Review

Human and orthologous gene nomenclature

Mathew W. Wright *, Elspeth A. Bruford

*HUGO Gene Nomenclature Committee, The Galton Laboratory, Department of Biology, University College London,
Wolfson House, 4 Stephenson Way, London, NW1 2HE, UK*

Abstract

Standardised human gene nomenclature is an essential tool for all researchers, and applying this nomenclature to orthologous genes in other species greatly extends its utility. As more genomes are sequenced and annotated the wealth of information in the public domain concerning homologous genes and their products increases greatly. Here we review many of the resources currently available for identifying orthologs and paralogs, and viewing regions of synteny between genomes, and discuss how these can and are being used in the process of nomenclature assignment. Whilst synchronised gene naming has been actively pursued between human and mouse for many years, and continues apace, it is hoped that this nomenclature will eventually be extended to many other comparative genomes.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Orthology; Paralogy; Synteny; Database

1. Introduction

It has long been accepted that standardised gene nomenclature is vital for effective scientific communication. The provision of a unique symbol in each genome for every gene is essential, since it ensures that everyone, not just geneticists, is able to retrieve and discuss relevant information concerning specific genes. The HUGO Gene Nomenclature Committee (HGNC) fulfill this need by providing unique 'approved' symbols and names for human genes.

The HGNC, formed in the 1970s (Shows et al., 1979), is a committee working under the auspices of the Human Genome Organisation (HUGO). It has grown from a single force, Phyllis J. McAlpine who was largely responsible for setting up the human gene nomenclature committee and devising its original guidelines, to the current team of post-docs and bioinformaticians chaired by Sue Povey. Now based at University College London, the HGNC team maintains the online searchable HGNC Database which contains every approved human gene symbol and name, and related information for each gene including sequence accessions, chromosomal location and links to other databases. Approved symbols must be unique, and are short-form representations of the descriptive gene name which, where possible, is based on the structure, function or homology of the gene. Gene symbols are written in italics and must only contain Latin letters and Arabic numerals. We research each gene individually, through a combination of sequence analysis and database searching, and consult with researchers, and specialist advisors on specific gene families (http://www.gene.ucl.ac.uk/nomenclature/advisors.html). We seek advice from our International Advisory Committee (IAC) on policy issues and work closely with a number of databases to ensure human approved gene nomenclature is widely disseminated.

We also have a close collaboration with the Mouse Genomic Nomenclature Committee (MGNC), based at the Mouse Genome Informatics (MGI) database at The Jackson Laboratory, Maine, the authoritative source of official names for mouse genes, alleles and strains. The HGNC and MGNC

Abbreviations: HUMOT, human and mouse orthologous gene nomenclature; HCOP, HGNC comparison of orthology predictions; *PON*, paraoxonase; *pe*, pale ear phenotype; *le*, light ear phenotype; *HPS*, Hermansky Pudlak Syndrome; *xid*, X-linked immune deficiency; *BTK*, Bruton agammaglobulinemia tyrosine kinase; *REX2*, RNA Exonuclease 2; *REXO2*, REX2 RNA exonuclease 2 homolog (*S. cerevisiae*).

* Corresponding author. Tel.: +44 20 7679 7410; fax: +44 20 7387 3496.
  E-mail address: nome@galton.ucl.ac.uk (M.W. Wright).

communicate on a daily basis via email, and ensure our nomenclature rules and guidelines are, for the most part, identical (Maltais et al., 2002; Wain et al., 2002). We make every effort to reduce interspecies nomenclature confusion where possible by assigning the equivalent gene symbol to orthologous genes and have always co-ordinated closely to achieve this aim. Human and mouse gene symbols are easily distinguished since the letters in human gene symbols are all uppercase, whereas in mouse only the first letter is uppercase and the rest are all lowercase (e.g. paraoxonase 2: human *PON2* and mouse *Pon2*).

2. Ortholog and paralog

Classification of genes as orthologs and paralogs is essential when defining the evolutionary relationship of two genes (Koonin, 2001). Orthologs are genes in different species that derive from a common ancestor, and generally share the same function. Hence, comparative genomics frequently provides an insight into the putative functions of genes in different species.

Paralogs are genes related by duplication from a common ancestor. For example, the chicken *PON2* gene is the ortholog of human *PON2* (Draganov and La Du, 2004), whereas human *PON2* is the paralog of human *PON1* and *PON3* (Fig. 1). Paralogs are not restricted within a genome, and so are either inparalogs, which arise through a gene duplication event after speciation, or outparalogs which arise following a gene duplication before speciation (O'Brien et al., 2005; Remm et al., 2001). Generally inparalog pairs are more likely to have the same function and are less diverged than outparalog pairs.

The utility of standardised orthologous gene names is perhaps one of the strongest arguments for approved nomenclature and cooperation between nomenclature committees. Without this resource the comparison of genomes would be made far more difficult.

3. HUMOT Project

The HGNC and MGNC recently consolidated our collaboration with the initiation of the HUMOT (Human and Mouse Orthologous Gene Nomenclature) Project. The curators from both groups attempt to identify and assign parallel nomenclature to all human and mouse orthologs and paralogs, and resolve situations where orthology assertions may be impossible to assign, e.g. olfactory receptor genes. We also aim to assign parallel nomenclature to human and mouse genes in ortholog pairs that currently have different approved symbols in each species.

When defining an orthologous relationship we examine the percentage identity of the protein and nucleotide sequences and then look for conserved synteny. Despite rearrangement of genomes during evolution, whole blocks of genes have remained located together. Conserved synteny involves two or more genes being grouped in a specific chromosomal region in one species, and also being grouped in an equivalent chromosomal region in another species. These 'syntenic' chromosomal regions simplify mapping across the genomes of different species. For example, syntenic evidence can determine the correct ortholog when there are a number of highly related paralogs at different genomic locations. When orthology is difficult to define we also look at conserved exon–
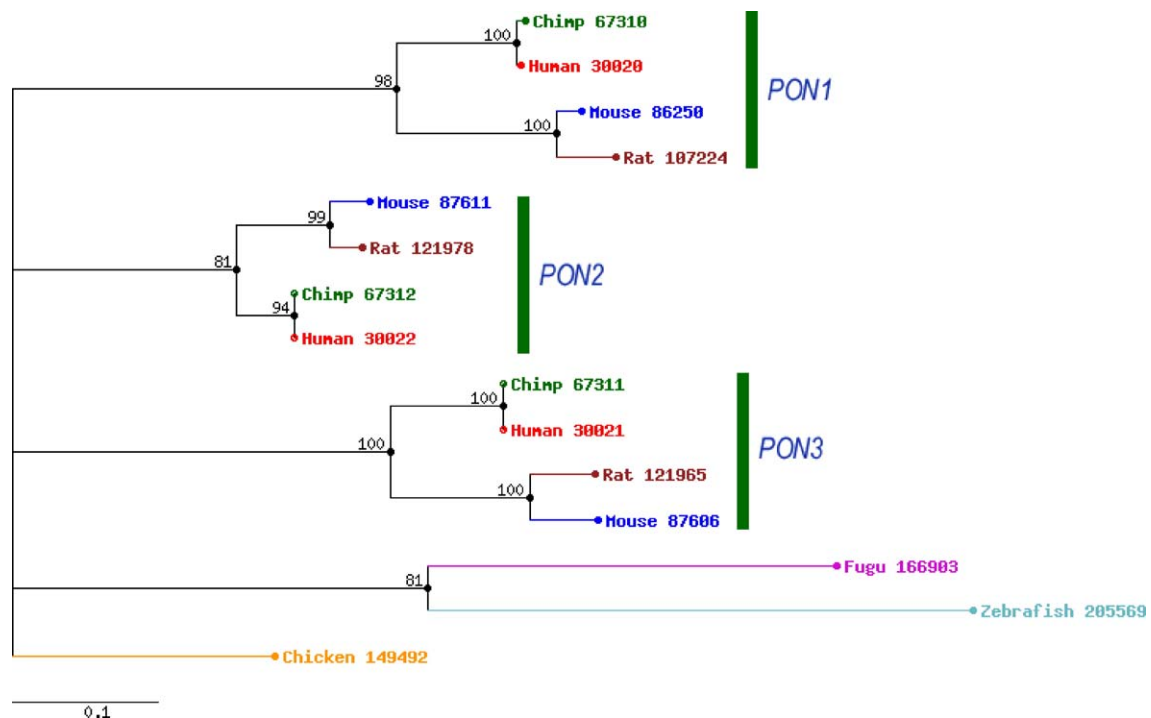

Fig. 1. Partial screenshot from the PhIGS database of a TREE-PUZZLE JTT tree for the PON family (cluster 217014). Figure has been annotated with green bars and gene symbols to indicate the subfamilies.

intron structure, check relevant publications of comparative assays and study phylogenetic relationships. Particular care is exercised in making homology assignments between genes identified as members of large gene families or within specific chromosomal clusters. Ultimately, if orthology is still unclear then we will not make an orthology assertion, since it is better to make no assertion than to disseminate potentially incorrect information in the public databases. We always check our orthology assertions by referring to online resources that contain information relating to mouse and human genes and their orthologous relationships. For instance, ortholog predictions can be found in MGI's Mammalian Orthology Search, NCBI's HomoloGene, the PhIGs database at the Joint Genome Institute (JGI), the 'orthologue predictions' section in Ensembl's GeneView, and in the InParanoid database at the Karolinska Institute.

If a human/mouse ortholog pair has been curated by the HGNC, then the MGI ID for the predicted mouse ortholog is shown in the Symbol Report for the human gene in the HGNC database. Each of these MGI IDs links directly to the gene report page for that gene in the MGI database.

4. Orthology resources

Each 'Gene Detail' page at MGI has a 'Mammalian Orthology' link to the predicted human and rat orthologs. The approved symbols for the human, mouse and rat orthologs are shown, along with the chromosomal location, database IDs that are also links to the gene entry in HGNC, MGI, RGD and Entrez Gene, and the 'orthology criteria' used in defining these orthologous relationships. Orthology criteria include AA for 'Amino acid sequence comparison,' CL for 'Conserved map location,' NT for 'Nucleotide sequence comparison' and FC for 'Functional complementation.' MGI regard some of these criteria as stronger evidence than others, e.g. FC is stronger evidence than NT.

HomoloGene predicts orthologous relationships between the genes of 21 completely sequenced genomes (Wheeler et al., 2005). It begins its orthology analyses by comparing the protein sequences encoded by these genomes using BLASTP and then matching up the sequences using a phylogenetic tree built from sequence similarity; a phylogenetic tree is a branched diagram showing the evolutionary relationships between various species or homologous genes that are thought to have a common ancestor. The protein alignments are then mapped back to their corresponding nucleotide sequences, and are also matched using synteny where possible. Any sequences that are not covered by this analysis are then matched using an algorithm with stringent cutoffs to prevent unrelated genes being grouped as orthologs. Genes that are more homologous within a species than to a gene in another species are predicted as paralogs.

Ensembl (Hubbard et al., 2005) first uses BLASTP+SW (Smith–Waterman algorithm) to identify putative orthologs by reciprocal BLAST analysis. Each best BLAST hit protein pair represents an orthologous gene pair between the two genomes being analysed. Further putative orthologs are found using synteny evidence. All putative orthologs are displayed in the 'orthologue prediction' section in GeneView. The supporting evidence for each prediction is also displayed, e.g. UBRH for '(U)nique (B)est (R)eciprocal (H)it' and RHS for 'Reciprocal Hit based on Synteny around BRH.' In addition, the evidence of DWGA for '(D)erived from (W)hole (G)enome (A)lignment' is given for orthologs between human and chimpanzee, as this is based on a whole genome alignment using BLASTZ.

The Ensembl and Homologene databases sometimes return more than one predicted ortholog, which usually indicates the presence of closely related paralogs. Inparanoid further differentiates paralogs as either inparalogs or outparalogs (O'Brien et al., 2005; Remm et al., 2001), by using the best hits from a BLAST analysis between sequences from two different species. The two best hits are called the 'main orthologs' and form an orthologous group to which other closely related sequences are added and called 'inparalogs.' Another useful paralogy tool is the ParaDB database (Leveugle et al., 2003), which predicts and maps entire duplicated paralogous regions in the human, mouse and fruitfly genomes.

However, not all databases begin their analyses with a BLAST analysis. The Phylogenetically Inferred Groups (PhIGs) database at JGI starts by looking at phylogeny based on the Ensembl dataset (Dehal and Boore, 2005). PhIGs uses known phylogenetic relationships to cluster genes from multiple organisms into groups that are all descended from a single ancestral gene. This is particularly useful in differentiating between orthology versus paralogy relationships (see phylogenetic tree in Fig. 1). For instance, if the ortholog of a gene in one genome has been deleted in another genome the orthology defined by a reciprocal BLAST will be incorrect.

Whilst incorrect orthology assertions can arise because a gene is species-specific, and hence has no equivalent in the other species, they can also arise because the gene in either species could be incorrectly mapped, the gene predicted as the ortholog may actually be a paralog, or (more rarely) they may potentially represent a single transposed gene in either genome. Incorrect orthology assertions may also have been made at a time when genomes were incomplete.

In an attempt to identify such problems, the HGNC now provides a new online orthology tool known as the HGNC Comparison of Orthology Predictions (HCOP) search (Wright et al., 2005). Previously, researchers had to check in each of the orthology prediction sites in order to determine whether an orthology assertion was agreed by all of these databases. However, the HCOP search collates these data and displays the known orthology assertions for each human and mouse gene as predicted by HGNC, MGI, HomoloGene, Ensembl, PhIGs and Inparanoid. Each search returns the official nomenclatures, sequence accessions, database identifiers, aliases, and chromosomal locations for each putative ortholog pair. HCOP facilitates orthology predictions between human and mouse genes. This resource could potentially be expanded to include other mammalian genomes.

In addition to these online resources there is OrthoParaMap (OPM), a freely available suite of programs. The position of the genes within two genomes is used to determine how the genes

were duplicated, and this information is then mapped on to phylogenetic trees and used to infer internal nodes within the trees. OPM is particularly useful in identifying which type of gene duplication may have occurred, e.g. local duplication such as unequal crossing-over, segmental duplication, or speciation (Cannon and Young, 2003). Further programs available which use phylogenetic trees to define orthology include GeneTree (Page, 1998) and RIO (Zmasek and Eddy, 2002).

5. Comparative genome resources

Conservation of synteny occurs throughout mammalian genomes and is a reliable indicator of orthology. Various online databases provide comparative maps or tables where conserved synteny can be observed.

Ensembl (Hubbard et al., 2005) provides a number of tools for visualising genomic information, especially 'MultiContig-View' which can be used for comparing genomic regions across multiple species. Ensembl's ortholog predictions in GeneView link directly to the MultiContigView display. With this tool, the genome sequences from each species are horizontally aligned in relation to each other. The user can zoom in or out depending on the level of detail required, and select as many of the available genomes as desired. Putative orthologs are linked between the alignments with blue lines; this is illustrated for the paraoxonase (PON) family in Fig. 2. Other features, such as alignments to cDNAs and proteins, can also be selected by using the menus.

NCBI's MapViewer allows you to view and search the chromosome maps of a subset of the organisms found in Entrez Genome. Regions of interest can be found via a simple text query (e.g. approved gene symbol or marker name) or by BLAST analysis. The region is viewed vertically and the user can progressively zoom in to see greater detail, even down to the level of genomic sequence for the region. Multiple organisms can be viewed at the same time, with each genome arrayed in parallel, and putative orthologs between each species linked with faint lines. Links are provided to related resources such as Entrez Gene (Wheeler et al., 2005).

MGI illustrate which regions in the human genome are syntenic with regions in the mouse genome with a whole genome orthology map. Human synteny blocks are aligned vertically against the corresponding regions in the mouse chromosomes. Each human region is a link to all of the MGI predicted human/mouse ortholog pairs that lie within that region. Also, linkage maps can be generated which use MGD data to compare a mouse chromosome or region with its syntenic chromosome/region in human, rat, cat, sheep or tammar wallaby. The chromosomal locations and approved symbols (where available) are shown for both species.

The Oxford Grid is particularly useful as a quick visual reference that two chromosomes share a syntenic region, and enables retrieval of information in a format that provides an overview of orthology between the genomes of two species (Edwards, 1991). MGI's Oxford Grid is a matrix of coloured cells, each cell representing two chromosomes, one from each species (chosen from human, mouse, rat, cat, cow, pig and sheep). The chromosomes of one species are arrayed as columns, with the chromosome numbers and species' name at the top and bottom of the grid. The second species' chromosomes are rows with the chromosome numbers and species' name appearing on the left and right sides of the grid. The number inside each cell gives the number of orthology assertions between the two chromosomes, and the colour of each indicates a range in the number of orthologies. Each coloured cell links through to a list of the orthology assertions represented by that cell, with associated information.

Whereas the MGI Oxford Grid provides an immediate overview of synteny between two chromosomes, the ANGIS Oxford Grid allows users to examine specific regions of interest between two chromosomes. The OXGRID project, hosted by the Australian National Genomic Information Service (ANGIS), provides detailed Oxford Grids for human, mouse, rat, cow, pig, sheep, goat, dog and cat, in which the length of the sides of each cell is proportional to the length of these two chromosomes. Where a coloured point lies within each cell indicates its location on the relevant chromosome of both species. This is useful, for instance, in visualising synteny
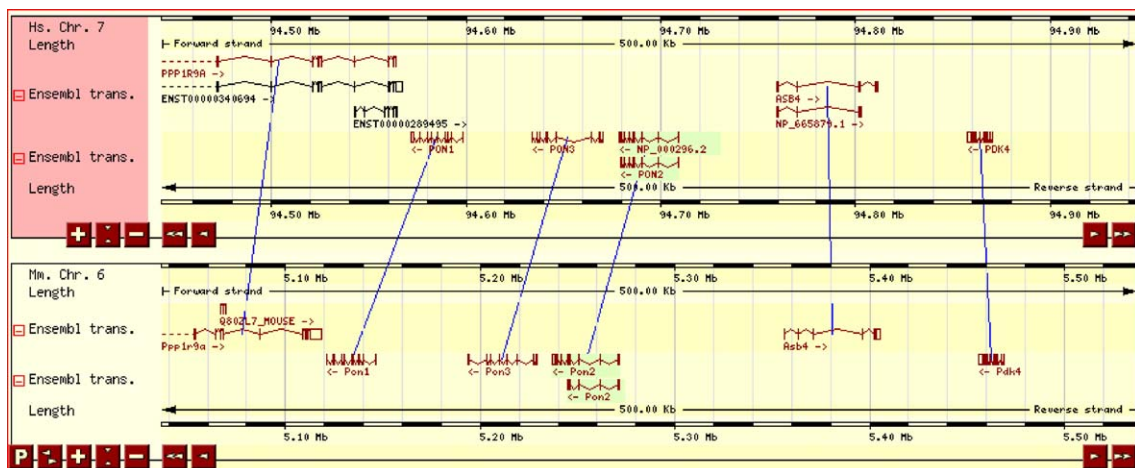

Fig. 2. Partial screenshot of the PON gene cluster as illustrated in MultiContigView in Ensembl. Human and mouse putative orthologs are connected via blue lines.

breakpoints, and the colour of each point indicates the transcriptional polarity.

6. Standardised nomenclature

Once human and mouse orthologous genes have been named in unison, this nomenclature can then proliferate out to other mammalian species as their genomic sequence becomes available. Such resources have already proved invaluable, especially since human disease genes are represented in 99.5% of human–rodent ortholog sets (Huang et al., 2004) and there is a wealth of phenotypic data for the mouse. Some human disease genes have been identified because the mouse ortholog expresses a similar phenotype. For instance, the human ortholog of the gene causing the mouse pale ear (*ep*) phenotype is mutated in a form of Hermansky Pudlak Syndrome (HPS1), a pigmentary disorder (Feng et al., 1997; Gardner et al., 1997). The ortholog of the gene causing the highly similar light ear (*le*) mouse phenotype was hence a candidate gene for another form of HPS, and mutations were subsequently identified in the human ortholog, *HPS4* (Suzuki et al., 2002). Human disease genes have also been used to predict genes responsible for mouse phenotypes. An example is the mouse *Btk* gene, mutations in which are responsible for X-linked immune deficiency (*xid*) in mice. Mutations in the human ortholog of this gene, *BTK*, are associated with X-linked agammaglobulinemia (Tsukada et al., 1993; Vetrie et al., 1993) and the mouse gene was discovered by mapping based on the location of the human gene (Rawlings et al., 1993). The rat is another excellent potential model for identifying the genetic causes of human or mouse phenotypes, since there is a wealth of physiological data for the rat. The Rat Genome Database (RGD) is the essential resource for rat genes and genomic data (Twigger et al., 2005) and they aim to adopt the approved nomenclature of human and mouse genes for each rat ortholog.

Since the human and chimpanzee genomes are so similar, identification of ortholog pairs in these species is much easier than with any other genomes. Indeed, human/chimp orthologs typically differ by only two amino acids, and approximately 29% of all orthologous proteins encoded in human and chimp are identical (The Chimpanzee Sequencing and Analysis Consortium, 2005). It is hence instinctive that researchers will use the approved human gene symbols to represent each chimp ortholog. To date, chimp chromosome 21 (previously designated as chimp chromosome 22) genes have directly adopted the HGNC approved symbol of the human chromosome 21 orthologs (Watanabe et al., 2004), and we envisage that approved human nomenclature will be used throughout the rest of the chimp genome, where applicable. Likewise ARKdb, a database which provides genome mapping data for several commercial species, including cat, cow, deer, horse, pig and sheep (Hu et al., 2001), already makes considerable efforts to name orthologous genes with the approved HGNC symbol.

When the function of a gene is unknown and it encodes no known intrinsic domain, then the human gene is sometimes named after its ortholog in lower eukaryotes. This is especially true for the model organism genomes that have their own genomic databases, such as ZFIN (Zebrafish Information Network) for zebrafish (Sprague et al., 2003), Flybase for *Drosophila* (Drysdale and Crosby, 2005), SGD for *S. cerevisiae* (Christie et al., 2004) and wormbase for *C. elegans* (Harris et al., 2004). Sometimes the exact approved symbol used by these databases needs to be altered in order to be unique in the HGNC database, e.g. the ortholog of yeast *REX2* for 'RNA Exonuclease 2' in SGD is approved as *REXO2* for human 'REX2, RNA exonuclease 2 homolog (*S. cerevisiae*)' in the HGNC database.

Standardised nomenclature provides an essential resource not just for human researchers but for all eukaryotic researchers, especially as more and more genomes are rapidly being sequenced and annotated. Comparative genomics is an essential constituent of the post-genomic era and a standardised nomenclature across species will ensure that comparative geneticists are all speaking the same language.

Further information can be found on the HGNC website (http://www.gene.ucl.ac.uk/nomenclature). You can also contact us directly if you have any orthology or nomenclature queries at nome@galton.ucl.ac.uk.

URLs

ANGIS Oxford Grid, http://oxgrid.angis.org.au/.

ARKdb, http://www.thearkdb.org/.

Ensembl, http://www.ensembl.org.

Flybase, http://flybase.bio.indiana.edu/.

HGNC, http://www.gene.ucl.ac.uk/nomenclature/.

HGNC HCOP, http://www.gene.ucl.ac.uk/cgi-bin/nomenclature/hcop.pl.

HGNC HUMOT, http://www.gene.ucl.ac.uk/nomenclature/ activities/humot/.

HUGO, http://www.hugo-international.org/.

Inparanoid, http://inparanoid.cgb.ki.se/.

NCBI Entrez Gene, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene.

NCBI HomoloGene, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi? DB=HomoloGene.

NCBI MapViewer, http://www.ncbi.nlm.nih.gov/mapview/.

MGI, http://www.informatics.jax.org/.

MGI Comparative Maps, http://www.informatics.jax.org/searches/ linkmap_form.shtml.

MGI Human/Mouse Orthology Map, http://www.informatics.jax.org/ reports/homologymap/mouse_human.shtml.

MGI Mammalian Orthology Query, http://www.informatics.jax.org/ searches/homology_form.shtml.

MGI Oxford Grid, http://www.informatics.jax.org/searches/ oxfordgrid_form.shtml.

MGNC, http://www.informatics.jax.org/mgihome//nomen/.

ParaDB, http://abi.marseille.inserm.fr/paradb/.

PhIGs, http://phigs.jgi-psf.org/.

RGD, http://rgd.mcw.edu/.
SGD, http://www.yeastgenome.org/.
UCSC, http://genome.ucsc.edu/.
Wormbase, http://www.wormbase.org/.
ZFIN, http://www.shigen.nig.ac.jp:6070/.

References

Cannon, S.B., Young, N.D., 2003. OrthoParaMap: distinguishing orthologs from paralogs by integrating comparative genome data and gene phylogenies. BMC Bioinformatics 4, 35.

Chimpanzee Sequencing and Analysis Consortium, 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437, 69–87.

Christie, K.R., et al., 2004. Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. Nucleic Acids Res. 32, D311–D314.

Dehal, P., Boore, J.L., 2005. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. 3, e314.

Draganov, D.I., La Du, B.N., 2004. Pharmacogenetics of paraoxonases: a brief review. Naunyn-Schmiedeberg's Arch. Pharmacol. 369, 78–88.

Drysdale, R.A., Crosby, M.A., 2005. FlyBase: genes and gene models. Nucleic Acids Res. 33, D390–D395.

Edwards, J.H., 1991. The Oxford grid. Ann. Hum. Genet. 55 (Pt 1), 17–31.

Feng, G.H., Bailin, T., Oh, J., Spritz, R.A., 1997. Mouse pale ear (ep) is homologous to human Hermansky–Pudlak syndrome and contains a rare 'AT–AC' intron. Hum. Mol. Genet. 6, 793–797.

Gardner, J.M., et al., 1997. The mouse pale ear (ep) mutation is the homologue of human Hermansky–Pudlak syndrome. Proc. Natl. Acad. Sci. U. S. A. 94, 9238–9243.

Harris, T.W., et al., 2004. WormBase: a multi-species resource for nematode biology and genomics. Nucleic Acids Res. 32, D411–D417.

Hu, J., et al., 2001. The ARKdb: genome databases for farmed and other animals. Nucleic Acids Res. 29, 106–110.

Huang, H., et al., 2004. Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes. Genome Biol. 5, R47.

Hubbard, T., et al., 2005. Ensembl 2005. Nucleic Acids Res. 33, D447–D453.

Koonin, E.V., 2001. An apology for orthologs — or brave new memes. Genome Biol. 2 (COMMENT1005).

Leveugle, M., Prat, K., Perrier, N., Birnbaum, D., Coulier, F., 2003. ParaDB: a tool for paralogy mapping in vertebrate genomes. Nucleic Acids Res. 31, 63–67.

Maltais, L.J., Blake, J.A., Chu, T., Lutz, C.M., Eppig, J.T., Jackson, I., 2002. Rules and guidelines for mouse gene, allele, and mutation nomenclature: a condensed version. Genomics 79, 471–474.

O'Brien, K.P., Remm, M., Sonnhammer, E.L., 2005. Inparanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res. 33, D476–D480.

Page, R.D., 1998. GeneTree: comparing gene and species phylogenies using reconciled trees. Bioinformatics 14, 819–820.

Rawlings, D.J., et al., 1993. Mutation of unique region of Bruton's tyrosine kinase in immunodeficient XID mice. Science 261, 358–361.

Remm, M., Storm, C.E., Sonnhammer, E.L., 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J. Mol. Biol. 314, 1041–1052.

Shows, T.B., et al., 1979. International system for human gene nomenclature (1979) ISGN (1979). Cytogenet Cell Genet. 25, 96–116.

Sprague, J., et al., 2003. The Zebrafish Information Network (ZFIN): the zebrafish model organism database. Nucleic Acids Res. 31, 241–243.

Suzuki, T., et al., 2002. Hermansky–Pudlak syndrome is caused by mutations in HPS4, the human homolog of the mouse light-ear gene. Nat. Genet. 30, 321–324.

Tsukada, S., et al., 1993. Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia. Cell 72, 279–290.

Twigger, S.N., et al., 2005. Tools and strategies for physiological genomics — the rat genome database. Physiol. Genomics 23, 246–256.

Vetrie, D., et al., 1993. The gene involved in X-linked agammaglobulinaemia is a member of the src family of protein–tyrosine kinases. Nature 361, 226–233.

Wain, H.M., Bruford, E.A., Lovering, R.C., Lush, M.J., Wright, M.W., Povey, S., 2002. Guidelines for human gene nomenclature. Genomics 79, 464–470.

Watanabe, H., et al., 2004. DNA sequence and comparative analysis of chimpanzee chromosome 22. Nature 429, 382–388.

Wheeler, D.L., et al., 2005. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 33, D39–D45.

Wright, M.W., Eyre, T.A., Lush, M.J., Povey, S., Bruford, E.A., 2005. HCOP: The HGNC Comparison of Orthology Predictions Search Tool. Mamm Genome 16, 827–828.

Zmasek, C.M., Eddy, S.R., 2002. RIO: analyzing proteomes by automated phylogenomics using resampled inference of orthologs. BMC Bioinformatics 3, 14.