Measuring the evolutionary rate of protein–protein interaction

Wenfeng Qian[a], Xionglei He[a,1], Edwin Chan[a,2], Huailiang Xu[a,b], and Jianzhi Zhang[a,3]

[a]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109; and [b]College of Animal Science and Technology, Sichuan Agricultural University, Ya'an, Sichuan 625000, China

Edited by Masatoshi Nei, Pennsylvania State University, University Park, PA, and approved April 14, 2011 (received for review March 23, 2011)

Despite our extensive knowledge about the rate of protein sequence evolution for thousands of genes in hundreds of species, the corresponding rate of protein function evolution is virtually unknown, especially at the genomic scale. This lack of knowledge is primarily because of the huge diversity in protein function and the consequent difficulty in gauging and comparing rates of protein function evolution. Nevertheless, most proteins function through interacting with other proteins, and protein–protein interaction (PPI) can be tested by standard assays. Thus, the rate of protein function evolution may be measured by the rate of PPI evolution. Here, we experimentally examine 87 potential interactions between *Kluyveromyces waltii* proteins, whose one to one orthologs in the related budding yeast *Saccharomyces cerevisiae* have been reported to interact. Combining our results with available data from other eukaryotes, we estimate that the evolutionary rate of protein interaction is $(2.6 \pm 1.6) \times 10^{-10}$ per PPI per year, which is three orders of magnitude lower than the rate of protein sequence evolution measured by the number of amino acid substitutions per protein per year. The extremely slow evolution of protein molecular function may account for the remarkable conservation of life at molecular and cellular levels and allow for studying the mechanistic basis of human disease in much simpler organisms.

The rate of protein sequence evolution has been of enduring interest to evolutionary biologists (1–7) ever since the primary sequences of homologous proteins became available about 50 years ago (8). Estimation and comparison of the rate of protein sequence evolution led to several major discoveries, including the establishment of the molecular clock concept (1), the application of the molecular clock concept to molecular dating of evolutionary events (9), and the proposal of the neutral theory of molecular evolution (2, 10, 11), a paradigm-shifting episode in the history of evolutionary biology (12). In the last decade, studies of the rate of protein sequence evolution have exploded because of the availability of hundreds of complete genome sequences from diverse organisms. Despite some controversies, much has been learned from these studies, such as the identification of various determinants of the rate of protein sequence evolution (6, 13–22) and the estimation of the fraction of proteins subject to positive selection in human and ape evolution (23–26).

Surprisingly, however, very little is known about the rate of protein function evolution, despite the fact that such information could be invaluable for answering a number of important questions. For instance, if most amino acid changes are adaptive, one would predict a positive correlation between the rate of protein function change and the rate of protein sequence change. By contrast, this correlation is not expected if most amino acid changes do not affect protein function and are neutral. Knowing the rate of protein function evolution also helps us understand the speed and frequency with which new functions originate in evolution. Although the rate of protein function evolution can be calculated anecdotally for a few functionally well-characterized proteins, such as the vertebrate hemoglobin and opsin, there has been no systematic effort to estimate this rate from many proteins. This is probably because of the huge diversity in protein function, which makes functional characterizations of many proteins from multiple species both technically challenging and financially costly. Furthermore, even if such functional data are available, it would be difficult to quantify functional differences among homologous proteins and compare such differences among nonhomologous proteins. Thus, it would be ideal to have a universal functional measure that can be applied to a large number of proteins for estimating and comparing the rates of protein function evolution. Because most proteins function through interacting with other proteins and protein–protein interactions (PPIs) can be tested by standard methods such as the yeast two-hybrid (Y2H) assay (27), we propose to use PPI as one universal measure of protein function in estimating the evolutionary rate of protein function. In this study, we experimentally probe PPI evolution between two yeast species using Y2H assays. Combining our data with existing PPI data from multiple other eukaryotes, we estimate that the evolutionary rate of PPI is three orders of magnitude lower than that of the protein sequence. The striking conservation of protein molecular function has important implications for evolutionary biology and biomedicine.

Results and Discussion

Measuring the Rate of PPI Evolution in Two Yeasts. Because large-scale PPI data are available for a number of model organisms such as the budding yeast *Saccharomyces cerevisiae* (28), nematode worm *Caenorhabditis elegans* (29), fruit fly *Drosophila melanogaster* (30), and human *Homo sapiens* (31), one might think that the rate of PPI evolution can be estimated directly from these existing data. Such estimation, however, would be highly unreliable, because these data were generated by high-throughput methods that have high false-negative and false-positive rates whose exact values are either unknown or not known with any precision (28). Instead, we decided to individually examine the interactions between two proteins whose respective one to one orthologs in another species are known to interact. The requirement for one to one orthologs minimizes the influence of gene duplication, which is known to induce changes of protein function (32–34), including PPIs (35, 36). We chose to compare the budding yeast *S. cerevisiae* (*Sce*), a genetic model organism with abundant PPI data (37), with its relative *Kluyveromyces waltii* (*Kwa*), which diverged ~150 MYA (38). This intermediate level of divergence provides time for potential evolutionary changes in PPI, but it also ensures accuracy in identifying one to one orthologs (*Materials and Methods*).

We started by identifying the subset of *Sce* proteins that have one to one orthologs in *Kwa* (Fig. 1). For this subset, we then identified from BioGRID (http://thebiogrid.org/) 335 PPIs with at least two Y2H reports and 481 PPIs with only one small-scale Y2H report. We focused on previous Y2H reports, because different types of PPIs have variable rates of detection by different methods (39) and Y2H is our method of choice. We

Author contributions: W.Q., X.H., and J.Z. designed research; W.Q., X.H., E.C., and H.X. performed research; J.Z. contributed new reagents/analytic tools; W.Q. and J.Z. analyzed data; and W.Q. and J.Z. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]Present address: State Key Laboratory of Biocontrol, College of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China.

[2]Present address: College of Medicine, University of Illinois, Chicago, IL 60612.

[3]To whom correspondence should be addressed. E-mail: jianzhi@umich.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1104695108/-/DCSupplemental.

disregarded PPIs with only one large-scale Y2H report because of the high false-positive rates of high-throughput studies. We refer to these 335 + 481 = 836 PPIs as putative *Sce* PPIs. They should have relatively low probabilities to be false positives, because they have been reported in either one small-scale experiment or at least two experiments (Fig. 1). We selected *Kwa* genes orthologous to the *Sce* genes involved in 115 randomly chosen putative PPIs after setting several criteria to lessen the effort for gene cloning (Fig. 1), and we were able to clone 87 pairs of them for a standard Y2H assay (*Materials and Methods*).

We found that, among the 87 protein pairs (Table S1), one or both members of 11 pairs showed self-activation (Fig. 1; see also row II, column B in Fig. 2) and hence, could not be evaluated for PPI; 33 pairs had no PPI (Fig. 1; see also column E in Fig. 2), and the remaining 43 had PPIs (Fig. 1; see also column C in Fig. 2). To validate that the 33 *Kwa* noninteractive pairs represent true evolutionary changes between *Sce* and *Kwa*, we need to confirm that their corresponding *Sce* proteins interact in our Y2H assay. We were able to clone the *Sce* genes for 29 of 33 pairs (Fig. 1). Excluding 2 self-activation cases, we found 0 of the remaining 27 *Sce* pairs to interact (Fig. 1; see also column F in Fig. 2), which is owing to either false-positive errors in the *Sce* PPI database or the known variation in PPI detection by different variants of Y2H (40). Thus, 0 of 33 *Kwa* noninteractive pairs can be confirmed to have resulted from true evolutionary changes.

If two proteins have been reported to interact in *Sce* and their orthologs are confirmed by us to interact in *Kwa*, the probability that the reported *Sce* interaction is not genuine is lower than 0.005 (*SI Materials and Methods*). Thus, for the 43 *Kwa* PPIs, their corresponding *Sce* PPIs are most likely true. Indeed, we were able to experimentally validate each of 20 randomly selected corresponding *Sce* PPIs (Fig. 1; see also column D in Fig. 2 and Table S1).

Taken together, our experiments showed that 43 of 43 *Sce* PPIs are conserved in *Kwa*. Although PPIs detected by Y2H may not be biological, those detected here are highly likely to be biological, because nonspecific or artificial interactions are not expected to be evolutionarily conserved. Our experiment was designed to identify *Sce* PPIs that are absent in *Kwa* because of either gains of interactions in *Sce* or losses of interactions in *Kwa* after the species separation. Hence, we effectively measured the total gains and losses that occurred in one lineage during 150 MY. With this consideration, we estimated that the 95% confidence interval of the total rate of PPI evolution is between 0 and 4.6×10^{-10} per PPI per year, with the maximum likelihood estimate being 0 (Fig. 3 and *SI Materials and Methods*).

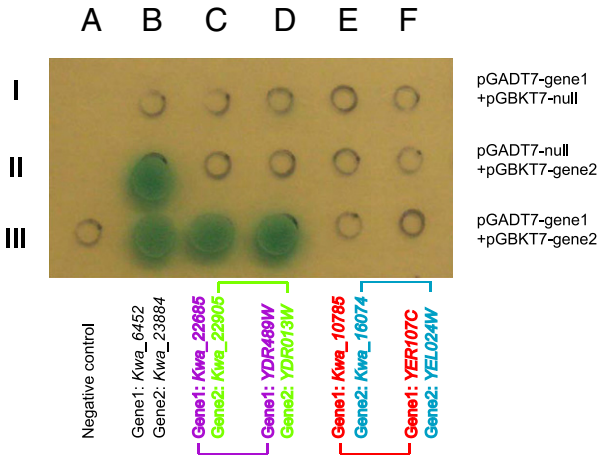

Fig. 2. Examples of yeast two-hybrid (Y2H) experimental results for *S. cerevisiae* (*Sce*) and *K. waltii* (*Kwa*) genes. Yeast cells for the Y2H assays were placed on to the test plates (adenine, histidine, leucine, and tryptophan dropout synthetic media with X-α-gal) in regions marked with black circles. Plasmids in the yeast cells are indicated. pGADT7-gene1 and pGBKT7-gene2 are the Y2H plasmids with genes 1 and 2 inserted, respectively. pGADT7-null and pGBKT7-null are the empty plasmids without gene inserts. Row III, column A is the negative control (pGADT7-null + pGBKT7-null). Row II, column B shows an example of self-activation of *Kwa_23884*. Columns C and D show examples of positive *Kwa* PPIs, whose corresponding *Sce* PPIs are also confirmed. Column E shows an example of negative *Kwa* PPI, whereas column F shows that the corresponding *Sce* PPI is also negative. Gene names starting with *Kwa* are *Kwa* genes; otherwise, they are *Sce* genes. One to one orthologous genes have the same color in gene name and are connected by lines.

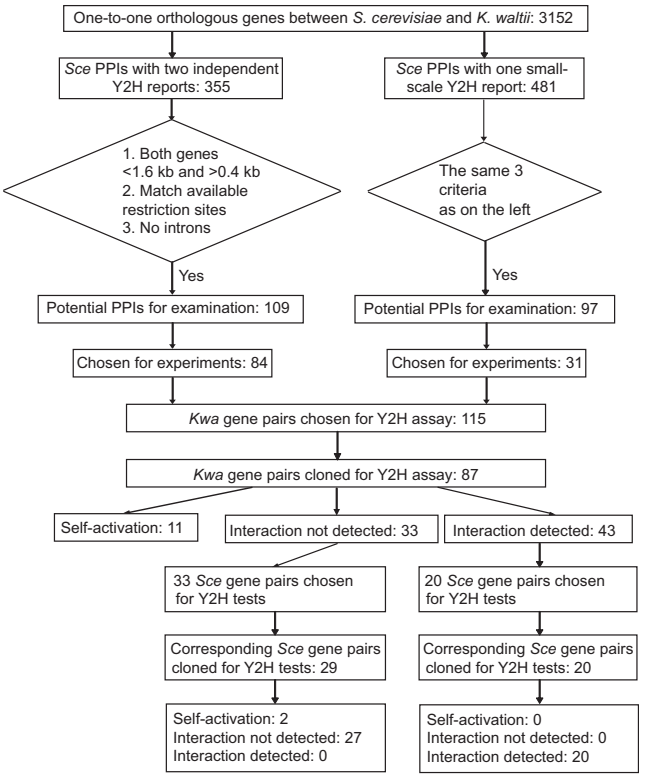

Fig. 1. Flowchart describing the selection of candidate protein–protein interactions (PPIs) for yeast two-hybrid (Y2H) assays and the experimental results of PPI conservation between *S. cerevisiae* (*Sce*) and *K. waltii* (*Kwa*).

Combined Estimate of the Rate of PPI Evolution. To investigate the generality of our estimate, we analyzed all previously reported

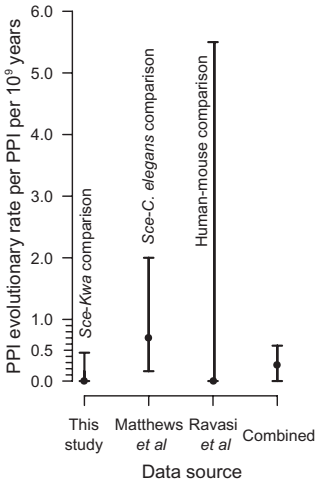

Fig. 3. Maximum likelihood estimates (dots) and 95% confidence intervals (error bars) of the evolutionary rate of protein–protein interaction (PPI).

between-species PPI differences for which false-positive and false-negative errors can be excluded. One study (41) reported that 6 of 19 *Sce* PPIs are conserved in *C. elegans*. However, the study did not use one to one orthologs and thus, could not exclude the influence of gene duplication (41). If only one to one orthologs are considered, their data indicate that two of five *Sce* PPIs are conserved in *C. elegans*. If the two species diverged 1,300 MYA, as suggested by molecular dating (9), the maximum likelihood estimate of the PPI evolutionary rate is 7.0×10^{-10}; and the 95% confidence interval is 1.6×10^{-10} to 2.0×10^{-9} (Fig. 3 and *SI Materials and Methods*). Another study used a high-throughput method to examine the PPIs between transcription factors and found 6 of 23 gold-standard mouse PPIs to be conserved in human (42). The purpose of using gold-standard PPIs was to avoid false positives. Because the rate of detection of a true PPI in that study (42) was 0.253, the actual fraction of mouse PPIs conserved in human is $6/(23 \times 0.253) \approx 100\%$ (i.e., six of six). If human and mouse diverged 90 MYA (9), the maximum likelihood estimate of the PPI evolutionary rate is 0, with the 95% confidence interval being 0 to 5.5×10^{-9} (Fig. 3 and *SI Materials and Methods*).

The confidence intervals of the PPI evolutionary rate estimates from the three datasets encompassing yeasts, worms, and mammals overlap (Fig. 3), although the estimate from our data is the most precise, because the size of our data is four times that of the previous data combined. Using the three datasets together, we derived a maximum likelihood estimate of the PPI evolutionary rate of 2.6×10^{-10} per function per year, with an SE of 1.6×10^{-10} (*SI Materials and Methods*). Our estimated rate of PPI evolution is extremely low. As a comparison, the rate of sequence evolution for the yeast, *C. elegans*, mouse, and human proteins involved in the calculation of the rate of PPI evolution in this study is, on average, 4.1×10^{-7} aa substitutions per protein per year (*SI Materials and Methods*). That is, ~1,558 aa substitutions, or ~5.0 per site, will happen in the time required for one PPI change in a protein.

Caveats. Although our yeast experiment has substantially increased the sample size for estimating the rate of PPI evolution, the number of PPIs examined is still small compared with the number of all PPIs in yeast (37). Thus, it is important to ask whether the PPIs that we studied are a representative sample of all yeast PPIs. For this purpose, we first plotted the frequency distribution of PPI degrees (i.e., the number of BioGRID-recorded Y2H-based PPIs per gene) for all 3,152 *Sce* genes that have one to one *Kwa* orthologs and the corresponding distribution for the 74 *Sce* genes involved in the 43 PPIs conserved between *Sce* and *Kwa*. Note that only PPIs among the 3,152 genes are counted to avoid the complication of PPI changes after gene duplication. We found that both distributions cover similarly large degree variations among genes, although our sample of 74 genes tended to have higher degrees than the 3,152 genes ($P = 5 \times 10^{-8}$, Mann–

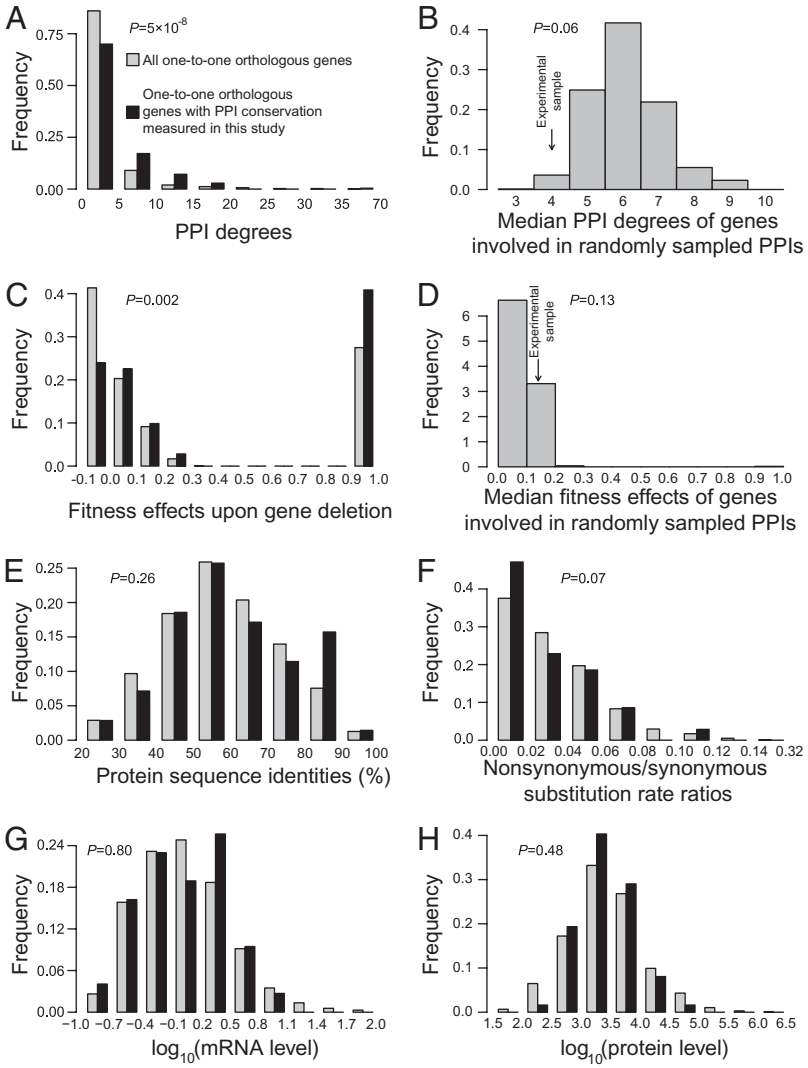

Fig. 4. Comparisons in various gene properties between all 3,152 *S. cerevisiae* genes that have one to one *K. waltii* orthologs (gray bars) and the 74 *S. cerevisiae* genes involved in the 43 PPIs measured for evolutionary conservation (black bars). (*A*) A comparison in PPI degree (i.e., the number of PPIs that a gene has with the rest of the 3,152 genes). (*B*) Frequency distribution of median PPI degrees of genes involved in 43 randomly sampled PPIs, which is derived from 1,000 simulations. (*C*) A comparison in gene importance, which is measured by the fitness reduction caused by gene deletion. (*D*) Frequency distribution of median importance of the genes involved in 43 randomly sampled PPIs, which is derived from 1,000 simulations. (*E*) A comparison in protein sequence identity between *S. cerevisiae* and *K. waltii*. (*F*) A comparison in the nonsynonymous to synonymous substitution rate ratio measured by comparing *S. cerevisiae* and *K. waltii* sequences. (*G*) A comparison in mRNA expression levels. (*H*) A comparison in protein expression levels. All *P* values are from Mann–Whitney tests except those in *B* and *D*, which are from simulation tests.

Whitney test) (Fig. 4*A*). This disparity, however, is not unexpected, because genes with higher degrees are more likely to be chosen when PPIs are randomly picked. To illustrate this point, we randomly sampled 43 PPIs from all of the PPIs among the 3,152 genes and calculated the median degree of the genes involved in the sampled PPIs. We repeated this process 1,000 times to obtain a frequency distribution of the median degree (Fig. 4*B*). Interestingly, the median degree of the 74 genes studied is even lower than that of randomly selected ones, although their difference is not significant ($P = 0.06$, simulation test).

We also plotted the frequency distribution of gene importance measured by the fitness reduction caused by the deletion of the gene for all 3,152 genes and the 74 genes that we studied. The 74 genes cover the whole range of gene importance, although they tend to be more important than average genes ($P = 0.002$) (Fig. 4*C*). This finding is expected, because PPI degree and gene importance are known to correlate positively with each other (43, 44). Nevertheless, there is no significant difference between the median importance of the 74 genes that we studied and the genes involved in the randomly selected PPIs of the above simulation ($P = 0.13$, simulation test) (Fig. 4*D*). Thus, the apparent bias in the degree and importance of the 74 genes that we studied (Fig. 4 *A* and *C*) is the byproduct of random PPI selection. Because the PPIs were randomly selected in our experiment, the above bias in some gene properties is unlikely to affect our estimation of the rate of PPI evolution. Furthermore, there is no indication that important genes or genes with higher degrees tend to have PPIs that are evolutionarily more conserved.

We also compared the frequency distributions for the 3,152 genes and the 74 studied genes in terms of protein sequence conservation (Fig. 4*E*) and nonsynonymous to synonymous substitution rate ratio (Fig. 4*F*), but we did not find significant differences. If the rate of PPI evolution is primarily determined by the rate of protein sequence evolution, our results suggest that our sample is unbiased for estimating the rate of PPI evolution.

Because mRNA and protein expression levels affect the evolution of protein structure, stability, and propensity for nonspecific protein interactions (6, 22, 45), we also compared our 74 studied genes with the 3,152 genes in terms of mRNA expression (Fig. 4*G*) and protein expression levels (Fig. 4*H*), but we did not find significant differences.

We also examined gene ontology (GO) (46) differences between the two groups of genes. Although one to three functional categories were found to be significantly depleted or enriched (at a false discovery rate of 5%) among the 74 genes for each of the three aspects of GO (cellular component, molecular function, and biological process) (47), the 74 genes are not limited to a small number of GO categories (Fig. 5). Furthermore, even for the GO categories with significant discrepancies between the 74 genes and 3,152 genes, the discrepancies are moderate when the entire distribution of genes across all GO categories is considered, and thus, they are unlikely to have a major impact on the estimation of the rate of PPI evolution (Fig. 5).

We found that at least 27 of 87 putative PPIs of *Sce* cannot be confirmed by our Y2H assay. Because we defined putative PPIs relatively rigorously, with the requirement that they had been reported by two Y2H experiments or one small-scale Y2H experiment, one may wonder why many of them cannot be confirmed in our Y2H assay. One reason is that a PPI may not be detectable by all variants of Y2H (40). Furthermore, our Y2H

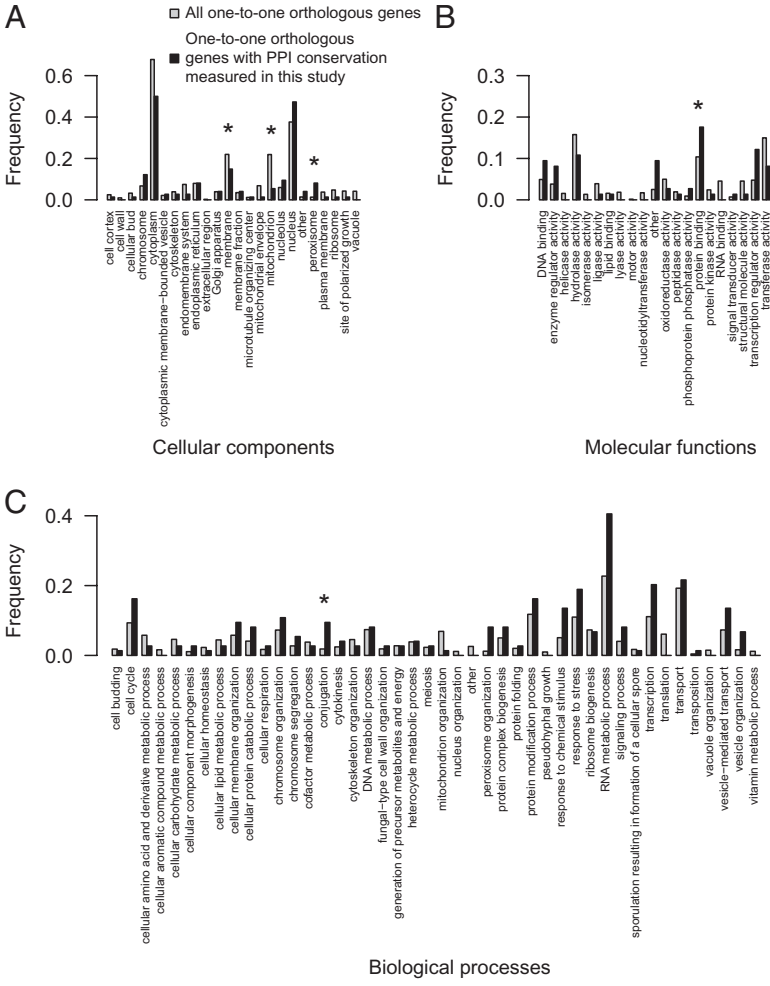

Fig. 5. Comparisons in gene ontology (GO) slim distributions between all 3,152 *S. cerevisiae* genes that have one to one *K. waltii* orthologs (gray bars) and the 74 *S. cerevisiae* genes involved in the 43 PPIs measured for evolutionary conservation (black bars). Bins with significant discrepancies at the false discovery rate of 5% are marked by asterisks.

assay uses three reporter genes, and the PPI will only be scored when all three genes are activated. This stringent design guards against false positives caused by spurious gene activation without PPI, which can happen occasionally. In fact, our validation rate is much higher than those reported in the literature (28, 42), presumably because of the small-scale nature of our experiment.

Because both *Sce* and *Kwa* PPIs were examined in *Sce* cells, one wonders whether our experimental design would cause an overestimation of the evolutionary rate of PPI because of the possibility that naturally interacting *Kwa* proteins may not interact well in *Sce* cells. This concern is unnecessary here, because we found no validated *Sce* PPIs whose *Kwa* orthologs do not interact in our Y2H assay.

Implications. In this work, we estimated the evolutionary rate of protein molecular function by measuring the conservation of PPIs between species, and we found the rate to be strikingly low in the absence of gene duplication. Our finding has a number of important implications. First, it suggests a high similarity in molecular function between one to one orthologs from even distantly related species. For instance, based on our estimated rate, an interaction between two human proteins is expected to be present between their respective one to one orthologs in mouse, fish, fly, worm, fungi, and plants, with probabilities of 98%, 89%, 79%, 77%, 71%, and 66%, respectively (Fig. 6 and *SI Materials and Methods*). Life is fundamentally conserved at molecular and cellular levels, because most biological processes at these levels are similar among divergent species (12). Given the prevalence and importance of PPIs in almost all cellular processes, the extreme conservation of PPIs is likely one of the bedrocks of the conservative nature of life. Note, however, that our Y2H assay is qualitative rather than quantitative. Hence, we cannot exclude the possibility that the strength of a protein interaction evolves much faster than the presence or absence of the interaction.

Second, although molecular functions of proteins are conserved in evolution, the physiological roles of proteins and their contributions to organismal fitness can change quickly and substantially, which is evident from frequent observations of the huge diversity in the phenotypic effect of orthologous gene deletions (48) and the great disparity in the dispensability of orthologous genes in different species (13, 49, 50). For instance, some mouse defects in blood vessel formation and yeast hypersensitivity to the hypercholesterolemia drug lovastatin are caused by mutations of orthologous genes (48). In another example, *Arabidopsis* orthologs of human genes implicated in Waardenburg syndrome (deafness and neutral crest anomalies) are involved in gravitropism (48). A systematic comparison between phenotypes of human and mouse mutations found that over 20% of mouse one to one orthologs of human essential genes are nonessential (49). However, if the molecular functions of one to one orthologous genes are highly conserved in evolution, which is suggested by this study, the molecular underpinnings of human disease may be studied in much simpler model organisms that do not even have the disease or relevant tissue/organ.

Third, previous analyses of high-throughput PPI data revealed a substantial amount of subfunctionalization and neofunctionalization after gene duplication (35, 36). The contrast between these results and the present finding in one to one orthologous genes suggests that the majority of molecular function changes in protein evolution are associated with gene duplication. However, because of the unreliability of high-throughput PPI data, previous results on duplicate genes (35, 36) should be verified in the future. It would be highly desirable to conduct a study on duplicate genes similar to the present one to quantify the difference in the rate of protein function evolution in the presence and absence of gene duplication (51); this is similar to a study that has been conducted recently on the rate of protein subcellular relocalization (52). In this respect, the *Sce* and *Kwa* comparison will also be appropriate, because *Sce* retains ~500 pairs of duplicate genes generated by a whole-genome duplication that occurred since the separation of *Sce* from *Kwa* (38, 53).

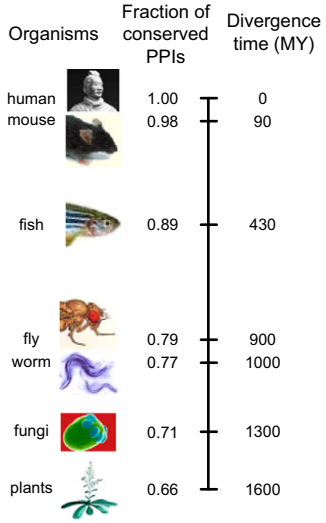

Fig. 6. Fraction of human PPIs expected to be conserved in various widely used model organisms based on previously estimated divergence times (9) and our estimated rate of PPI evolution.

Materials and Methods

Identification of Putative PPIs for Experimental Tests. Gene sequences of *Sce* were downloaded from *Saccharomyces* Genome Database (SGD; http://yeastgenome.org/), and gene sequences of *Kwa* were downloaded from the supplementary materials of ref. 53 (http://www.nature.com/nature/journal/v428/n6983/extref/nature02424-s1.htm). To identify one to one orthologous genes between the two species, we combined the genomes of *Sce* and *Kwa*, conducted all against all BlastP searches with an E-value cutoff of 1×10^{-20}, and removed self-hits. If (*i*) proteins A and B are reciprocal best hits in the above search, (*ii*) they do not belong to the same species, and (*iii*) the aligned region of the two proteins is longer than 80% of the shorter one, we classify them as a pair of one to one orthologs.

Protein interaction data in *Sce* were downloaded from BioGRID (http://thebiogrid.org/) at the beginning of our study in 2006 (GRID-ORGANISM-Saccharomyces_cerevisiae-2.0.20.tab.txt). Among the *Sce* proteins that have one to one *Kwa* orthologs, we identified 355 PPIs with at least two independent Y2H reports. Among the remaining PPIs, we identified 481 that had one small-scale Y2H report (i.e., with <30 PPIs per report). These two sets of PPIs were treated as putative *Sce* PPIs subject to further analysis.

To test the interaction between the *Kwa* orthologs of a pair of *Sce* proteins that are known to interact, we need to clone the *Kwa* orthologs. To reduce the difficulty in gene cloning, we selected *Sce* PPIs for which the *Kwa* orthologs have ORFs between 400 and 1,600 nt long. We excluded intron-containing genes so that the full coding region could be amplified from genomic DNA in one piece. We also eliminated genes incompatible with the restriction sites available on plasmids pGADT7 or pGBKT7. The majority of *Sce* PPIs selected had been discovered in at least two independent Y2H reports (73%), whereas a minority of them (27%) had been identified in only one small-scale Y2H experiment (Table S1).

Y2H Assays. After cloning the *Kwa* genes, we performed the Y2H assay using the Matchmaker GAL4 Two-Hybrid System 3 (Clonetch). The two PPI partners were cloned into pGADT7 and pGBKT7 plasmids, respectively, through the following procedure. We first amplified the genes by PCR using PfuUltra high-fidelity polymerase (Stratagene) to minimize PCR errors during the amplification. We purified the PCR products (Qiagen), digested them with two of five restriction enzymes (*EcoRI*, *BamHI*, *NdeI*, *PstI*, and *ClaI*), and cloned them into pGADT7 and/or pGBKT7 by T4 DNA ligase (Promega). We transformed the ligation products into TOP10 chemically competent cells (Invitrogen), extracted the plasmid (Qiagen), and confirmed the clones by DNA sequencing at the University of Michigan DNA Sequencing Core.

We transformed the two plasmids into *S. cerevisiae* AH109 cells (*MATa*, *trp1-901*, *leu2-3*, *112*, *ura3-52*, *his3-200*, *gal4Δ*, *gal80Δ*, *LYS2::GAL1*$_{UAS}$-*GAL1*$_{TATA}$-*HIS3*, *GAL2*$_{UAS}$-*GAL2*$_{TATA}$-*ADE2*, *URA3::MEL1*$_{UAS}$-*MEL1*$_{TATA}$-*lacZ*, *MEL1*), which were selected on synthetic dextrose plates with leucine and tryptophan dropped out (SD-Leu-Trp). The colonies were also pinned onto synthetic dextrose plates with adenine, histidine, leucine, and tryptophan

EVOLUTION

dropped out and with 20 μg/mL X-α-gal added (SD-Ade-His-Leu-Trp/X-α-gal). Because *MEL1* encodes a secreted enzyme α-galactosidase, its presence can be assayed directly on X-α-gal–containing plates without cell lysis. If the transformed yeast can grow on the dropout plates (SD-Ade-His-Leu-Trp/X-α-gal) and appears blue, the proteins are considered to be interacting with each other. For a strain to grow on SD-Ade-His-Leu-Trp/X-α-gal and be blue, all three reporter genes (*HIS3*, *ADE2*, and *MEL1* under promoters *pGAL1*, *pGAL2*, and *pMEL1*, respectively) must be activated. Hence, our Y2H assay is quite stringent. The high stringency implies that some weak PPIs may not be detected by our Y2H assay. Self-activation was tested by cotransformation of a gene-containing plasmid (pGADT7 or pGBKT7) and an empty plasmid (pGBKT7 or pGADT7, respectively). We excluded a gene pair from further consideration if either gene showed self-activation.

We selected 10 random protein pairs and tested their interaction [(Kwa_10129, Kwa_23895), (Kwa_12079, Kwa_9492), (Kwa_12518, Kwa_5419), (Kwa_13638, Kwa_23894), (Kwa_15314, Kwa_15321), (Kwa_16145, Kwa_18622), (Kwa_1973, Kwa_21767), (Kwa_2064, Kwa_10342), (Kwa_2079, Kwa_17326), and (Kwa_21273, Kwa_23528)] in our Y2H assay, and none of them showed PPI, suggesting a low false-positive rate in our experiment.

The potential PPI pairs that did not show positive results in *Kwa* by our Y2H experiment were examined in *Sce*. We cloned the corresponding *Sce* genes into the same plasmids (pGBKT7 and pGADT7) and then conducted the Y2H assay as described above. We also randomly selected a subset of positive *Kwa* PPIs and examined whether their *Sce* orthologs interact in our Y2H assay.

Examination of Potential Biases of the Experimentally Studied Genes. *Sce* PPI information was from BioGRID. Only Y2H PPIs between *Sce* genes that have one to one *Kwa* orthologs were counted in PPI degree calculations. The fitness effect of gene deletion was obtained from ref. 54. The mRNA expression levels were from ref. 55, and the protein expression levels were from ref. 56. One to one orthologous genes between *Sce* and *Kwa* were aligned by ClustalW (57), and the nonsynonymous to synonymous substitution rate ratios were calculated by Codeml in Phylogenetic Analysis by Maximum Likelihood (PAML) (58). To examine the impact of random sampling of PPIs on the bias of the selected genes, we randomly sampled 43 PPIs from all PPIs among the 3,152 *Sce* genes that have one to one *Kwa* orthologs and repeated this process 1,000 times.

ACKNOWLEDGMENTS. We thank Meg Bakewell, Calum Maclean, Jian-Rong Yang, and two anonymous reviewers for valuable comments. H.X. was supported in part by a fellowship from the Chinese government. This work was supported by a research grant from the National Institutes of Health (to J.Z.).

1. Zukerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. *Evolving Genes and Proteins*, eds Bryson V, Vogel HJ (Academic, New York), pp 97–166.
2. Kimura M (1968) Evolutionary rate at the molecular level. *Nature* 217:624–626.
3. Li W (1997) *Molecular Evolution* (Sinauer, Sunderland, MA).
4. Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics* (Oxford University Press, New York).
5. Koonin E, Galperin M (2003) *Sequence—Evolution—Function: Computational Approaches in Comparative Genomics* (Kluwer, Boston).
6. Drummond DA, Wilke CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134:341–352.
7. Wolf YI, Novichkov PS, Karev GP, Koonin EV, Lipman DJ (2009) The universal distribution of evolutionary rates of genes and distinct characteristics of eukaryotic genes of different apparent ages. *Proc Natl Acad Sci USA* 106:7273–7280.
8. Doolittle RF (2010) The roots of bioinformatics in protein evolution. *PLoS Comput Biol* 6:e1000875.
9. Hedges SB, Dudley J, Kumar S (2006) TimeTree: A public knowledge-base of divergence times among organisms. *Bioinformatics* 22:2971–2972.
10. King JL, Jukes TH (1969) Non-Darwinian evolution. *Science* 164:788–798.
11. Kimura M (1983) *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, UK).
12. Zhang J (2010) Evolutionary genetics: Progress and challenges. *Evolution Since Darwin: The First 150 Years*, eds Bell MA, Futuyma DJ, Eanes WF, Levinton JS (Sinauer, Sunderland, MA), pp 87–118.
13. Zhang J, He X (2005) Significant impact of protein dispensability on the instantaneous rate of protein evolution. *Mol Biol Evol* 22:1147–1155.
14. Liao BY, Scott NM, Zhang J (2006) Impacts of gene essentiality, expression pattern, and gene compactness on the evolutionary rate of mammalian proteins. *Mol Biol Evol* 23:2072–2080.
15. Liao BY, Weng MP, Zhang J (2010) Impact of extracellularity on the evolutionary rate of mammalian proteins. *Genome Biol Evol* 2:39–43.
16. Pál C, Papp B, Hurst LD (2001) Highly expressed genes in yeast evolve slowly. *Genetics* 158:927–931.
17. Jordan IK, Rogozin IB, Wolf YI, Koonin EV (2002) Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res* 12:962–968.
18. Hirsh AE, Fraser HB (2001) Protein dispensability and rate of evolution. *Nature* 411: 1046–1049.
19. Wang Z, Zhang J (2009) Why is the correlation between gene importance and gene evolutionary rate so weak? *PLoS Genet* 5:e1000329.
20. Liao BY, Weng MP, Zhang J (2010) Contrasting genetic paths to morphological and physiological evolution. *Proc Natl Acad Sci USA* 107:7353–7358.
21. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW (2002) Evolutionary rate in the protein interaction network. *Science* 296:750–752.
22. Yang JR, Zhuang SM, Zhang J (2010) Impact of translational error-induced and error-free misfolding on the rate of protein evolution. *Mol Syst Biol* 6:421.
23. Bakewell MA, Shi P, Zhang J (2007) More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proc Natl Acad Sci USA* 104:7489–7494.
24. Clark AG, et al. (2003) Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* 302:1960–1963.
25. Nielsen R, et al. (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol* 3:e170.
26. Nozawa M, Suzuki Y, Nei M (2009) Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. *Proc Natl Acad Sci USA* 106:6700–6705.
27. Fields S, Song O (1989) A novel genetic system to detect protein–protein interactions. *Nature* 340:245–246.
28. Yu H, et al. (2008) High-quality binary protein interaction map of the yeast interactome network. *Science* 322:104–110.
29. Li S, et al. (2004) A map of the interactome network of the metazoan C. elegans. *Science* 303:540–543.
30. Giot L, et al. (2003) A protein interaction map of Drosophila melanogaster. *Science* 302:1727–1736.
31. Rual JF, et al. (2005) Towards a proteome-scale map of the human protein–protein interaction network. *Nature* 437:1173–1178.
32. Ohno S (1970) *Evolution by Gene Duplication* (Springer, Berlin).
33. Zhang J (2003) Evolution by gene duplication: An update. *Trends Ecol Evol* 18:292–298.
34. Conant GC, Wolfe KH (2008) Turning a hobby into a job: How duplicated genes find new functions. *Nat Rev Genet* 9:938–950.
35. He X, Zhang J (2005) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* 169:1157–1164.
36. Wagner A (2001) The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Mol Biol Evol* 18:1283–1292.
37. He X, Zhang J (2009) On the growth of scientific knowledge: Yeast biology as a case study. *PLoS Comput Biol* 5:e1000320.
38. Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387:708–713.
39. Braun P, et al. (2009) An experimentally derived confidence score for binary protein-protein interactions. *Nat Methods* 6:91–97.
40. Chen YC, Rajagopala SV, Stellberger T, Uetz P (2010) Exhaustive benchmarking of the yeast two-hybrid system. *Nat Methods* 7:667–668.
41. Matthews LR, et al. (2001) Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs." *Genome Res* 11:2120–2126.
42. Ravasi T, et al. (2010) An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* 140:744–752.
43. Jeong H, Mason SP, Barabási AL, Oltvai ZN (2001) Lethality and centrality in protein networks. *Nature* 411:41–42.
44. He X, Zhang J (2006) Why do hubs tend to be essential in protein networks? *PLoS Genet* 2:e88.
45. Vavouri T, Semple JI, Garcia-Verdugo R, Lehner B (2009) Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell* 138:198–208.
46. Ashburner M, et al. (2000) Gene ontology: Tool for the unification of biology. *Nat Genet* 25:25–29.
47. Al-Shahrour F, et al. (2007) FatiGO +: A functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. *Nucleic Acids Res* 35:W91–W96.
48. McGary KL, et al. (2010) Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc Natl Acad Sci USA* 107:6544–6549.
49. Liao BY, Zhang J (2008) Null mutations in human and mouse orthologs frequently result in different phenotypes. *Proc Natl Acad Sci USA* 105:6987–6992.
50. Kim DU, et al. (2010) Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. *Nat Biotechnol* 28:617–623.
51. Studer RA, Robinson-Rechavi M (2009) How confident can we be that orthologs are similar, but paralogs differ? *Trends Genet* 25:210–216.
52. Qian W, Zhang J (2009) Protein subcellular relocalization in the evolution of yeast singleton and duplicate genes. *Genome Biol Evol* 1:198–204.
53. Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature* 428:617–624.
54. Deutschbauer AM, et al. (2005) Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* 169:1915–1925.
55. Holstege FC, et al. (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95:717–728.
56. Ghaemmaghami S, et al. (2003) Global analysis of protein expression in yeast. *Nature* 425:737–741.
57. Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
58. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.