Orthology and Functional Conservation in Eukaryotes

Kara Dolinski and David Botstein

Department of Molecular Biology, Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544;
email: kara@genomics.princeton.edu, botstein@genomics.princeton.edu

Annu. Rev. Genet. 2007. 41:465–507

First published online as a Review in Advance on August 3, 2007

The *Annual Review of Genetics* is online at http://genet.annualreviews.org

This article's doi:
10.1146/annurev.genet.40.110405.090439

0066-4197/07/1201-0465$20.00

Key Words

comparative genomics, protein evolution, interspecies complementation

Abstract

In recent years, it has become clear that all of the organisms on the Earth are related to each other in ways that can be documented by molecular sequence comparison. In this review, we focus on the evolutionary relationships among the proteins of the eukaryotes, especially those that allow inference of function from one species to another. Data and illustrations are derived from specific comparison of eight species: *Homo sapiens*, *Mus musculus*, *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Danio rerio*, *Saccharomyces cerevisiae*, and *Plasmodium falciparum*.

INTRODUCTION

When the Human Genome Project was first proposed, it was controversial, in no small part because it was unclear, even to many scientists, how knowledge of the human genome sequence would advance understanding of biology and medicine. Twenty years later, there is no longer any question that the advent of the era of genomic sequences has provided insights and tools that have completely transformed all of the biomedical sciences. The top-level insight, of course, is that all the organisms on Earth are related to each other by descent in ways that can now readily be documented by molecular sequence comparisons. From the point of view of the working biologist, the most important evolutionary relationships are those among the amino acid sequences of proteins. Although other sequences (notably those of conserved RNAs, large and small, and many kinds of regulatory sequences found in genomic DNA) are also important, the bulk of biological functions are carried out by proteins. In this review we focus exclusively on the evolutionary relationships among the proteins of eukaryotes, especially those that allow inference of function from one species to another.

Biologists are interested in two kinds of organism. We study the human organism (*Homo sapiens*) largely out of self-interest: In the past century it has become manifest that out of biological understanding of ourselves as biological entities comes the power to ameliorate, and sometimes even to eliminate, our diseases. Most of the funding for biological research today is provided in the expectation that what will be learned will ultimately be useful in dealing with human disease; this expectation has largely been met. In addition, the practice of medicine involves much study of individual people, both when they are healthy and when they are not. Doctors obtain highly detailed information both from objective measurements and from the self-reported observations of their patients, a source not available for any experimental organism. However, experiments with humans are often impractical and/or unethical. Thus for experimentation the medical community has turned to various surrogate species, the most widely studied of which is the mouse (*Mus musculus*). From the point of view of protein sequence evolution, the mouse and the human are very similar, as we shall see.

EIGHT EUKARYOTIC SPECIES COMPARED

The genome sequences have allowed, as indicated above, inference about functions, especially of proteins, based on the evolutionary relatedness. Since even mice are not among the most tractable experimental species, most of what is known about the basic biological functions of proteins in eukaryotes derives not from the study of human or mouse genes and proteins directly, but has been inferred from studies with other simpler eukaryotes, collectively referred to as model organisms. The model organisms all have in common small size, a short life cycle, a small genome, ease of manipulation and analysis by genetic and biochemical means, and, not least, a substantial community of researchers devoted to the study of their basic biology. For this reason, we focus in this review on the protein sequences encoded in the genomes of five of them: a yeast (*Saccharomyces cerevisiae*), a nematode worm (*Caenorhabditis elegans*), a fruitfly (*Drosophila melanogaster*), a flowering plant (*Arabidopsis thaliana*), and a vertebrate fish (*Danio rero*). They are the leading experimental organisms for modern biologists, and among them span much of the evolutionary tree of the eukaryotes. Also of particular interest from the point of view of health is the malaria parasite *Plasmodium falciparum*, an organism that is also very difficult to study in the laboratory and which, although it is a eukaryote, has a relatively exotic parasitic lifestyle. Thus, in this review, we focus on a total of eight eukaryotic species: the five model organisms, mouse, human, and the malaria parasite.

Table 1 Public resources that provide comparative genomics data and tools

Name	Description	URL
Clusters of Orthologous Groups (COG) (44)	Provides groups of orthologous proteins for seven eukaryotic species; the COG construction protocol involves manual curation	**www.ncbi.nlm.nih.gov/COG/**
Eukaryotic Gene Orthologs (EGO) (27)	Displays predicted orthologs derived from several eukaryotic genomes based on gene alignments	**www.tigr.org/tdb/tgi/ego/ego.shtml**
Homologene (48)	Provides automated predictions of homologs among the genes of several eukaryotes	**www.ncbi.nlm.nih.gov/HomoloGene/**
Inparanoid (32)	Houses pair-wise groups of orthologous proteins for multiple species	**http://inparanoid.cgb.ki.se/**
OrthoDisease (33)	Uses the Inparanoid algorithm to generate pair-wise orthologs between human disease genes and genes from other species	**http://orthodisease.cgb.ki.se**
OrthoMCL-DB (5, 28)	Utilizes a Markov Cluster algorithm to predict orthologous groups of proteins for multiple species simultaneously	**http://orthomcl.cbil.upenn.edu/**
Sybil (S. Angiuoli & O. White, personal communication)	Uses Jaccard clustering to group sequences based on pair-wise BLAST analysis	**http://Sybil.sourceforge.net/**
YOGY (36)	Retrieves orthologous proteins from four different resources: COG, Inparanoid, Homologene, and OrthoMCL-DB	**www.sanger.ac.uk/PostGenomics/S_pombe/YOGY/**
P-POD (12)	Orthologs and Jaccard clusters	**http://ortholog.princeton.edu/**

Many thousands of proteins, representing many hundreds of homologous families, are encoded in each of these genomes. In this review we cannot enumerate even a tiny fraction of the evolutionary and functional relationships that might be of interest. So the discussion here is necessarily limited to just a few illustrative examples. For more detail, and for information on evolutionary and functional relationships among proteins and species not mentioned here, the reader is referred to the public databases (**Table 1**). Specifically, we refer to P-POD, a database maintained at Princeton [**http://ortholog.princeton.edu**; (12)], which contains the functionalities and curated information (for the same eight species) that we believe are particularly useful to those interested in the conserved functions of eukaryotic proteins. In this review, we have extracted our figures, tables, and examples from P-POD, in the way we imagine readers might when they become interested in a particular function or gene family.

RETENTION OF BIOLOGICAL FUNCTIONS BY ORTHOLOGS

Two protein sequences in different species are orthologous if their amino acid sequences remain similar to each other (homologous) because they originated from a common ancestor, having been separated in evolutionary time only by speciation event(s). When orthology is unambiguous, the two proteins in the two different species generally (but not necessarily) retain the same function, the more so when the function is important or essential to evolutionary fitness. For instance, where there is a single actin gene in each of two species, they are orthologous and likely to have extremely similar functions. It is the relative confidence with which we can infer something about biological function when

P-POD: Princeton Protein Orthology Database

Ortholog: proteins from different species that are homologous and have evolved from a common ancestor by speciation; often they retain the same biological function

Homologs: proteins (or genes) that are related based on sequence similarity. Homologs may be either orthologs or paralogs

Paralogs: proteins that are homologous due to gene duplication after a speciation event; often they evolve new biological functions

orthology is established that makes it such a valuable tool for the working biologist. By knowing the function of a protein in an experimentally tractable organism, one can generally get a very good idea of the function in orthologous proteins in other species, even ones quite distant from the model organism in which experiments were done. There is, of course, no absolute certainty about this inference: Exceptions exist, and ambiguities about which sequence is the true ortholog among several similar homologs in a species are common, as we shall see.

In contrast, paralogs are the result of duplications that happen after speciation event(s). Although paralogous proteins often retain similar biochemical functions, even when the ancestor protein's function is essential, they are free to diverge after the duplication, and generally do so. For example, the several kinds of human hemoglobins are paralogs of each other and also paralogs of myoglobin, as all of these proteins are derived from a common ancestor's heme-carrier protein. It is not always possible to decide which of the paralogs is the "true" ortholog to the ancestor protein. In the globin example, all retain the heme-carrier function(s), and in the eons of duplication and divergence no one of them may have retained all of the ancestor's essential function(s); it sometimes turns out that these are shared among the more specialized paralogous descendants.

Just about anyone who studies yeast, worms, or flies in their research has written (at least in grant proposals) that studying the basic biology of model organisms is likely to be relevant to human health. In the past decade, the availability of complete protein sets from the most commonly studied eukaryotes has provided a way to identify these orthologs on a global scale and has shown that not just the individual proteins, but how those proteins work in concert within the context of larger biological processes are shared among all eukaryotes (6, 23, 40, 47). Even processes that at first seem specific to "higher" eukaryotes have analogous pathways in yeast. A particularly impressive example is the striking similarity between neuronal signaling in mammals and vesicular transport along the secretory pathway in yeast [for excellent reviews, see (34, 49)]. What began as independent parallel efforts in humans (driven by the desire to understand how neurotransmitters are delivered to synapses) and yeast (driven by the desire to understand intracellular trafficking of proteins) turned into a single field of science.

Of course, efficient and accurate identification of orthologs is key to enabling valid information transfer from experimentally tractable systems to the less tractable ones. As described in detail below, P-POD includes a system for finding and evaluating orthologies among the genomes of our eight eukaryotic species. The basic output of the OrthoMCL option is our best approximation of the set of orthologs (see below) and is shown as an arbitrarily rooted similarity tree.

MAMMALIAN ORTHOLOGS OF YEAST PROTEINS CAN PROVIDE MISSING FUNCTION(S) TO YEAST MUTANTS

As indicated above, there is an expectation, but no guarantee, that orthologous proteins will retain the same function over arbitrarily long evolutionary distances. This expectation, very important for inference of human gene functions in particular, has been strongly buttressed by experiments that show directly that in many instances a mammalian protein can indeed functionally substitute for the lack of the corresponding protein in one of the model organisms. The first such "complementation" experiment, published in 1985 (16), showed that expression in *S. cerevisiae* of a mammalian RAS protein (one of several paralogs) results in restoration of viability to a double-mutant strain defective in both of the paralogous yeast *RAS* genes. **Figure 1*a*** shows the P-POD output for the query "*ras1*" with the OrthoMCL option (see below for OrthoMCL description). It is easy to see that the two yeast genes

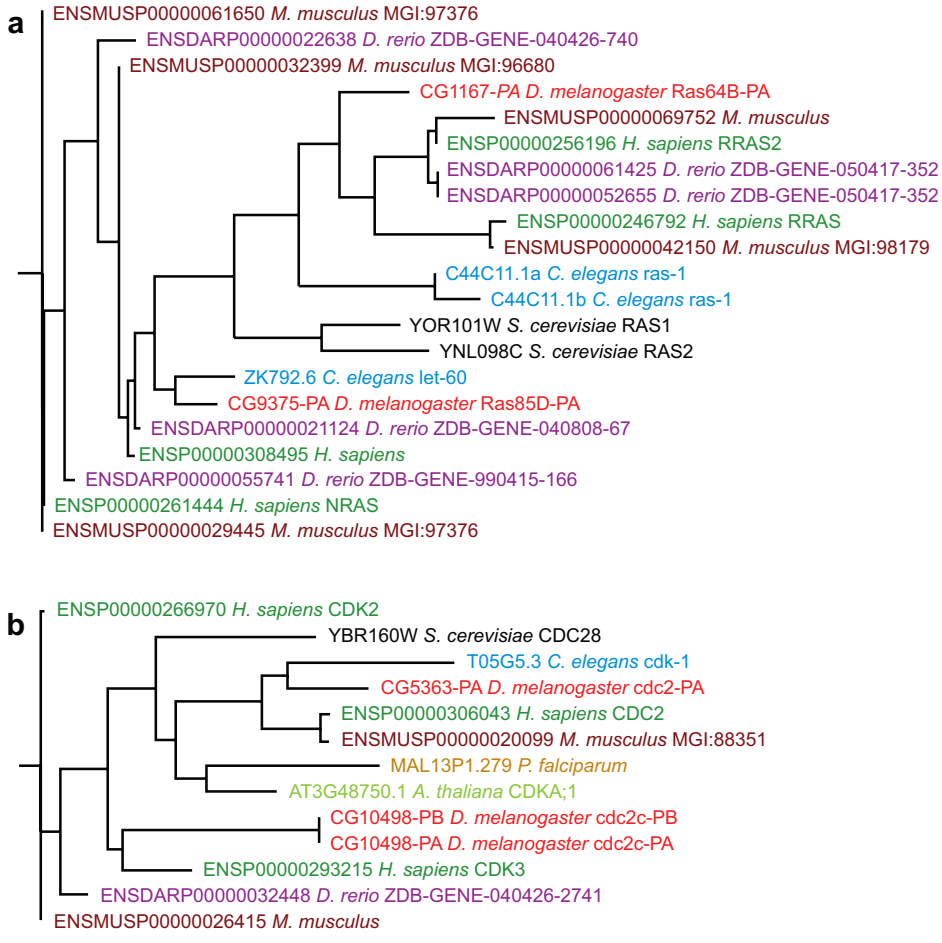


Figure 1

Conservation of ras and Cdc28 proteins. Parts (*a*) and (*b*): ortholog families of ras and Cdc28, respectively, generated by OrthoMCL as described.

RAS1 and *RAS2* are recently duplicated paralogs, as is the fact that the human genome contains several paralog relatives of the yeast *RAS* genes. It is interesting, although not surprising, that the malaria parasite, distant both evolutionarily and functionally from the other seven species, has no sequence similar enough to satisfy the algorithm. It is also worth noting, in this and subsequent figures, the clustering together of the mammals, mouse, and human, relative to the other eukaryotes.

Another early example that has had great influence was the cloning, by complementation, of a fission yeast (*Schizosaccharomyces pombe*) *cdc2* mutant (defective in the catalytic subunit of the primary cell-cycle–dependent protein kinase) of the human ortholog [called *CDK1*; (26)] as well as the budding yeast ortholog *CDC28* (2). **Figure 1*b*** shows the P-POD for the query "*cdc28*" with the OrthoMCL option. Here there are fewer paralogs overall (the human CDK1 sequence that was recovered in fission yeast is marked "CDC2"). As might be expected from the nature of the function (master regulator of the cell cycle), every one of the eight divergent eukaryotic species has an ortholog.

There are today very many published instances of this kind of functional substitution. **Table 2** illustrates a small subset of curated instances from P-POD in which a well-established human disease gene itself (and not a paralog), when expressed in yeast, successfully complements a mutation abolishing

Table 2 Complementation of yeast mutants by their orthologous human genes associated with disease phenotypes

Yeast gene[a]	Human gene (HGNC)	OMIM phenotypes[b]	SGD description
ALG1	ALG1	Congenital disorder of glycosylation, Type Ik	Mannosyltransferase, involved in asparagine-linked glycosylation in the ER
IRA1	NF1	Neurofibromatosis, Watson syndrome, myelomonocytic leukemia	GTPase-activating protein that negatively regulates RAS
TRS20	ZNF547	Spondyloepiphyseal dysplasia tarda, X-linked	One of 10 subunits of the TRAPP complex of the *cis*-Golgi which mediates vesicle docking and fusion
GLK1	GCK	Maturity-onset diabetes of the young, hyperinsulinemic hypoglycemia, PNDM diabetes mellitus, permanent neonatal	Glucokinase, catalyzes the phosphorylation of glucose at C6 in the first irreversible step of glucose metabolism
ERS1	CTNS	Cystinosis, late-onset juvenile or adolescent nephropathic type	Protein with similarity to human cystinosin, which is a H^+-driven transporter involved in L-cystine export from lysosomes
YFH1	FXN	Friedreich ataxia 1	Frataxin, regulates mitochondrial iron accumulation
CCC2	ATP7A	Cutis laxa, X-linked, Menkes disease	Cu^{+2}-transporting P-type ATPase
RAD3		Trichothiodystrophy, photosensitive, erythroderma with hair abnormality and mental and growth retardation	5′ to 3′ DNA helicase, subunit of RNA polymerase II transcription initiation factor TFIIH and Nucleotide Excision Repair Factor 3 (NEF3)
LPD1		Leigh syndrome; LS Leigh syndrome due to mitochondrial complex I deficiency	Dihydrolipoamide dehydrogenase, the lipoamide dehydrogenase component (E3) of the pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase multienzyme complexes
HXK1	GCK	Diabetes mellitus, permanent neonatal, hyperinsulinemic hypoglycemia, familial	Hexokinase isoenzyme 1, a cytosolic protein that catalyzes phosphorylation of glucose during glucose metabolism
MET13		Homocystinuria due to deficiency of N(5,10)-methylenetetrahydrofolate reductase activity MTHFR deficiency	Isozyme of methylenetetrahydrofolate reductase, catalyzes the reduction of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate in the methionine biosynthesis pathway
PMR1	ATP2C1	Benign chronic pemphigus; BCPM	High affinity Ca^{2+}/Mn^{2+} P-type ATPase required for Ca^{2+} and Mn^{2+} transport into Golgi; involved in Ca^{2+} dependent protein sorting and processing
PFK1	PFKM	Phosphofructokinase, red cell	Alpha subunit of heterooctameric phosphofructokinase involved in glycolysis
FOL2	GCH1	GTP cyclohydrolase I deficiency, dystonia, progressive, with diurnal variation	GTP-cyclohydrolase I, catalyzes the first step in the folic acid biosynthetic pathway

Table 2 (*Continued*)

Yeast gene[a]	Human gene (HGNC)	OMIM phenotypes[b]	SGD description
PUT2	ALDH4A1	Hyperprolinemia, Type II	Delta-1-pyrroline-5-carboxylate dehydrogenase, nuclear-encoded mitochondrial protein involved in utilization of proline as sole nitrogen source
DNA2	WRN	Werner syndrome; WRN	Essential tripartite DNA replication factor with single-stranded DNA-dependent ATPase, ATP-dependent nuclease, and helicase activities; required for Okazaki fragment processing; involved in DNA repair pathways; potential Cdc28p substrate
YHC3	CLN3	Ceroid lipofuscinosis, neuronal, 3 (Batten disease)	Vacuolar membrane protein involved in the ATP-dependent transport of arginine into the vacuole and possibly in balancing ion homeostasis
SOD1	SOD1	Amyotrophic lateral sclerosis 1; ALS1	Cytosolic superoxide dismutase
STE24	ZMPSTE24	Mandibuloacral dysplasia with Type B lipodystrophy; MADB	Highly conserved zinc metalloprotease that functions in two steps of a-factor maturation, C-terminal CAAX proteolysis and the first step of N-terminal proteolytic processing
TIM8	TIMM8A	Opticoacoustic nerve atrophy with dementia	Mitochondrial intermembrane space protein mediating import and insertion of polytopic inner membrane proteins
TRZ1	ELAC2	Prostate cancer	tRNase Z, involved in RNA processing, has two putative nucleotide triphosphate-binding motifs (P-loop) and a conserved histidine motif
UNG1	UNG	Immunodeficiency with hyper-IgM, Type 5	Uracil-DNA glycosylase, required for repair of uracil in DNA formed by spontaneous cytosine deamination, not required for strand-specific mismatch repair, cell-cycle regulated, expressed in late G1, localizes to mitochondria and nucleus
SGS1	BLM	Bloom syndrome; BLM	Nucleolar DNA helicase of the RecQ family involved in maintenance of genome integrity, regulates chromosome synapsis and meiotic crossing over
PFK2	PFKM	Phosphofructokinase, red cell	Beta subunit of heterooctameric phosphofructokinase involved in glycolysis, indispensable for anaerobic growth, activated by fructose-2,6-bisphosphate and AMP, mutation inhibits glucose induction of cell cycle-related genes
RAS2	HRAS	Costello syndrome, bladder cancer, thyroid carcinoma, follicular; FTC	GTP-binding protein that regulates the nitrogen starvation response, sporulation, and filamentous growth; farnesylation and palmitoylation required for activity and localization to plasma membrane

(*Continued*)

Table 2 (*Continued*)

Yeast gene[a]	Human gene (HGNC)	OMIM phenotypes[b]	SGD description
ALG9	ALG9	Congenital disorder of glycosylation, Type Il	Mannosyltransferase, involved in N-linked glycosylation; catalyzes the transfer of mannose from Dol-P-Man to lipid-linked oligosaccharides
ALG12	ALG12	Congenital disorder of glycosylation, Type Ig	Alpha-1,6-mannosyltransferase localized to the ER; responsible for the addition of the alpha-1,6 mannose to dolichol-linked Man7GlcNAc2, acts in the dolichol pathway for N-glycosylation
COQ2	COQ2	Coenzyme Q10 deficiency	Para hydroxybenzoate: polyprenyl transferase, catalyzes the second step in ubiquinone (coenzyme Q) biosynthesis
GSH2	GSS	Glutathione synthetase deficiency	Glutathione synthetase, catalyzes the ATP-dependent synthesis of glutathione (GSH) from gamma-glutamylcysteine and glycine; induced by oxidative stress and heat shock
IRA2	NF1	Neurofibromatosis, familial spinal, Watson syndrome, juvenile myelomonocytic leukemia; JMML	GTPase-activating protein that negatively regulates RAS by converting it from the GTP- to the GDP-bound inactive form, required for reducing cAMP levels under nutrient limiting conditions
RAS1	HRAS	Costello syndrome, bladder cancer, thyroid carcinoma, follicular; FTC	GTPase involved in G-protein signaling in the adenylate cyclase activating pathway, plays a role in cell proliferation; localized to the plasma membrane
RAD53	CHEK2	Li-Fraumeni syndrome 2; LFS2, osteogenic sarcoma, prostate cancer, breast cancer	Protein kinase, required for cell-cycle arrest in response to DNA damage

[a]This table contains instances where a human disease gene complemented the corresponding yeast gene mutation, and that the disease gene itself (not a paralog) is the one used in the experiment. All cases of experimentally tested orthologs and their citations are available from the P-POD database.
[b]We considered a human gene related to a disease if it is associated with an OMIM phenotype in the OMIM database.

the function of its yeast ortholog. The table gives also the brief description of the function from the point of human biology as given by the OMIM (Online Mendelian Inheritance in Man) database and from the point of yeast as found in SGD (Saccharomyces Genome Database).

COMPUTATIONAL METHODS TO IDENTIFY ORTHOLOGS

BLAST: basic local alignment search tool

Two basic schemes underlie the many individual computational approaches generally used to identify orthologs: phylogenetic analysis and all-against-all BLASTp reciprocal best hits. The former is the more direct and rigorous approach, where homologs are grouped and aligned, and phylogenetic trees are generated based on these alignments and subsequently examined to identify the orthologs. Several substantial studies have successfully used this approach [for example, (6, 38, 54)]. The disadvantage of strictly phylogenetic approaches is that they are computationally expensive and generally also seem to require manual intervention.

The all-against-all reciprocal BLASTp approach, where each sequence from one species is used as a BLAST query against the genome of another and vice versa, requires much less computation, and lends itself to full automation. Ortholog pairs are identified as reciprocal best hits, meaning that two sequences in different organisms have higher BLAST (basic local alignment search tool) scores with each other than with any other sequence in the other genome. Several very useful ortholog databases, such as the Clusters of Orthologous Groups (COG) and TIGR Orthologous Gene Alignments (TOGA), have been generated in this way (27, 44). In general they appear to be efficient in finding orthologs.

Nevertheless, the reciprocal BLAST method (and the assumptions that underlie it) is a compromise. The most serious problem is the large number of false positives that are produced. For example, in a study that identified orthologous pairs of transmembrane proteins in *C. elegans* and *H. sapiens*, Remm & Sonnhammer (38) compared results from a rigorous phylogenetic approach to results generated from all-against-all BLAST. The phylogenetic method yielded 174 high-confidence orthologs; the simple two-way BLAST identified 168 of them. However, the BLAST method generated 34 additional ortholog pairs. When the authors examined these more closely, 17 were found to be false positives clearly traceable to problems with the BLAST method, whereas the remaining 17 were unable to be resolved, and thus could be real orthologs not found by the phylogenetic method (or may be due to another problem with the two-way BLAST).

The false-positive problem is most pronounced in eukaryotic genomes, where gene duplication and functional redundancy is commonplace. When duplication occurs after speciation, the duplicated gene often still retains the function of the ortholog; these homologs are referred to as "recent" or in-paralogs. "Ancient" paralogs (or out-paralogs) are homologs that arose from duplication events before speciation and are thus likely to be functionally diverged. The simple two-way BLAST method often cannot distinguish between these types of paralogs and includes them inappropriately in orthologous groups, thus leading to many false positives. Similarity between multidomain proteins that are not functionally related but share common domains can also contribute to false positives in all-versus-all BLAST methods.

Out-paralog algorithms have been devised that aim to deal with at least some of the false-positives generated by all-by-all BLAST methods. For example, the Inparanoid and OrthoMCL algorithms identify orthologous groups such that in-paralogs are retained while out-paralogs are excluded (28, 39). The first step in the Inparanoid method is an all-versus-all BLAST to find all possible pairwise similarities. To avoid problems with multidomain proteins, an overlap cutoff of 50% of the entire sequence length is enforced. Ortholog groups are generated by first seeding the group with a reciprocal best hit ortholog pair (the "main ortholog" pair), then building the group by adding in-paralogs and excluding out-paralogs, as appropriate. In-paralogs are clustered around the main ortholog from each species independently, with the criteria used for adding the in-paralogs being that the main ortholog is more similar to in-paralogs from the same species than to any other sequence in the other species. Additional rules are then are applied to merge, delete, or separate overlapping groups.

OrthoMCL similarly distinguishes between recent and ancient paralogs by using BLAST both within and between species, but requires that recent paralogs are included in an orthologous group only when they are more similar to each other than to any other sequence from other species. The main difference between OrthoMCL and Inparanoid is that while Inparanoid generates pairwise orthologous families, OrthoMCL generates orthologous families that contain sequences from more than two species. This is accomplished by using the Markov CLuster (MCL)

COG: Clusters of Orthologous Groups

TOGA: TIGR Orthologous Gene Alignments

In-paralogs: homologs that arise when duplication occurs after speciation, and the duplicated gene often still retains the function of the ortholog

Out-paralogs: homologs that arise from duplication events before speciation and are thus likely to be functionally diverged

MCL: Markov CLuster algorithm

algorithm, which was originally utilized for graph clustering using flow simulation (7). The MCL algorithm takes as input a similarity matrix, which is based on a graph where sequences are nodes and edges are relationships between them; all of the relationships in the matrix are considered simultaneously during clustering, which allows for distinction of recent versus ancient paralogs and can avoid the problem caused by similarity shared by multidomain proteins that are mistakenly identified as orthologs by other two-way BLAST methods. Recently, considerations of synteny have been added to the programs to further improve upon the basic BLAST approach [see for example, (19, 55)]. Several of these ortholog-finding methods have been implemented to create useful resources for researchers, and **Table 1** includes Web resources that provide access to one or more instantiations of each of these algorithms.

P-POD offers two views of the relationships among proteins. One of these uses the OrthoMCL algorithm to produce orthologous groups from the eight organisms listed above. The other uses a Jaccard Clustering algorithm (also based on graph theory) originated at TIGR (S. Angiuoli & O. White, personal communication). This algorithm uses the reciprocal BLAST input to produce groups of very highly related proteins, but not necessarily orthologs, by grouping proteins into the same family if they share a signficant number of homologs (12). Being based on sequence similarity without direct reference to phylogenetic lineage, it allows users to see the orthologous groups provided by OrthoMCL in the fuller context of larger gene families. For instance, in P-POD the OrthoMCL option, when queried with an alpha-tubulin sequence (for example, yeast *tub1*), will return, for each of the eight eukaryotic species, only alpha-tubulins (the true orthologs and inparalogs; **Figure 2**), whereas the JACCARD option will return the larger tubulin family, including the out-paralogous beta-tubulins and gamma-tubulins (not shown).

INTERSPECIES COMPLEMENTERS AND ESSENTIAL FUNCTIONS ARE ENRICHED AMONG ORTHOLOGOUS GROUPS FOUND BY ORTHOMCL

Since deletions are now available for every gene in the yeast genome, it is possible to study very many potential orthologs for complementing activity. All one needs is a selectable phenotype in yeast; for proteins essential in yeast, this phenotype is simply viability as in the historical examples described above. With 643 experimentally confirmed orthologs that were collected from the literature, we were able to assess the accuracy of the OrthoMCL algorithm as implemented in P-POD. OrthoMCL generated 445 orthologous groups that agreed with the experimental result. In the remaining cases, OrthoMCL either left out an ortholog that should have been present in a group (153 false negatives) or included an ortholog that should not have been present (45 false positives) (12).

Since deletions of every yeast gene have been assayed for viability, we were able to examine whether essential yeast genes are more likely to be conserved than nonessential ones. We found that indeed essential yeast genes are more likely to be found in an OrthoMCL family than not ($\chi^2 = 78$, $p = 1.1 \times 10^{-18}$). We also examined whether essential genes in yeast are more likely to be orthologous to human disease genes. Here we found no statistically significant enrichment for essentiality among disease-related yeast genes. Surprising at first, this result can easily be rationalized by the realization that genes essential to yeast may in fact be lethal in human cells as well, and thus be inconsistent with the full development required to make even a diseased mammal.

CONSERVATION OF PROTEIN NETWORKS

Proteins do not function alone. They function as parts of pathways, macromolecular structures, and regulatory networks. In order

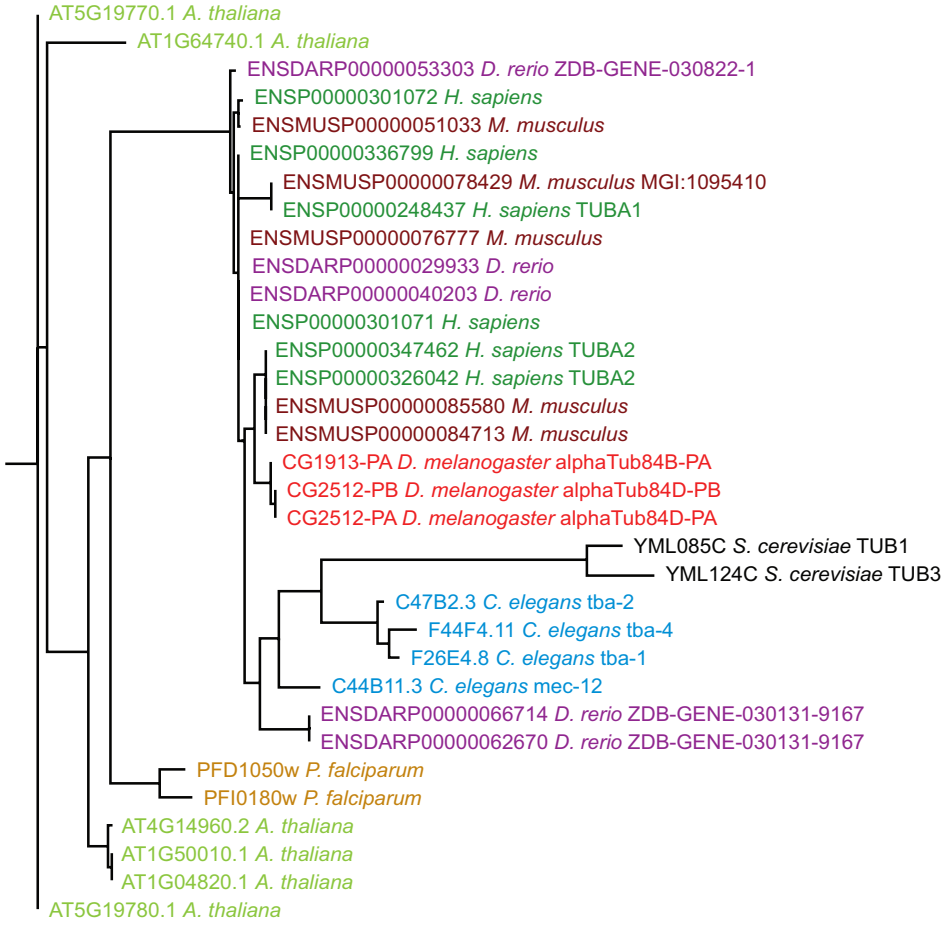


Figure 2

Conservation of the alpha tubulins. Ortholog family generated by OrthoMCL.

for their biological function to be conserved, they must retain not only their elemental activity (e.g., enzyme or ligand binding) but also the ability to interact with other proteins. Thus one might expect to observe conservation at the level of protein complexes and pathways. Nowhere is this more clear than in the case of an interspecies complementation experiment: If an ortholog from a distantly related species has retained a biological function in the context of the foreign species, it must retain at least minimal ability to carry out the interactions required.

PathBLAST is an algorithm that addresses this by aligning protein networks, where high scoring alignments are subnetworks that contain the same set of proteins that also share sequence similarity (17, 18). Reguly et al. (37) applied this method to examine protein complexes that yeast and flies have in common. Over 1400 conserved complexes were identified in this manner. Other methods that compare protein networks across species have also been developed [for example, see (14, 20, 21, 29)].

For reasons similar to the reasons why highly conserved proteins are more likely to be essential, hubs in a protein interaction network, i.e., proteins that interact with many other proteins, are expected to be more likely essential (1, 15, 37) and conserved (15, 51, 52). Wuchty and colleagues explored the correlation between conservation and connectivity in detail (53). They started with a yeast

physical interaction network and tagged orthologous proteins using the Inparanoid algorithm; this resulted in a set of about 2000 yeast proteins with orthologs in *H. sapiens*, *A. thaliana*, *C. elegans*, *M. musculus*, and *D. melanogaster*. Next, proteins were grouped based on their connectivity, and evolutionary retention was determined by counting the fraction of orthologs in each group. They found that most connected proteins were indeed more conserved. The authors also determined the fraction of interacting protein pairs that are conserved (i.e., both interacting partners have orthologs) in groups based on the local topology of the network; they found that highly clustered interactions are also more conserved.

Sharan and colleagues (42) extended the PathBLAST algorithm so that networks from multiple species can be directly aligned with each other. They compared the protein interaction networks of yeast, worm, and fly, and found many highly conserved network regions. A representative example is the conservation of the proteasome complex, specifically the physical interactions among subunits that comprise the lid of the 19S regulatory particle. The lid of the 19S regulatory particle recognizes the polyubiquitinated substrate, cleaves the polyubiquitin into monomers, opens the pore, and denatures the protein substrate. **Figure 3** illustrates the conservation at the sequence and network level, and that, not surprisingly, much more interaction information is available from yeast, where several high-throughput interaction assays have been applied. Rri1 is included in this subnetwork. Rri1 is the metalloendopeptidase subunit of the COP9 signalosome complex and is not thought to be part of the proteasome. However, it shares significant sequence homology with Rpn11, the corresponding metalloendopeptidase of the proteasome. In two high-throughput mass spectrometry studies (10, 22), Rri1 was reported to interact with Rpn5, which actually forms a complex with Rpn11; one explanation for this result is that there was misidentification of Rri1 based on its amino acid similarity to Rpn11. The rest of highly conserved COP9 signalosome complex shares significant similarity to the proteasome complex, suggesting a common evolutionary origin of these two important protein complexes [reviewed in (8)].

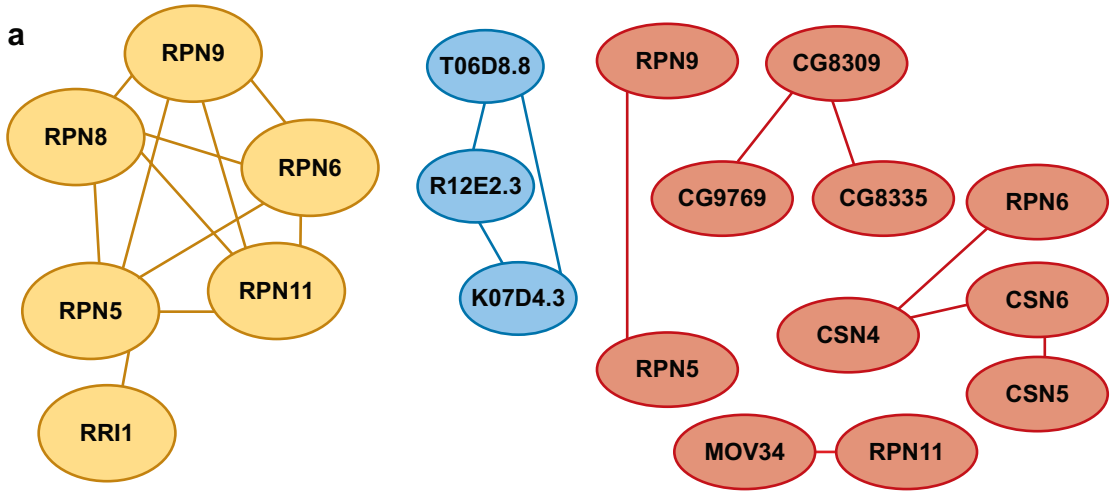
USING ORTHOLOGY TO LEARN ABOUT BIOLOGY: ILLUSTRATIVE EXAMPLES

Inference of function from studies of protein orthology can provide insights in many different ways. In each of the examples below, information from a heterologous species provided the basis for experiments that led to important advances in understanding.

Mismatch Repair and Colon Cancer

In 1993, a set of papers appeared showing that inherited nonpolyposis colon cancers have a cellular phenotype: instability of short repeated sequences in the tumor cells (25, 30, 35). Stimulated by this result, and even before the human genes had been cloned, yeast researchers isolated mutations in yeast genes with the same phenotype (including mutations in *MSH2* and *MLH1*), predicting that the colon cancer gene(s) were likely to be at their homologs (43).

Figure 4*a* and ***b*** shows the orthologs of the yeast *MLH1* and *MSH2* proteins using OrthoMCL algorithm in P-POD. The *MLH1* and *MSH2* proteins correspond in sequence and activity to the the biochemically well-characterized mutL and MutS DNA mismatch repair proteins of *E. coli*, respectively. Each of the eight species, including the human, has an *MSH2* and *MLH1* ortholog. Using the DNA sequence of these two yeast genes, it was possible to clone the human orthologs, which indeed turned out to be genes whose inheritance predisposes to colon cancer [(25, 43; and others, reviewed in 11)]. **Figure 4*c*** shows the Jaccard clustering output for the query MSH2. In this family there are five human paralogs. Each one of these is



b

Description	Yeast	Worm	Fly
Proteasome subunit of the 19S regulatory lid	RPN9	T06D8.8	CG8309, RPN9
Essential proteasome subunit of the 19S regulatory lid	RPN8	R12E2.3	CG8335, MOV34, CSN6, CG9769
Essential proteasome subunit of the 19S regulatory lid	RPN6		RPN6
Metalloprotease proteasome subunit of the 19S regulatory lid	RPN11	K07D4.3	RPN11, CSN5
Essential proteasome subunit of the 19S regulatory lid	RPN5		RPN5, CSN4
Metalloendopeptidase subunit of the COP9 signalosome complex	RRI1	K07D4.3	RPN11, CSN5

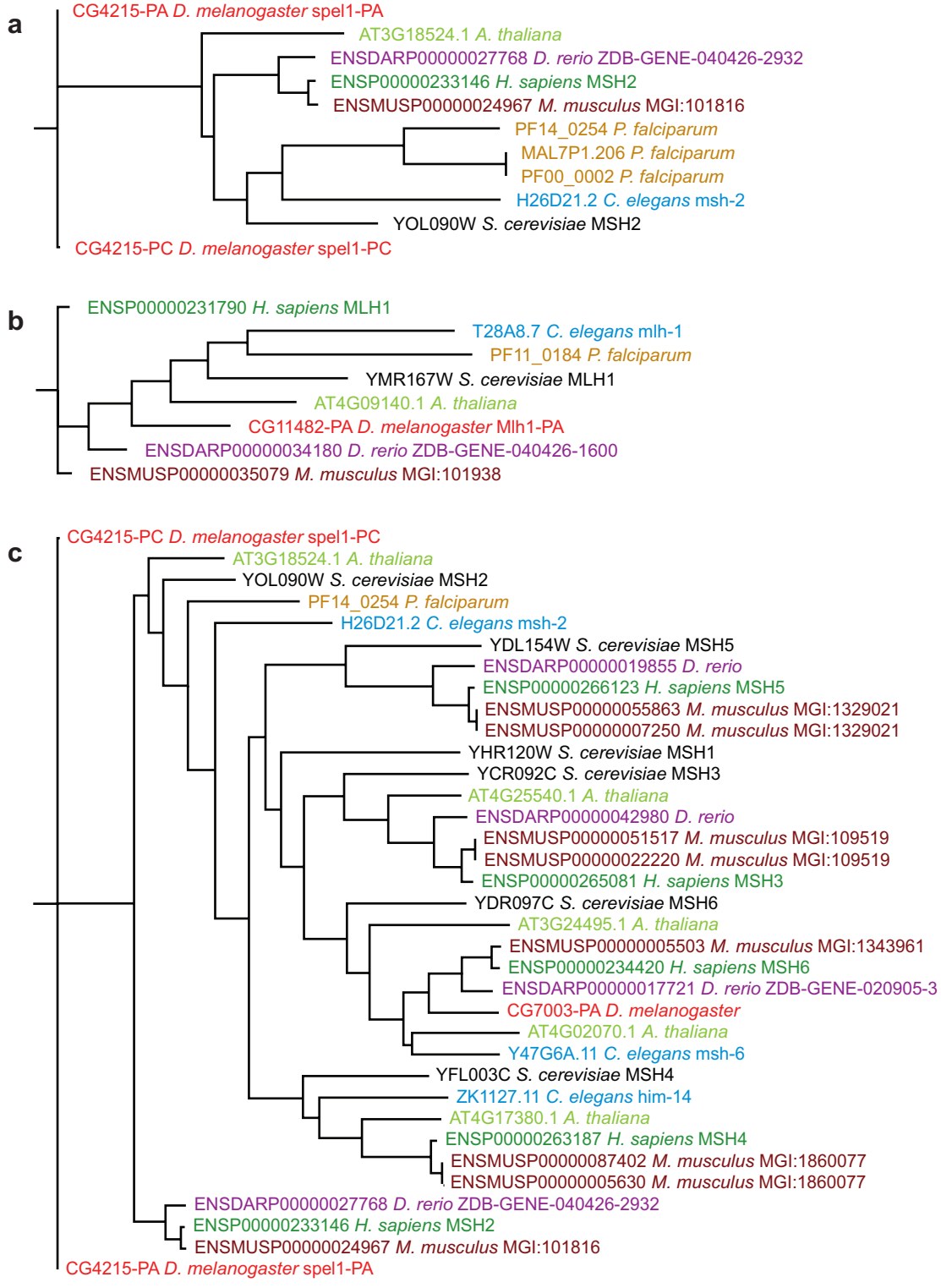
Figure 3

Conserved proteins of the proteasome: the lid components of the 19S regulatory particle of the 26S proteasome [adapted from **Figure 2*j*** from (42)]. (*a*) Nodes are proteins and edges are interactions between them. Yellow nodes are yeast proteins, blue are worm, and red are fly. (*b*) Conservation of the proteins. Each row shows the equivalent subunits from each species.

implicated in OMIM as associated with cancer, although not nearly as frequently as the human orthologs of yeast *MSH2* and *MLH1*.

Role of AMP-Activated Kinases

It is by no means always the case that information essential to understanding biology travels one way, from the model systems to the human. Often the information from each organism supplements what is found in the other, to make a unitary understanding applicable to both. For example, one of the central regulators of metabolism was identified in yeast on the basis of its effects on carbon catabolism, and then shown, in 1986, to be a protein kinase



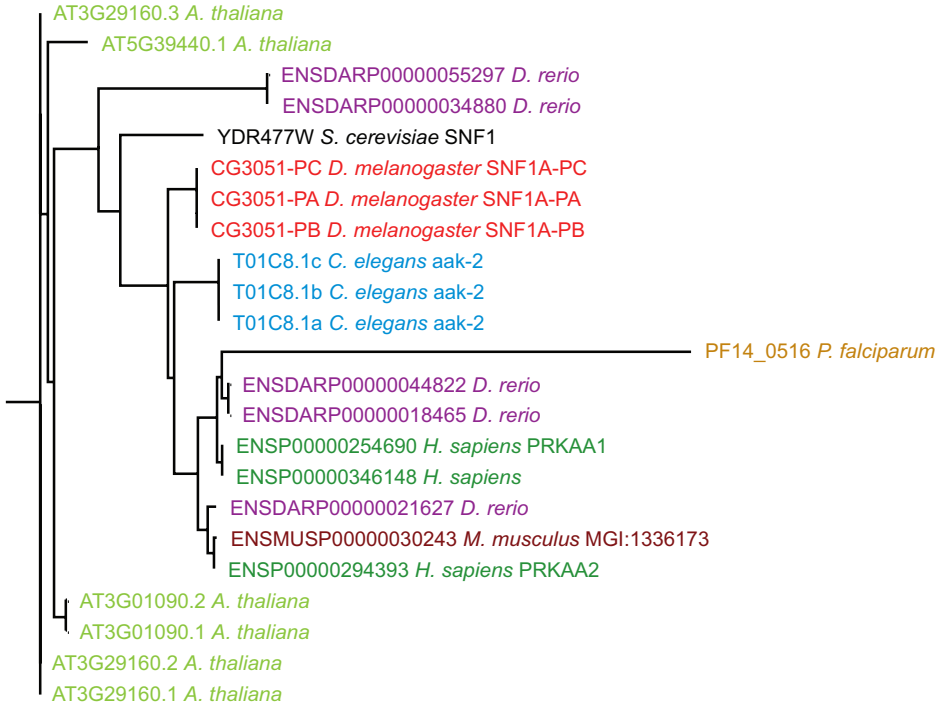


Figure 5

Conservation of Snf1/AMPK protein kinases. Ortholog family generated by OrthoMCL.

by its autophosphorylation activity (4). This kinase turned out to be strongly conserved and has orthologs in all the other eukaryotes (**Figure 5**).

As the yeast ortholog was characterized further, it became known that the human Snf1/AMP-activated kinase (AMPK) family is involved in metabolism and stress response [reviewed in (41)], and that AMPK effectively conserves ATP when AMP levels are high. Because of the sequence conservation between yeast Snf1 and rat AMPK, in particular in the protein kinase domains, Woods and colleagues showed that Snf1 and AMPK are functionally interchangeable (by showing yeast Snf1 can specifically phosphorylate a known mammalian AMPK target) and that the biological roles and mechanism of action between the two orthologs are remarkably conserved (50).

The kinases that activate Snf1 are also conserved. In yeast, Sak1, Tos3, and Elm1 are the Snf1-activating kinases, and there are corresponding orthologs in mammals. One such mammalian AMPK-activating kinase, LKB1, which was identified by sequence similarity (13), is involved in Peutz-Jeghers cancer-susceptibility syndrome. The gene encoding the third AMPK-activating kinase to be identified in mammals, TAK1, was cloned by screening a mammalian cDNA library for genes that complement the *snf1* growth phenotype of a *sak1 tos3 elm1* triple mutant (31). TAK1 would not have been discovered through sequence similarity because identity to other AMPK-activating kinases is not

Figure 4

Conservation of MutS and MutL proteins. Parts (*a*) and (*b*): ortholog families of MSH2 (MutS) and MLH1 (MutL), respectively, generated by OrthoMCL; part (*c*): larger MSH2 protein family generated by Jaccard Clustering.

AMPK: AMP-activated kinase

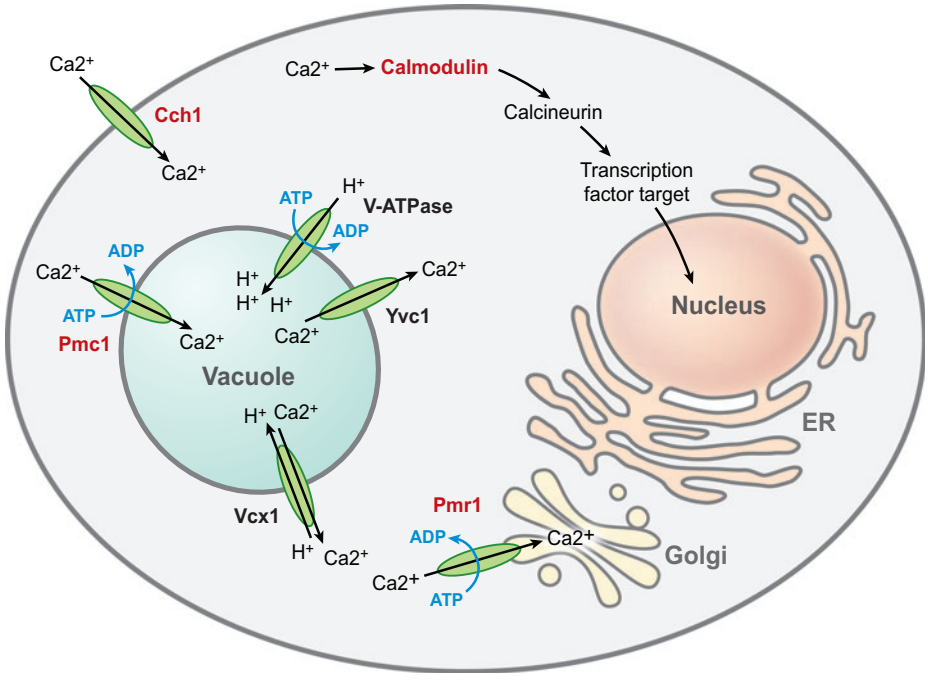


Figure 6

Calcium signaling and homeostasis in yeast. Each yeast protein has a mammalian ortholog. Genes for which direct functional complementation from another species has been exhibited are indicated with text outlined in red [adapted from **Figure 1** in (46)].

significant enough for it to be uncovered through a BLAST search (31).

Calcium Homeostasis and Signaling

Calcium homeostasis and signaling, critical for several processes in the eukaryotic cell, is yet another highly conserved pathway (**Figure 6**). All of the major players in calcium signaling are conserved from yeast to mammals. This fact has been leveraged to better characterize several calcium signaling pathways in mammals, from T cell activation (9) to Hailey-Hailey disease (HHD) (45). What we have learned about this disease from yeast is reviewed in detail in Reference 46; we briefly summarize here. A breakthrough in the study of HHD at the molecular level occurred when mutations in patients with the disease were mapped to ATP2C1/hSPCA1, which was found to share high sequence similarity (49% identity) with yeast Pmr1, a Ca^{2+}/Mn^{2+} P-type ATPase on the Golgi membrane. The phenotypes of the *pmr1* null mutant in yeast are very similar to those observed in patients' keratinocytes, and the human ATP2C1 gene can complement the calcium and manganese transport defects of the yeast *pmr1* mutant. This assay for functional complementation has provided an easy method for detailed structure-function analysis by screening site-directed mutations for the ability to restore wild-type function in the yeast mutant.

HHD: Hailey-Hailey disease

ORTHOLOGOUS INTERACTING PROTEINS CARRY OUT THE COMMON CELL BIOLOGY OF EUKARYOTES

Table 3 is a compilation of groups of orthologous proteins whose yeast protein member has been shown to have more than 20 interactions; they are listed alphabetically by yeast gene. From this table, one can see quickly that the basic common cell biology of eukaryotes is dominated by proteins whose structure, function, and interactions have been conserved over most, if not always all, the eukaryotes. The examples used here are illustrative, and appear early in the sorted list: *CDC28*

Table 3 Conservation of highly interacting proteins in yeast

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
ACT1	15	91 (36, 55)	88.9	Cytokinesis: actin	–
ADA2	97	49 (11, 38)	30.2	Protein modification process: transcription coactivator	–
AGP1	192	37 (1, 36)	–	Transport: low-affinity amino acid permease with broad substrate range	–
ALG1	169	39 (4, 35)	35.1	Protein modification process: mannosyltransferase	*H. sapiens/ S. cerevisiae*
ALG5	270	31 (8, 23)	40	Protein modification process: UDP-glucose:dolichyl-phosphate glucosyltransferase	–
APN1	217	36 (35, 1)	–	DNA metabolic process: major apurinic/apyrimidinic endonuclease	*C. elegans/ S. cerevisiae*
ARF1	93	50 (31, 19)	74.1	Vesicle-mediated transport: ADP-ribosylation factor	*H. sapiens/ S. cerevisiae*
ARP4	255	32 (3, 29)	–	DNA metabolic process: nuclear actin-related protein involved in chromatin remodeling	–
ASF1	50	62 (41, 21)	59.9	Protein modification process: Nucleosome assembly factor	–
BCK1	202	37 (33, 4)	41	Signal transduction: mitogen-activated protein (MAP) kinase kinase kinase acting in the protein kinase C signaling pathway	*P. carinii/S. cerevisiae*
BCK2	338	23 (21, 2)	–	Cell cycle: protein rich in serine and threonine residues involved in protein kinase C signaling pathway	–
BEM1	46	64 (24, 40)	29.2	Conjugation: Protein containing SH3-domains	
BEM2	315	27 (25, 2)	43.4	Cell wall organization and biogenesis: Rho GTPase activating protein (RhoGAP) involved in the control of cytoskeleton organization and cellular morphogenesis	–
BET1	105	48 (21, 27)	24.2	Vesicle-mediated transport: Type II membrane protein required for vesicular transport between the endoplasmic reticulum and Golgi complex	–
BMH1	240	34 (21, 13)	66.6	Pseudohyphal growth: 14-3-3 protein	*T. gondii/S. cerevisiae*

(*Continued*)

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
BNI1	94	50 (32, 18)	33.1	Pseudohyphal growth: formin	*A. thaliana/ S. cerevisiae*
BOS1	252	33 (12, 21)	–	Vesicle-mediated transport: v-SNARE (vesicle specific SNAP receptor)	–
BRR2	128	44 (2, 42)	36.7	RNA metabolic process: RNA-dependent ATPase RNA helicase involved in the facilitation and disruption of snRNA interactions	–
BSD2	115	46 (8, 38)	–	Transport: heavy metal ion homeostasis protein	–
BUB2	166	40 (34, 6)	22.8	Cell cycle: mitotic exit network regulator	–
CAN1	19	80 (33, 47)	–	Transport: plasma membrane arginine permease	–
CBF2	54	61 (33, 28)	–	Process unknown: essential kinetochore protein	–
CCR4	62	58 (37, 21)	36.3	RNA metabolic process: component of the CCR4-NOT transcriptional complex	–
CDC11	305	28 (7, 21)	–	Cytokinesis: component of the septin ring of the mother-bud neck that is required for cytokinesis	–
CDC12	152	41 (23, 18)	41	Cytokinesis: component of the septin ring of the mother-bud neck that is required for cytokinesis	–
CDC13	87	51 (36, 15)	–	Organelle organization and biogenesis: single-stranded DNA-binding protein found at TG1-3 telomere G-tails	–
CDC14	100	49 (25, 24)	32.7	Protein modification process: protein phosphatase required for mitotic exit	*C. albicans/ S. cerevisiae*
CDC15	121	46 (36, 10)	–	Cytokinesis: Protein kinase of the Mitotic Exit Network that is localized to the spindle pole bodies at late anaphase	–
CDC2	113	47 (42, 5)	49.7	DNA metabolic process: catalytic subunit of DNA polymerase delta	–
CDC20	131	44 (17, 27)	–	Cell cycle: cell-cycle–regulated activator of anaphase-promoting complex/cyclosome (APC/C)	–

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
CDC23	236	34 (13, 21)	37.5	Cell cycle: subunit of the anaphase-promoting complex/cyclosome (APC/C)	–
CDC24	295	29 (21, 8)	24.8	Pseudohyphal growth: guanine nucleotide exchange factor (GEF or GDP-release factor) for Cdc42p	–
CDC28	4	124 (57, 67)	60.2	Transcription: catalytic subunit of the main cell cycle cyclin–dependent kinase	*O. tauri/S. cerevisiae*
CDC34	64	58 (43, 15)	35.5	Cell cycle: ubiquitin-conjugating enzyme or E2	–
CDC42	20	79 (52, 27)	80.1	Pseudohyphal growth: small rho-like GTPase	*D. melanogaster/ S. cerevisiae*
CDC47	196	37 (9, 28)	49.1	DNA metabolic process: component of the hexameric MCM complex	–
CDC48	212	36 (9, 27)	68.7	Cell cycle: ATPase in ER	–
CDC5	33	70 (48, 22)	44.3	Protein modification process: Polo-like kinase with similarity to *Xenopus* Plx1 and *S. pombe* Plo1p	*H. sapiens/ S. cerevisiae*
CDC53	117	46 (19, 27)	40.2	Cell cycle: cullin	*A. thaliana/ S. cerevisiae*
CDC54	321	25 (3, 22)	50.6	DNA metabolic process: essential helicase component of heterohexameric MCM2-7 complexes which bind prereplication complexes on DNA and melt the DNA prior to replication	–
CDC55	329	24 (21, 3)	68.3	Translation: nonessential regulatory subunit B of protein phosphatase 2A	–
CDC6	146	42 (27, 15)	26.1	DNA metabolic process: ATP-binding protein required for DNA replication	–
CHC1	61	58 (35, 23)	49.5	Vesicle-mediated transport: clathrin heavy chain	–
CHS7	327	24 (2, 22)	–	Vesicle-mediated transport: protein of unknown function	–
CIN8	300	29 (28, 1)	–	Cell cycle: kinesin motor protein	–
CLA4	44	66 (43, 23)	58.3	Cytokinesis: Cdc42p activated signal transducing kinase of the PAK (p21-activated kinase) family	–

(*Continued*)

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
CLB2	30	71 (36, 35)	34.7	Cell cycle: B-type cyclin involved in cell cycle progression	–
CLB5	190	38 (28, 10)	31.4	Cell cycle: B-type cyclin involved in DNA replication during S phase	–
CLF1	76	55 (15, 40)	–	Cell cycle: essential splicesome assembly factor	*D. melanogaster/ S. cerevisiae*
CLN1	200	37 (28, 9)	–	Cell cycle: G1 cyclin involved in regulation of the cell cycle	*M. sativa/S. cerevisiae*
CLN2	56	60 (45, 15)	–	Conjugation: G1 cyclin involved in regulation of the cell cycle	*M. sativa/S. cerevisiae*
CLN3	122	46 (39, 7)	–	Organelle organization and biogenesis: G1 cyclin involved in cell cycle progression	*M. sativa/S. cerevisiae*
CNB1	336	23 (21, 2)	58.9	Cell wall organization and biogenesis: calcineurin B	–
COG1	296	29 (21, 8)	–	Vesicle-mediated transport: essential component of the conserved oligomeric Golgi complex (Cog1p through Cog8p)	–
CRM1	70	56 (5, 51)	46.7	Ribosome biogenesis and assembly: major karyopherin	–
CSE2	328	24 (3, 21)	–	Transcription: component of the Med9/10 module	–
CSG2	43	66 (24, 42)	–	Process unknown: endoplasmic reticulum membrane protein	–
CTF13	275	31 (22, 9)	–	Organelle organization and biogenesis: subunit of the CBF3 complex	–
CTK1	134	44 (29, 15)	44.2	Protein modification process: catalytic (alpha) subunit of C-terminal domain kinase I (CTDK-I)	–
CYC8	188	38 (17, 21)	35.3	Transcription: general transcriptional corepressor	–
DAM1	126	45 (24, 21)	–	Cell cycle: essential subunit of the Dam1 complex (aka DASH complex)	–
DBF2	22	77 (18, 59)	–	Protein modification process: Ser/Thr kinase involved in transcription and stress response	–
DCP1	238	34 (13, 21)	–	RNA metabolic process: subunit of the Dcp1p-Dcp2p decapping enzyme complex	*S. pombe/S. cerevisiae*

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
DMA1	316	26 (5, 21)	–	Cytoskeleton organization and biogenesis: protein involved in regulating spindle position and orientation	–
DOA1	165	40 (30, 10)	41.3	Protein modification process: WD repeat protein required for ubiquitin-mediated protein degradation	–
DOA4	247	34 (29, 5)	–	Vesicle-mediated transport: ubiquitin hydrolase	–
DST1	141	43 (34, 9)	32.4	Meiosis: general transcription elongation factor TFIIS	–
DUN1	311	27 (22, 5)	–	Protein modification process: cell-cycle checkpoint serine-threonine kinase required for DNA damage-induced transcription of certain target genes	–
ELG1	245	34 (27, 7)	–	DNA metabolic process: protein required for S phase progression and telomere homeostasis	–
ELM1	312	27 (23, 4)	–	Pseudohyphal growth: serine/threonine protein kinase that regulates cellular morphogenesis	*A. thaliana/ S. cerevisiae*
ELO1	221	35 (3, 32)	29.9	Lipid metabolic process: elongase I	*D. melanogaster/ S. cerevisiae*
EMP24	90	50 (2, 48)	32.8	Organelle organization and biogenesis: integral membrane component of endoplasmic reticulum-derived COPII-coated vesicles	–
EMP47	301	28 (1, 27)	–	Vesicle-mediated transport: integral membrane component of endoplasmic reticulum-derived COPII-coated vesicles	–
ERD1	136	43 (1, 42)	–	Protein modification process: predicted membrane protein required for the retention of lumenal endoplasmic reticulum proteins	–
ERG25	74	55 (3, 52)	33.4	Lipid metabolic process: C-4 methyl sterol oxidase	–
ERG6	285	30 (25, 5)	–	Lipid metabolic process: Delta(24)-sterol C-methyltransferase	–

(*Continued*)

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
ERV29	142	42 (1, 41)	35.8	Vesicle-mediated transport: protein localized to COPII-coated vesicles	–
EXO1	337	23 (21, 2)	40.7	DNA metabolic process: 5′-3′ exonuclease and flap-endonuclease involved in recombination	–
FEN1	24	76 (9, 67)	28.1	Vesicle-mediated transport: fatty acid elongase	–
FET3	194	37 (7, 30)	–	Process unknown: ferro-O2-oxidoreductase required for high-affinity iron uptake and involved in mediating resistance to copper ion toxicity	*S. cerevisiae/ M. musculus*
FKS1	250	34 (32, 2)	–	Cell wall organization and biogenesis: 1,3-beta-D-glucan synthase subunit	*A. thaliana/ S. cerevisiae*
GAL11	80	54 (28, 26)	–	Transcription: component of the Mediator complex	–
GAL4	144	42 (4, 38)	–	Carbohydrate metabolic process: DNA-binding transcription factor required for the activation of the GAL genes in response to galactose	–
GAS1	201	37 (29, 8)	–	Cell wall organization and biogenesis: Beta-1,3-glucanosyltransferase	*A. fumigatus/ S. cerevisiae*
GCN4	107	47 (11, 36)	–	Amino acid and derivative metabolic process: transcriptional activator of amino acid biosynthetic genes in response to amino acid starvation	–
GCN5	3	126 (54, 72)	46.5	Protein modification process: histone acetyltransferase	*P. falciparum/ S. cerevisiae*
GCS1	135	44 (37, 7)	27.1	Vesicle-mediated transport: ADP-ribosylation factor GTPase activating protein (ARF GAP)	*O. sativa/S. cerevisiae*
GIC1	263	32 (25, 7)	–	Cell budding: protein of unknown function involved in initiation of budding and cellular polarization	–
GIC2	281	30 (21, 9)	–	Cell budding: protein of unknown function involved in initiation of budding and cellular polarization	–
GLC7	8	109 (27, 82)	84.4	Meiosis: catalytic subunit of type 1 serine/threonine protein phosphatase	–

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
GPI8	48	62 (3, 59)	56.4	Protein modification process: ER membrane glycoprotein subunit of the glycosylphosphatidylinositol transamidase complex	–
GRR1	164	40 (29, 11)	24.9	Response to stress: F-box protein component of the SCF ubiquitin-ligase complex	–
GSF2	58	59 (2, 57)	–	Transport: ER localized integral membrane protein that may promote secretion of certain hexose transporters	–
GSP1	84	51 (9, 42)	82.4	Nuclear organization and biogenesis: GTP binding protein (mammalian Ranp homolog) involved in the maintenance of nuclear organization	–
GTS1	258	32 (9, 23)	–	Transcription: protein containing a zinc-finger in the N terminus and a long Gln-rich region in the C terminus	–
HHF1	42	66 (24, 42)	92.2	DNA metabolic process: one of two identical histone H4 proteins	–
HHF2	269	31 (8, 23)	92.2	DNA metabolic process: one of two identical histone H4 proteins	–
HHT1	36	69 (25, 44)	90.4	DNA metabolic process: one of two identical histone H3 proteins	–
HHT2	234	34 (8, 26)	90.4	DNA metabolic process: one of two identical histone H3 proteins	–
HOG1	92	50 (26, 24)	46.4	Signal transduction: mitogen-activated protein kinase involved in osmoregulation via three independent osmosensors	*A. adeninivorans*/ *S. cerevisiae*
HPR1	55	61 (41, 20)	–	Response to stress: subunit of THO/TREX complexes that couple transcription elongation with mitotic recombination and with mRNA metabolism and export	–
HPR5	5	119 (36, 83)	–	DNA metabolic process: DNA helicase and DNA-dependent ATPase involved in DNA repair	–
HSF1	317	26 (21, 5)	48	Response to stress: trimeric heat shock transcription factor	*H. sapiens*/ *S. cerevisiae*

(*Continued*)

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
HSP82	99	49 (25, 24)	63	Response to stress: cytoplasmic chaperone (Hsp90 family) required for pheromone signaling and negative regulation of Hsf1p	*C. elegans*/ *S. cerevisiae*
HTA1	82	52 (21, 31)	78.2	DNA metabolic process: one of two nearly identical (see also HTA2) histone H2A subtypes	–
HTB1	235	34 (9, 25)	72.4	DNA metabolic process: one of two nearly identical (see HTB2) histone H2B subtypes required for chromatin assembly and chromosome function	–
HTZ1	79	54 (24, 30)	69.2	Transcription: histone variant H2AZ	–
IFA38	10	104 (2, 102)	33.6	Lipid metabolic process: microsomal beta-keto-reductase	–
IPL1	243	34 (25, 9)	43.1	Cytoskeleton organization and biogenesis: aurora kinase involved in regulating kinetochore-microtubule attachments	–
IRE1	262	32 (22, 10)	44.2	Protein modification process: serine-threonine kinase and endoribonuclease	*T. reesei*/*S. cerevisiae*
KAP123	331	23 (1, 22)	23.5	Transport: karyopherin beta	–
KAP95	233	34 (1, 33)	35.7	Nuclear organization and biogenesis: karyopherin beta	–
KAR2	174	39 (16, 23)	67.5	Response to stress: ATPase involved in protein import into the ER	*A. oryzae*/*S. cerevisiae*
KAR9	69	57 (47, 10)	–	Cytoskeleton organization and biogenesis: karyogamy protein required for positioning of the mitotic spindle and cytoplasmic microtubules	–
KEM1	229	35 (27, 8)	52.8	Ribosome biogenesis and assembly: evolutionarily conserved 5′-3′ exonuclease component of cytoplasmic processing (P) bodies involved in mRNA decay	*D. melanogaster*/ *S. cerevisiae*
KRR1	287	29 (2, 27)	64.1	Ribosome biogenesis and assembly: involved in 18S rRNA synthesis and 40S ribosomal subunit assembly	–
LAC1	288	29 (3, 26)	27	Process unknown: ceramide synthase component	*H. sapiens*/ *S. cerevisiae*

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
LAS17	187	38 (15, 23)	30.7	Cytokinesis: actin assembly factor	–
LTE1	18	81 (23, 58)	–	Vesicle-mediated transport: putative GDP/GTP exchange factor required for mitotic exit at low temperatures	–
MAD2	63	58 (39, 19)	43.1	Organelle organization and biogenesis: component of the spindle-assembly checkpoint complex	–
MAS6	265	31 (4, 27)	27.3	Transport: essential protein of the mitochondrial inner membrane	–
MCD1	139	43 (18, 25)	29.2	Cell cycle: Required for sister chromatid cohesion in mitosis and meiosis	–
MCM2	280	30 (7, 23)	52.7	DNA metabolic process: protein involved in DNA replication	*E. histolytica/ S. cerevisiae*
MEC1	27	75 (66, 9)	38.5	DNA metabolic process: genome integrity checkpoint protein and PI kinase superfamily member	–
MEC3	297	29 (24, 5)	–	DNA metabolic process: DNA damage and meiotic pachytene checkpoint protein	–
MED6	339	22 (1, 21)	47.7	Transcription: subunit of the RNA polymerase II mediator complex	–
MED7	340	22 (1, 21)	39.5	Transcription: subunit of the RNA polymerase II mediator complex	–
MIG1	179	39 (29, 10)	–	Transcription: transcription factor involved in glucose repression	–
MLP2	273	31 (10, 21)	22.3	Transcription: myosin-like protein associated with the nuclear envelope	–
MRE11	111	47 (34, 13)	35.1	DNA metabolic process: subunit of RMX complex that functions in repair of DNA double-strand breaks and in telomere stability	–
MSH2	154	41 (29, 12)	45.4	DNA metabolic process: protein that forms heterodimers with Msh3p and Msh6p that bind to DNA mismatches to initiate the mismatch repair process	–
MSN5	197	37 (16, 21)	27.1	Transport: karyopherin involved in nuclear import and export	–
MST27	205	36 (2, 34)	–	Organelle organization and biogenesis: putative integral membrane protein	–

(*Continued*)

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
MUC1	183	39 (39, 0)	31.6	Pseudohyphal growth: GPI-anchored cell surface glycoprotein required for diploid pseudohyphal formation and haploid invasive growth	–
MYO2	95	50 (33, 17)	40.9	Vesicle-mediated transport: one of two type V myosin motors (along with MYO4) involved in actin-based transport of cargos	–
NGG1	185	38 (8, 30)	27.2	Protein modification process: transcriptional regulator involved in glucose repression of Gal4p-regulated genes	–
NHP6A	276	31 (25, 6)	46.5	DNA metabolic process: high-mobility group nonhistone chromatin protein	–
NOG1	309	27 (2, 25)	44.4	Ribosome biogenesis and assembly: putative GTPase required for 60S ribosomal subunit biogenesis	–
NOP1	145	42 (5, 37)	67.4	Ribosome biogenesis and assembly: nucleolar protein	*H. sapiens*/ *S. cerevisiae*
NOP58	193	37 (3, 34)	47.4	Ribosome biogenesis and assembly: protein involved in prer RNA processing	–
NOP7	91	50 (3, 47)	41.5	Ribosome biogenesis and assembly: nucleolar protein involved in rRNA processing and 60S ribosomal subunit biogenesis	–
NPL3	149	41 (10, 31)	54.6	Pseudohyphal growth: RNA-binding protein that carries poly(A)$^+$ mRNA from the nucleus into the cytoplasm	–
NRD1	302	28 (4, 24)	23.1	Transcription: RNA-binding protein that interacts with the C-terminal domain of the RNA polymerase II large subunit (Rpo21p)	–
NSP1	274	31 (10, 21)	29.7	Nuclear organization and biogenesis: essential component of the nuclear pore	–
NUF2	103	48 (5, 43)	23.6	Process unknown: component of the evolutionarily conserved kinetochore-associated Ndc80 complex (Ndc80p-Nuf2p-Spc24p-Spc25p)	–

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
NUP1	237	34 (13, 21)	28.6	Nuclear organization and biogenesis: nuclear pore complex (NPC) subunit	–
ORC1	251	33 (9, 24)	62.5	DNA metabolic process: largest subunit of the origin recognition complex	–
ORC2	175	39 (17, 22)	22	DNA metabolic process: subunit of the origin recognition complex	–
OST4	319	25 (2, 23)	–	Protein modification process: subunit of the oligosaccharyl-transferase complex of the ER lumen	–
PAB1	116	46 (17, 29)	54.3	Translation: Poly(A) binding protein	*A. thaliana/ S. cerevisiae*
PBS2	282	30 (22, 8)	48.8	Signal transduction: MAP kinase kinase that plays a pivotal role in the osmosensing signal-transduction pathway	*D. hansenii/ S. cerevisiae*
PCL1	158	40 (13, 27)	–	Cell cycle: Pho85 cyclin	–
PDE2	218	36 (36, 0)	34.1	Signal transduction: high-affinity cyclic AMP phosphodiesterase	*H. sapiens/ S. cerevisiae*
PDR5	293	29 (8, 21)	–	Transport: short-lived membrane ABC (ATP-binding cassette) transporter	–
PDS1	226	35 (24, 11)	–	Cell cycle: securin that inhibits anaphase by binding separin Esp1p	–
PEP12	104	48 (16, 32)	26.4	Vesicle-mediated transport: Target membrane receptor (t-SNARE) for vesicular intermediates traveling between the Golgi apparatus and the vacuole	–
PGD1	291	29 (7, 22)	–	Transcription: subunit of the Mediator global transcriptional cofactor complex	–
PHO2	310	27 (2, 25)	26	Transcription: homeobox transcription factor	–
PHO4	1	146 (4, 142)	–	Response to stress: basic helix-loop-helix (bHLH) transcription factor	–
PHO85	29	72 (23, 49)	52.7	Cell cycle: cyclin-dependent kinase	*M. musculus/ S. cerevisiae*
PHO86	137	43 (5, 38)	–	Transport: endoplasmic reticulum (ER) resident protein required for ER exit of the high-affinity phosphate transporter Pho84p	–

(*Continued*)

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
PHO88	9	108 (1, 107)	–	Transport: probable membrane protein	–
PIF1	264	32 (32, 0)	43.6	DNA metabolic process: DNA helicase involved in telomere formation and elongation	–
PIK1	330	24 (23, 1)	44.3	Cytokinesis: phosphatidylinositol 4-kinase	–
PKC1	23	77 (62, 15)	45.5	Signal transduction: protein serine/threonine kinase essential for cell wall remodeling during growth	–
PLC1	283	30 (24, 6)	34.2	Signal transduction: phosphoinositide-specific phospholipase C	–
PMA1	106	48 (23, 25)	–	Transport: plasma membrane H^+-ATPase	*A. thaliana/ S. cerevisiae*
PMR1	32	70 (30, 40)	39.9	Transport: high-affinity Ca^{2+}/Mn^{2+} P-type ATPase required for Ca^{2+} and Mn^{2+} transport into Golgi	*A. fumigatus/ S. cerevisiae*
POL1	162	40 (25, 15)	35.8	DNA metabolic process: DNA polymerase alpha-primase catalytic subunit	–
POL2	225	35 (23, 12)	51.8	DNA metabolic process: catalytic subunit of DNA polymerase epsilon	–
POL30	49	62 (36, 26)	35.5	DNA metabolic process: proliferating cell nuclear antigen (PCNA)	–
POL32	307	28 (21, 7)	–	DNA metabolic process: third subunit of DNA polymerase delta	–
POM34	259	32 (11, 21)	–	Transport: integral membrane protein of the nuclear pore	–
POP2	160	40 (22, 18)	39.6	RNA metabolic process: RNase of the DEDD superfamily	*M. musculus/ S. cerevisiae*
PPA1	220	35 (2, 33)	55.6	Organelle organization and biogenesis: subunit c″ of the vacuolar ATPase	–
PRE1	96	49 (10, 39)	44.8	Response to stress: 20S proteasome beta-type subunit	–
PRE2	268	31 (7, 24)	62.6	Protein catabolic process: 20S proteasome beta-type subunit	–
PRP8	71	56 (14, 42)	62.5	Process unknown: component of the U4/U6-U5 snRNP complex	–

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
PSE1	125	45 (4, 41)	29.8	Transport: karyopherin/importin that interacts with the nuclear pore complex	–
PTA1	279	30 (2, 28)	30.2	RNA metabolic process: subunit of holo-CPF	–
RAD1	47	63 (58, 5)	42.3	Cell cycle: single-stranded DNA endonuclease (with Rad10p)	*A. thaliana/ S. cerevisiae*
RAD14	199	37 (25, 12)	31.3	Organelle organization and biogenesis: protein that recognizes and binds damaged DNA during nucleotide excision repair	–
RAD17	298	29 (24, 5)	–	Meiosis: checkpoint protein	–
RAD18	147	42 (35, 7)	25.2	DNA metabolic process: protein involved in postreplication repair	–
RAD23	108	47 (17, 30)	30.7	Protein catabolic process: protein with ubiquitin-like N terminus	–
RAD24	68	57 (44, 13)	28.3	Meiosis: checkpoint protein	–
RAD27	191	38 (35, 3)	58.3	DNA metabolic process: 5′ to 3′ exonuclease	*S. cerevisiae/E. coli*
RAD5	215	36 (29, 7)	53.5	DNA metabolic process: single-stranded DNA-dependent ATPase	–
RAD50	249	34 (32, 2)	28.7	Meiosis: subunit of MRX complex	–
RAD51	39	68 (56, 12)	67.1	DNA metabolic process: strand exchange protein	*P. angusta/ S. cerevisiae*
RAD52	6	116 (106, 10)	43.1	Cell cycle: protein that stimulates strand exchange by facilitating Rad51p binding to single-stranded DNA	*C. albicans/ S. cerevisiae*
RAD53	12	100 (61, 39)	40.8	DNA metabolic process: protein kinase	*H. sapiens/ S. cerevisiae*
RAD55	231	35 (30, 5)	–	DNA metabolic process: protein that stimulates strand exchange by stabilizing the binding of Rad51p to single-stranded DNA	–
RAD6	67	57 (42, 15)	69.5	DNA metabolic process: ubiquitin-conjugating enzyme (E2)	*A. thaliana/ S. cerevisiae*
RAD9	34	70 (63, 7)	–	DNA metabolic process: DNA damage-dependent checkpoint protein	–
RAI1	170	39 (5, 34)	26.6	Ribosome biogenesis and assembly: nuclear protein that binds to and stabilizes the exoribonuclease Rat1p	–

(*Continued*)

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
RAS2	7	113 (102, 11)	57.5	Pseudohyphal growth: GTP-binding protein that regulates the nitrogen starvation response	*H. sapiens/ S. cerevisiae*
REV3	203	37 (35, 2)	40.4	DNA metabolic process: catalytic subunit of DNA polymerase zeta	–
RGR1	209	36 (7, 29)	22.8	Transcription: component of RNA polymerase II holoenzyme/mediator complex	–
RHO1	60	59 (41, 18)	72.7	Cell wall organization and biogenesis: GTP-binding protein of the rho subfamily of Ras-like proteins	*S. mansoni/ S. cerevisiae*
RLF2	89	51 (41, 10)	32.3	Transcription: largest subunit (p90) of the Chromatin Assembly Complex (CAF-I) with Cac2p and Msi1p that assembles newly synthesized histones onto recently replicated DNA	–
ROM2	284	30 (25, 5)	26.5	Response to stress: GDP/GTP exchange protein (GEP) for Rho1p and Rho2p	–
RPB2	223	35 (13, 22)	68.8	Transcription: RNA polymerase II second largest subunit B150	–
RPD3	85	51 (29, 22)	63.2	Protein modification process: histone deacetylase	–
RPN1	290	29 (5, 24)	41.4	Protein catabolic process: non-ATPase base subunit of the 19S regulatory particle of the 26S proteasome	–
RPO21	2	135 (38, 97)	54.9	Transcription: RNA polymerase II largest subunit B220	–
RPT1	256	32 (4, 28)	83.5	Protein catabolic process: one of six ATPases of the 19S regulatory particle of the 26S proteasome involved in the degradation of ubiquitinated substrates	*A. thaliana/ S. cerevisiae*
RPT6	129	44 (8, 36)	76.6	Protein catabolic process: one of six ATPases of the 19S regulatory particle of the 26S proteasome involved in the degradation of ubiquitinated substrates	*A. thaliana/ S. cerevisiae*
RRM3	322	25 (23, 2)	42.5	DNA metabolic process: DNA helicase involved in rDNA replication and Ty1 transposition	–

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
RRP1	123	45 (1, 44)	31.5	Ribosome biogenesis and assembly: nucleolar protein necessary for 60S ribosomal subunit biogenesis and prerRNAs maturation	–
RRP6	254	33 (25, 8)	37.9	Ribosome biogenesis and assembly: exonuclease component	–
RSP5	26	76 (43, 33)	47.5	Transcription: ubiquitin ligase involved in ubiquitin-mediated degradation	–
RTF1	244	34 (26, 8)	23.6	Transcription: subunit of the RNA polymerase II-associated Paf1 complex	–
RTT107	118	46 (19, 27)	–	DNA metabolic process: protein implicated in Mms22-dependent DNA repair during S phase	–
RVS161	198	37 (22, 15)	26.8	Vesicle-mediated transport: amphiphysin-like lipid raft protein	–
RVS167	35	69 (19, 50)	29.5	Vesicle-mediated transport: actin-associated protein	–
SAC6	335	23 (21, 2)	41.8	Vesicle-mediated transport: fimbrin	*H. sapiens/ S. cerevisiae*
SAS2	306	28 (7, 21)	–	Transcription: histone acetyltransferase (HAT) catalytic subunit of the SAS complex (Sas2p-Sas4p-Sas5p)	–
SEC1	261	32 (22, 10)	29.3	Vesicle-mediated transport: Sm-like protein involved in docking and fusion of exocytic vesicles through binding to assembled SNARE complexes	–
SEC10	25	76 (21, 55)	25.2	Cytokinesis: essential 100-kDa subunit of the exocyst complex	–
SEC15	140	43 (31, 12)	21.2	Cytokinesis: essential 113-kDa subunit of the exocyst complex	–
SEC17	257	32 (7, 25)	34.9	Vesicle-mediated transport: peripheral membrane protein required for ER-Golgi vesicular transport and the priming step in homotypic vacuole fusion	–
SEC2	182	39 (34, 5)	–	Vesicle-mediated transport: guanyl-nucleotide exchange factor for the small G-protein Sec4p	–
SEC21	214	36 (24, 12)	40.8	Vesicle-mediated transport: gamma subunit of coatomer	–
SEC22	37	69 (26, 43)	39.5	Vesicle-mediated transport: R-SNARE protein	–

(*Continued*)

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
SEC23	133	44 (19, 25)	50.1	Vesicle-mediated transport: GTPase-activating protein	–
SEC3	156	41 (32, 9)	–	Cytokinesis: nonessential subunit of the exocyst complex	–
SEC4	77	55 (31, 24)	55.9	Cytokinesis: secretory vesicle-associated Rab GTPase essential for exocytosis	–
SEC5	227	35 (26, 9)	21	Cytokinesis: essential 107-kDa subunit of the exocyst complex	–
SEC6	242	34 (24, 10)	23.6	Cytokinesis: essential 88-kDa subunit of the exocyst complex	–
SEC61	186	38 (9, 29)	56.6	Transport: essential subunit of Sec61 complex	–
SEC63	210	36 (7, 29)	35.8	Transport: essential subunit of Sec63 complex	–
SEC72	325	24 (1, 23)	–	Transport: nonessential subunit of Sec63 complex	–
SEC8	239	34 (21, 13)	19.6	Cytokinesis: essential 121-kDa subunit of the exocyst complex	–
SEC9	110	47 (32, 15)	38.4	Vesicle-mediated transport: t-SNARE protein important for fusion of secretory vesicles with the plasma membrane	–
SED5	17	81 (12, 69)	48.1	Vesicle-mediated transport: *cis*-Golgi t-SNARE syntaxin	–
SGS1	31	71 (59, 12)	35.7	Response to stress: nucleolar DNA helicase of the RecQ family involved in maintenance of genome integrity	*H. sapiens*/ *S. cerevisiae*
SHO1	267	31 (6, 25)	25.7	Pseudohyphal growth: transmembrane osmosensor	–
SHR3	75	55 (5, 50)	–	Vesicle-mediated transport: endoplasmic reticulum packaging chaperone	*A. nidulans*/ *S. cerevisiae*
SIC1	73	56 (39, 17)	–	Cell cycle: inhibitor of Cdc28-Clb kinase complexes	*H. sapiens*/ *S. cerevisiae*
SIN3	159	40 (16, 24)	47.2	Protein modification process: component of the Sin3p-Rpd3p histone deacetylase complex	–
SIN4	127	45 (31, 14)	–	Transcription: subunit of the Mediator complex	–
SIR1	228	35 (26, 9)	–	Transcription: protein involved in repression of transcription at the silent mating-type loci HML and HMR	–

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
SIR2	86	51 (30, 21)	50.5	DNA metabolic process: conserved NAD^+-dependent histone deacetylase of the Sirtuin family involved in regulation of lifespan	*S. cerevisiae/S. pombe*
SIR4	176	39 (17, 22)	–	DNA metabolic process: silent information regulator	–
SIT4	120	46 (34, 12)	66	Cell wall organization and biogenesis: Type 2A-related serine-threonine phosphatase that functions in the G1/S transition of the mitotic cycle	–
SKP1	16	90 (12, 78)	57.2	Cytokinesis: kinetochore protein	–
SLA2	57	60 (48, 12)	24.3	Cell wall organization and biogenesis: transmembrane actin-binding protein involved in membrane cytoskeleton assembly and cell polarization	–
SLG1	216	36 (34, 2)	–	Cell wall organization and biogenesis: sensor-transducer of the stress-activated PKC1-MPK1 kinase pathway involved in maintenance of cell wall integrity	–
SLT2	21	79 (63, 16)	45	Signal transduction: serine/threonine MAP kinase involved in regulating the maintenance of cell wall integrity and progression through the cell cycle	–
SLY1	163	40 (27, 13)	33.2	Vesicle-mediated transport: hydrophilic protein involved in vesicle trafficking between the ER and Golgi	–
SML1	313	27 (23, 4)	–	Response to stress: ribonucleotide reductase inhibitor involved in regulating dNTP production	–
SNC2	177	39 (18, 21)	34.1	Vesicle-mediated transport: v-SNARE involved in the fusion between Golgi-derived secretory vesicles with the plasma membrane	–
SNF1	13	92 (55, 37)	59.3	Signal transduction: AMP-activated serine/threonine protein kinase	*S. tuberosum/ S. cerevisiae*

(*Continued*)

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
SNF2	38	68 (28, 40)	52.7	DNA metabolic process: catalytic subunit of the SWI/SNF chromatin remodeling complex involved in transcriptional regulation	–
SNL1	124	45 (3, 42)	–	Nuclear organization and biogenesis: protein involved in nuclear pore complex biogenesis and maintenance and protein folding	–
SNX4	323	24 (1, 23)	24.1	Transport: sorting nexin	–
SPA2	101	49 (32, 17)	–	Pseudohyphal growth: component of the polarisome	–
SPC1	148	41 (1, 40)	–	Transport: subunit of the signal peptidase complex (SPC)	–
SPC110	65	57 (13, 44)	–	Cytoskeleton organization and biogenesis: inner plaque spindle pole body (SPB) component	–
SPC24	266	31 (5, 26)	–	Process unknown: component of the evolutionarily conserved kinetochore-associated Ndc80 complex (Ndc80p-Nuf2p-Spc24p-Spc25p)	–
SPO11	314	27 (24, 3)	26.6	Cell cycle: meiosis-specific protein that initiates meiotic recombination	–
SPO12	53	61 (18, 43)	–	Meiosis: nucleolar protein of unknown function	–
SPO13	204	37 (36, 1)	–	Meiosis: meiosis-specific protein	–
SPT15	11	100 (36, 64)	78.8	Transcription: TATA-binding protein	–
SPT16	81	53 (38, 15)	36.5	RNA metabolic process: subunit of the heterodimeric FACT complex	–
SPT2	241	34 (23, 11)	28.8	RNA metabolic process: protein involved in negative regulation of transcription	–
SPT20	189	38 (17, 21)	–	Protein modification process: subunit of the SAGA transcriptional regulatory complex	–
SPT3	119	46 (24, 22)	27.1	Pseudohyphal growth: subunit of the SAGA and SAGA-like transcriptional regulatory complexes	–

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
SPT4	230	35 (29, 6)	43.4	Transcription: protein that forms a complex with Spt5p and mediates both activation and inhibition of transcription elongation	*H. sapiens*/ *S. cerevisiae*
SPT5	151	41 (22, 19)	43.7	Transcription: protein that forms a complex with Spt4p and mediates both activation and inhibition of transcription elongation	–
SPT7	224	35 (13, 22)	–	Conjugation: subunit of the SAGA transcriptional regulatory complex	–
SPT8	253	33 (12, 21)	–	Transcription: subunit of the SAGA transcriptional regulatory complex but not present in SAGA-like complex SLIK/SALSA	–
SRB4	66	57 (15, 42)	–	Transcription: subunit of the RNA polymerase II mediator complex	–
SRB5	52	61 (14, 47)	–	Transcription: subunit of the RNA polymerase II mediator complex	–
SRM1	130	44 (15, 29)	40.4	Ribosome biogenesis and assembly: nucleotide exchange factor for Gsp1p	*S. cerevisiae*/ *C. griseus*
SRP102	219	35 (1, 34)	26	Transport: signal recognition particle (SRP) receptor beta subunit	–
SSA1	72	56 (17, 39)	75.8	Translation: ATPase involved in protein folding and nuclear localization signal (NLS)-directed nuclear transport	–
SSD1	112	47 (39, 8)	–	Cell wall organization and biogenesis: protein involved in cellular integrity	–
SSF1	222	35 (6, 29)	37.8	Conjugation: constituent of 66S preribosomal particles	–
SSH1	332	23 (1, 22)	–	Transport: subunit of the Ssh1 translocon complex	–
SSN3	132	44 (19, 25)	50.4	Meiosis: cyclin-dependent protein kinase	–
STE11	41	67 (45, 22)	48.8	Pseudohyphal growth: signal transducing MEK kinase involved in pheromone response and pseudohyphal/invasive growth pathways	–
STE12	40	67 (42, 25)	–	Pseudohyphal growth: transcription factor that is activated by a MAP kinase signaling cascade	*C. glabrata*/ *S. cerevisiae*

(*Continued*)

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
STE20	102	49 (35, 14)	56.5	Pseudohyphal growth: PAK (p21-activated kinase) family kinase	*C. glabrata/ S. cerevisiae*
STE24	168	39 (1, 38)	39.1	Protein modification process: Highly conserved zinc metalloprotease that functions in two steps of a-factor maturation	*H. sapiens/ S. cerevisiae*
STE4	178	39 (26, 13)	37.2	Signal transduction: G protein beta subunit	–
STH1	98	49 (20, 29)	51.5	Meiosis: ATPase component of the RSC chromatin remodeling complex	–
STO1	308	28 (23, 5)	28.1	RNA metabolic process: large subunit of the nuclear mRNA cap-binding protein complex	–
STT4	277	31 (29, 2)	42.9	Organelle organization and biogenesis: phosphatidylinositol-4-kinase that functions in the Pkc1p protein kinase pathway	–
SUA7	171	39 (12, 27)	36.2	Transcription: transcription factor TFIIB	–
SUP35	294	29 (8, 21)	50.4	Translation: translation termination factor eRF3	*P. methanolica/ S. cerevisiae*
SUR2	211	36 (9, 27)	–	Lipid metabolic process: sphinganine C4-hydroxylase	–
SUR4	14	91 (12, 79)	28	Vesicle-mediated transport: elongase	–
SUS1	320	25 (3, 22)	31.1	Transcription: protein involved in mRNA export coupled transcription activation	–
SWE1	155	41 (32, 9)	33.3	Meiosis: protein kinase that regulates the G2/M transition by inhibition of Cdc28p kinase activity	*S. cerevisiae/S. pombe*
SWI4	78	55 (44, 11)	–	Transcription: DNA binding component of the SBF complex (Swi4p-Swi6p)	–
SWI6	83	52 (39, 13)	–	Transcription: transcription cofactor	–
SWP1	184	38 (3, 35)	–	Protein modification process: Delta subunit of the oligosaccharyl transferase glycoprotein complex	–
TAF1	157	40 (12, 28)	31.9	Protein modification process: TFIID subunit (145 kDa)	–

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
TAF10	208	36 (6, 30)	36.2	Protein modification process: TFIID and SAGA complex subunit	–
TAF12	207	36 (5, 31)	38.9	Protein modification process: TFIID and SAGA complex subunit	–
TAF13	303	28 (5, 23)	30.3	Cell cycle: TFIID subunit (19 kDa)	–
TAF14	138	43 (6, 37)	–	Cell cycle: subunit of TFIID	–
TAF6	114	46 (3, 43)	30.9	Protein modification process: subunit (60 kDa) of TFIID and SAGA complexes	–
TAF9	304	28 (5, 23)	33.3	Protein modification process: subunit (17 kDa) of TFIID and SAGA complexes	–
TEL1	246	34 (29, 5)	41.3	Response to stress: protein kinase involved in telomere length regulation	–
TFG2	318	25 (1, 24)	26.7	Transcription: TFIIF (Transcription Factor II) middle subunit	–
TID3	271	31 (9, 22)	23.3	Process unknown: component of the evolutionarily conserved kinetochore-associated Ndc80 complex (Ndc80p-Nuf2p-Spc24p-Spc25p)	*H. sapiens*/ *S. cerevisiae*
TLG1	292	29 (7, 22)	–	Vesicle-mediated transport: essential t-SNARE that forms a complex with Tlg2p and Vti1p and mediates fusion of endosome-derived vesicles with the late Golgi	–
TLG2	59	59 (8, 51)	26.8	Vesicle-mediated transport: t-SNARE that mediates fusion of endosome-derived vesicles with the late Golgi	–
TNA1	167	39 (1, 38)	–	Transport: high-affinity nicotinic acid plasma membrane permease	–
TOM70	289	29 (4, 25)	24.3	Membrane organization and biogenesis: component of the TOM (translocase of outer membrane) complex responsible for recognition and initial import steps for all mitochondrially directed proteins	–

(*Continued*)

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
TOP1	299	29 (25, 4)	44.7	DNA metabolic process: Topoisomerase I	*E. coli/S. cerevisiae*
TOP3	232	35 (31, 4)	46	Meiosis: DNA Topoisomerase III	–
TOR1	88	51 (37, 14)	55.8	Signal transduction: PIK-related protein kinase and rapamycin target	–
TOR2	161	40 (23, 17)	55.3	Signal transduction: PIK-related protein kinase and rapamycin target	–
TPK1	260	32 (22, 10)	49.7	Pseudohyphal growth: subunit of cytoplasmic cAMP-dependent protein kinase	–
TPK2	180	39 (30, 9)	52.2	Pseudohyphal growth: subunit of cytoplasmic cAMP-dependent protein kinase	–
TRA1	326	24 (2, 22)	43.4	DNA metabolic process: subunit of SAGA and NuA4 histone acetyltransferase complexes	–
TSC13	143	42 (3, 39)	31.7	Lipid metabolic process: enoyl reductase that catalyzes the last step in each cycle of very long chain fatty acid elongation	*A. thaliana/ S. cerevisiae*
TUB1	153	41 (27, 14)	74.7	Meiosis: alpha-tubulin	–
TUP1	213	36 (12, 24)	–	Transcription: general repressor of transcription	–
UBI4	173	39 (14, 25)	96	Response to stress: ubiquitin	–
UBP8	333	23 (2, 21)	32	Protein modification process: ubiquitin-specific protease that is a component of the SAGA (Spt-Ada-Gcn5-Acetyltransferase) acetylation complex	–
VAM3	195	37 (8, 29)	–	Vesicle-mediated transport: syntaxin-like protein required for vacuolar assembly	–
VAM7	286	29 (1, 28)	–	Vesicle-mediated transport: component of the vacuole SNARE complex involved in vacuolar morphogenesis	–
VPH1	334	23 (2, 21)	43.6	Organelle organization and biogenesis: subunit a of vacuolar-ATPase V0 domain	–
VTI1	272	31 (9, 22)	30.3	Vesicle-mediated transport: protein involved in *cis*-Golgi membrane traffic	*A. thaliana/ S. cerevisiae*

Table 3 (*Continued*)

Yeast gene	Rank[a]	Total interactions (genetic, physical)	% identity to human[b]	GO process: SGD description[c]	Cross complementation[d]
WBP1	51	61 (10, 51)	28.3	Cell cycle: Beta subunit of the oligosaccharyl transferase (OST) complex	–
YIF1	324	24 (1, 23)	29.4	Vesicle-mediated transport: integral membrane protein required for the fusion of ER-derived COPII transport vesicles with the Golgi	–
YIP1	150	41 (18, 23)	41.4	Vesicle-mediated transport: integral membrane protein required for the biogenesis of ER-derived COPII transport vesicles	–
YKT6	172	39 (13, 26)	47	Vesicle-mediated transport: vesicle membrane protein (v-SNARE) with acyltransferase activity	–
YKU70	248	34 (31, 3)	24.6	DNA metabolic process: subunit of the telomeric Ku complex (Yku70p-Yku80p)	–
YOP1	278	30 (2, 28)	38.1	Vesicle-mediated transport: membrane protein that interacts with Yip1p to mediate membrane traffic	–
YPC1	206	36 (2, 34)	38.9	Lipid metabolic process: alkaline ceramidase that also has reverse (CoA-independent) ceramide synthase activity	–
YPT1	45	66 (43, 23)	68.8	Vesicle-mediated transport: Ras-like small GTPase	*T. brucei/S. cerevisiae*
YPT6	109	47 (30, 17)	60.7	Vesicle-mediated transport: GTPase	–
YRA1	28	74 (18, 56)	41.3	Transport: nuclear protein that binds to RNA and to Mex67p	–
ZDS1	181	39 (31, 8)	–	Process unknown: protein that interacts with silencing proteins at the telomere	–

[a]The rank is based on the number of interactions reported for yeast, where the lower number has more interactions.
[b]Percent identity is based on the identity between the yeast and human proteins in the same OrthoMCL family. When more than one human protein is present, identity is averaged.
[c]Gene Ontology (GO) biological process annotation; GO terms used for granular annotations are mapped to the more general biological process term. GO annotation and gene descriptions were downloaded from the SGD web site in April 2006.
[d]When multiple experiments with different species are available for a yeast protein, only one example is shown. See the P-POD web site for the additional experiments and citations.

is the fourth in rank when sorted by interactions, *RAS2* is seventh, *SNF1* is thirteenth, and *ACT1* is fifteenth. Virtually every substantial aspect of cell biology is represented. Thus one can study whatever process one is interested in not only in the human or the mouse, but also in the tractable model organism and vice versa.

CONCLUSION

In the current version of P-POD, there are 984 orthologous groups that contain at least 1 member from each of the 8 species, and 555 that contain members from all but the parasite *P. falciparum*. Summing these together, we find more than 1500 functions that are likely to be universal among free-living organisms that have the standard eukaryotic lifestyle. The reason that so many are absent from *P. falciparum* may reflect its very different lifestyle, for instance, the highly synchronous waves of transcription (3, 24). For approximately 20% of these, at least one successful interspecies complementation has been reported in the literature. We see this as a very strong argument for the universality of the basic cell biology of all eukaryotes.

SUMMARY POINTS

1. Proteins are remarkably conserved from single-cellular to multicellular eukaryotes.
2. Many mammalian proteins, including disease-related proteins, can functionally replace their yeast counterparts, and vice versa.
3. With complete genomes available for the major model eukaryotes, orthologous protein sets can be identified on a global scale, and many such computational methods exist.
4. Proteins essential for viability and/or that interact with several other proteins are more likely to be conserved.
5. Not just individual proteins, but how those proteins work in concert within the context of larger biological processes are shared among all eukaryotes.
6. Taking the vast amount of evidence for sequence, network, and functional conservation together, the universality of the basic cell biology of eukaryotes is clear.

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Albert R, Jeong H, Barabasi AL. 2000. Error and attack tolerance of complex networks. *Nature* 406:378–82
2. Beach D, Durkacz B, Nurse P. 1982. Functionally homologous cell cycle control genes in budding and fission yeast. *Nature* 300:706–9
3. Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, DeRisi JL. 2003. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* 1:E5
4. Celenza JL, Carlson M. 1986. A yeast gene that is essential for release from glucose repression encodes a protein kinase. *Science* 233:1175–80

5. Chen F, Mackey AJ, Stoeckert CJ Jr, Roos DS. 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res.* 34:D363–68
6. Chervitz SA, Aravind L, Sherlock G, Ball CA, Koonin EV, et al. 1998. Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science* 282:2022–28
7. Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30:1575–84
8. Ferrell K, Wilkinson CR, Dubiel W, Gordon C. 2000. Regulatory subunit interactions of the 26S proteasome, a complex problem. *Trends Biochem. Sci.* 25:83–88
9. Foor F, Parent SA, Morin N, Dahl AM, Ramadan N, et al. 1992. Calcineurin mediates inhibition by FK506 and cyclosporin of recovery from alpha-factor arrest in yeast. *Nature* 360:682–84
10. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, et al. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415:141–47
11. **Hariharan IK, Haber DA. 2003. Yeast, flies, worms, and fish in the study of human disease. *N. Engl. J. Med.* 348:2457–63**
12. Heinicke S, Livstone MS, Lu C, Oughtred R, Kang F, et al. 2007. The Princeton Protein Orthology Database (P-POD): a comparative genomics analysis tool for biologists. *PLoS ONE* 2(8):e766
13. Hong SP, Leiper FC, Woods A, Carling D, Carlson M. 2003. Activation of yeast Snf1 and mammalian AMP-activated protein kinase by upstream kinases. *Proc. Natl. Acad. Sci. USA* 100:8839–43
14. Huang TW, Tien AC, Huang WS, Lee YC, Peng CL, et al. 2004. POINT: a database for the prediction of protein-protein interactions based on the orthologous interactome. *Bioinformatics* 20:3273–76
15. Jeong H, Mason SP, Barabasi AL, Oltvai ZN. 2001. Lethality and centrality in protein networks. *Nature* 411:41–42
16. **Kataoka T, Powers S, Cameron S, Fasano O, Goldfarb M, et al. 1985. Functional homology of mammalian and yeast RAS genes. *Cell* 40:19–26**
17. Kelley BP, Sharan R, Karp RM, Sittler T, Root DE, et al. 2003. Conserved pathways within bacteria and yeast as revealed by global protein network alignment. *Proc. Natl. Acad. Sci. USA* 100:11394–99
18. Kelley BP, Yuan B, Lewitter F, Sharan R, Stockwell BR, Ideker T. 2004. PathBLAST: a tool for alignment of protein interaction networks. *Nucleic Acids Res.* 32:W83–88
19. Kellis M, Patterson N, Birren B, Berger B, Lander ES. 2004. Methods in comparative genomics: genome correspondence, gene identification and regulatory motif discovery. *J. Comput. Biol.* 11:319–55
20. Kemmer D, Huang Y, Shah SP, Lim J, Brumm J, et al. 2005. Ulysses—an application for the projection of molecular interactions across species. *Genome Biol.* 6:R106
21. Koyuturk M, Kim Y, Topkara U, Subramaniam S, Szpankowski W, Grama A. 2006. Pairwise alignment of protein interaction networks. *J. Comput. Biol.* 13:182–99
22. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, et al. 2006. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440:637–43
23. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921
24. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, et al. 2003. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* 301:1503–8
25. Leach FS, Nicolaides NC, Papadopoulos N, Liu B, Jen J, et al. 1993. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* 75:1215–25

Summarizes several instances, from using zebrafish to study heart abnormalities to using yeast to learn about breast cancer, where basic research in the model organisms has shed light on the biology of human disease.

The first paper to show that a mammalian sequence can functionally replace the corresponding gene in yeast.

Presents the idea of interspecies complementation to clone the human gene encoding the catalytic subunit of the primary cell-cycle dependent protein kinase. This study is part of Paul Nurse's Nobel-prize winning work in characterizing the cell cyle; he shared the award with Lee Hartwell and Tim Hunt.

26. Lee MG, Nurse P. 1987. Complementation used to clone a human homologue of the fission yeast cell cycle control gene *cdc2*. *Nature* 327:31–35

27. Lee Y, Sultana R, Pertea G, Cho J, Karamycheva S, et al. 2002. Cross-referencing eukaryotic genomes: TIGR Orthologous Gene Alignments (TOGA). *Genome Res.* 12:493–502
28. Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13:2178–89
29. Liang Z, Xu M, Teng M, Niu L. 2006. NetAlign: a web-based tool for comparison of protein interaction networks. *Bioinformatics* 22:2175–77
30. Lothe RA, Peltomaki P, Meling GI, Aaltonen LA, Nystrom-Lahti M, et al. 1993. Genomic instability in colorectal cancer: relationship to clinicopathological variables and family history. *Cancer Res.* 53:5849–52
31. Momcilovic M, Hong SP, Carlson M. 2006. Mammalian TAK1 activates Snf1 protein kinase in yeast and phosphorylates AMP-activated protein kinase in vitro. *J. Biol. Chem.* 281:25336–43
32. O'Brien KP, Remm M, Sonnhammer EL. 2005. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res.* 33:D476–80
33. O'Brien KP, Westerlund I, Sonnhammer EL. 2004. OrthoDisease: a database of human disease orthologs. *Hum. Mutat.* 24:112–19
34. Pelham HR. 1999. SNAREs and the secretory pathway-lessons from yeast. *Exp. Cell Res.* 247:1–8
35. Peltomaki P, Lothe RA, Aaltonen LA, Pylkkanen L, Nystrom-Lahti M, et al. 1993. Microsatellite instability is associated with tumors that characterize the hereditary nonpolyposis colorectal carcinoma syndrome. *Cancer Res.* 53:5853–55
36. Penkett CJ, Morris JA, Wood V, Bahler J. 2006. YOGY: a web-based, integrated database to retrieve protein orthologs and associated Gene Ontology terms. *Nucleic Acids Res.* 34:W330–34
37. Reguly T, Breitkreutz A, Boucher L, Breitkreutz BJ, Hon GC, et al. 2006. Comprehensive curation and analysis of global interaction networks in *Saccharomyces cerevisiae*. *J. Biol.* 5:11
38. Remm M, Sonnhammer E. 2000. Classification of transmembrane protein families in the *Caenorhabditis elegans* genome and identification of human orthologs. *Genome Res.* 10:1679–89
39. Remm M, Storm CE, Sonnhammer EL. 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* 314:1041–52
40. Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, et al. 2000. Comparative genomics of the eukaryotes. *Science* 287:2204–15
41. Sanz P. 2003. Snf1 protein kinase: a key player in the response to cellular stress in yeast. *Biochem. Soc. Trans.* 31:178–81
42. Sharan R, Suthram S, Kelley RM, Kuhn T, McCuine S, et al. 2005. Conserved patterns of protein interaction in multiple species. *Proc. Natl. Acad. Sci. USA* 102:1974–79
43. Strand M, Prolla TA, Liskay RM, Petes TD. 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* 365:274–76
44. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, et al. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformat.* 4:41
45. Ton VK, Mandal D, Vahadji C, Rao R. 2002. Functional expression in yeast of the human secretory pathway Ca^{2+}, Mn^{2+}-ATPase defective in Hailey-Hailey disease. *J. Biol. Chem.* 277:6422–27
46. Ton VK, Rao R. 2004. Functional expression of heterologous proteins in yeast: insights into Ca^{2+} signaling and Ca^{2+}-transporting ATPases. *Am. J. Physiol. Cell Physiol.* 287:C580–89

47. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. 2001. The sequence of the human genome. *Science* 291:1304–51
48. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, et al. 2006. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 34:D173–80
49. Wickner W, Haas A. 2000. Yeast homotypic vacuole fusion: a window on organelle trafficking mechanisms. *Annu. Rev. Biochem.* 69:247–75
50. Woods A, Munday MR, Scott J, Yang X, Carlson M, Carling D. 1994. Yeast SNF1 is functionally related to mammalian AMP-activated protein kinase and regulates acetyl-CoA carboxylase in vivo. *J. Biol. Chem.* 269:19509–15
51. Wuchty S. 2002. Interaction and domain networks of yeast. *Proteomics* 2:1715–23
52. Wuchty S. 2004. Evolution and topology in the yeast protein interaction network. *Genome Res.* 14:1310–14
53. Wuchty S, Barabasi AL, Ferdig MT. 2006. Stable evolutionary signal in a yeast protein interaction network. *BMC Evol. Biol.* 6:8
54. Yuan YP, Eulenstein O, Vingron M, Bork P. 1998. Towards detection of orthologues in sequence databases. *Bioinformatics* 14:285–89
55. Zheng XH, Lu F, Wang ZY, Zhong F, Hoover J, Mural R. 2005. Using shared genomic synteny and shared protein functions to enhance the identification of orthologous gene pairs. *Bioinformatics* 21:703–10