vol. 5, pp. D821–D836.
12. Galat, A. (2003), 'Peptidylprolyl cis/trans isomerases (immunophilins): Biological diversity — targets — functions', *Curr. Top. Med. Chem.* Vol. 3, pp. 1315–1347.
13. He, Z., Li, L. and Luan, S. (2004), 'Immunophilins and parvulins. Superfamily of peptidyl prolyl isomerases in *Arabidopsis*', *Plant Physiol.* Vol. 134, pp. 1248–1267.
14. http://www.ncbi.nlm.nih.gov/LocusLink/.
15. http://www.gene.ucl.ac.uk/nomenclature/.
16. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db = unigene.
17. Osmani, S.A., May, G.S. and Morris, N.R. (1987), 'Regulation of the mRNA levels of *nimA*, a gene required for the G_2-M transition in *Aspergillus nidulans*', *J. Cell. Biol.* Vol. 104, pp. 1495–1504.
18. Osmani, S.A., Pu, R.T. and Morris, N.R. (1988), 'Mitotic induction and maintenance by over-expression of a G_2-specific gene that encodes a potential protein kinase', *Cell* Vol. 53, pp. 237–244.
19. Krien, M.J., West, R.R., John, U.P. *et al*. (2002), 'The fission yeast NIMA kinase Fin1p is required for spindle function and nuclear envelope integrity', *EMBO J.* Vol. 21, pp. 1713–1722.
20. Noguchi, K., Fukazawa, H., Murakami, Y. *et al*. (2002), 'Nek11, a new member of the NIMA family of kinases, involved in DNA replication and genotoxic stress responses', *J. Biol. Chem.* Vol. 277, pp. 39655–39665.
21. Shirane, M. and Nakayama, K.I. (2004), 'Immunophilin FKBP38, an inherent inhibitor of calcineurin, targets BCL2 to mitochondria and inhibits apoptosis', *Nippon Rinsho* Vol. 62, pp. 405–412.
22. Jin, L. and Harrison, S.C. (2002), 'Crystal structure of human calcineurin complexed with cyclosporine-A and human cyclophilin', *Proc. Natl. Acad. Sci. USA* Vol. 99, pp. 13522–13526.
23. Rycyzyn, M.A. and Clevenger, C.V. (2000), 'Role of cyclophilins in somatolactogenic action', *Ann. NY Acad. Sci.* Vol. 917, pp. 514–521.
24. Halestrap, A.P., McStay, G.P. and Clarke, S.J. (2002), 'The permeability transition pore complex: Another view', *Biochimie* Vol. 84, pp. 153–166.
25. Pemberton, T.J., Rulten, S.L. and Kay, J.E. (2003), 'Identification and characterization of *Schizosaccharomyces pombe* cyclophilin-3, a cyclosporin A-insensitive orthologue of human USA-CyP', *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* Vol. 786, pp. 81–91.
26. Vandenbroeck, K., Alloza, I. and Gadina, M. (2004), 'Inhibiting cytokines of the interleukin-12 family: Recent advances and novel challenges', *J. Pharm. Pharmacol.* Vol. 56, pp. 145–160.
27. Chappell, L.H. and Wastling, J.M. (1992), 'Cyclosporin A: Antiparasite drug, modulator of the host-parasite relationship, and immunosuppressant', *Parasitology* Vol. 105, pp. S25–S40.
28. Page, A.P., MacNiven, K. and Hengartner, M.O. (1996), 'Cloning and biochemical characterization of the cyclophilin homologues from the free-living nematode *Caenorhabditis elegans*', *Biochem. J.* Vol. 317, pp. 179–185.
29. Romano, P.G.N., Edvardsson, A., Ruban, A.V. *et al*. (2004), '*Arabidopsis* AtCYP20-2 is a light-regulated cyclophilin-type peptidyl-prolyl *cis-trans* isomerase associated with the photosynthetic membranes', *Plant Physiol.* Vol. 134, pp. 1244–1247.
30. Romano, P.G., Horton, P. and Gray, J.E. (2004), 'The *Arabidopsis* cyclophilin gene family', *Plant Physiol.* Vol. 134, pp. 1268–1282.
31. Reisdorph, N.A. and Small, G.D. (2004), 'The *CPH1* gene of *Chlamydomonas reinhardtii* encodes two forms of cryptochrome whose levels are controlled by light-induced proteolysis', *Plant Physiol.* Vol. 134, pp. 1546–1554.
32. http://www.gene.ucl.ac.uk/nomenclature/genefamily.shtml.
33. Wilkinson, D.M. and Huberman, B.A. (2004), 'A method for finding communities of related genes', *Proc. Natl. Acad. Sci. USA* Vol. 101, pp. S5241–S5248.