Conserved pathways within bacteria and yeast as revealed by global protein network alignment

Brian P. Kelley*, Roded Sharan†, Richard M. Karp†, Taylor Sittler*, David E. Root*, Brent R. Stockwell*, and Trey Ideker*‡

*Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142; and †International Computer Science Institute, 1947 Center Street, No. 600, Berkeley, CA 94704

Contributed by Richard M. Karp, July 25, 2003

We implement a strategy for aligning two protein–protein interaction networks that combines interaction topology and protein sequence similarity to identify conserved interaction pathways and complexes. Using this approach we show that the protein–protein interaction networks of two distantly related species, *Saccharomyces cerevisiae* and *Helicobacter pylori*, harbor a large complement of evolutionarily conserved pathways, and that a large number of pathways appears to have duplicated and specialized within yeast. Analysis of these findings reveals many well characterized interaction pathways as well as many unanticipated pathways, the significance of which is reinforced by their presence in the networks of both species.

Evolution is driven by biological variation at many levels. Mutations and rearrangements in genomic DNA lead to changes in protein structures, abundances, and modification states. Variations at the protein level, in turn, impact how proteins interact with one another, with DNA, and with small molecules to form signaling, regulatory, and metabolic networks. Changes in network organization have sweeping implications for cellular function, tissue-level responses, and the behavior and morphology of whole organisms.

Gene and protein sequences have long received the most attention as metrics for evolutionary change, both because they represent a fundamental level of biological variation and because they are readily available through automated sequencing technology. However, recent technological advances also enable us to characterize networks of protein interactions. Protein interactions are crucial to cellular function both in assembling protein complexes and in signal transduction cascades. Among the most direct and systematic methods for measuring protein interactions are coimmunoprecipitation (1) and the two-hybrid system (2), which have defined large protein–protein interaction networks for organisms including *Saccharomyces cerevisiae* (3–5), *Helicobacter pylori* (6), and *Caenorhabditis elegans* (7). Although the quality of data from these experiments has been mixed, pooling of multiple studies and integration with other data types such as gene expression have been used to reduce the number of false-positive interactions (8).

The rapid growth of protein network information raises a host of new questions in evolutionary and comparative biology. Given that protein sequences and structures are conserved in and among species, are networks of protein interactions conserved as well? Is there some minimal set of interaction pathways required for all species? Can we measure evolutionary distance at the level of network connectivity rather than at the level of DNA or protein sequence? Mounting evidence suggests that conserved protein interaction pathways indeed exist and may be ubiquitous: For example, proteins in the same pathway are typically present or absent in a genome as a group (9), and several hundred protein–protein interactions in the yeast network have also been identified for the corresponding protein orthologs in worms (10).

To explore interspecies pathway conservation on a global scale, we performed a series of whole-network comparisons using the protein–protein interaction networks of the budding yeast *S. cerevisiae* and the bacterial pathogen *H. pylori*. Comparative network analysis has proven powerful in a number of related domains including metabolic pathway analysis (11–14), motif finding (15), and correlation of biological networks with gene expression (16). Here we systematically search for and prioritize conserved interaction pathways in yeast vs. bacteria, yeast vs. yeast, and yeast vs. specific "queries" formulated to uncover homologous mitogen-activated protein kinase (MAPK) signaling and ubiquitin ligation machinery.

Methods

Network Comparison Overview. We developed an efficient computational procedure for aligning two protein interaction networks to identify their conserved interaction pathways.§ This procedure, which we named PATHBLAST because of its conceptual similarity to sequence alignment algorithms such as BLAST (17), searches for high-scoring pathway alignments involving two paths, one from each network, in which proteins of the first path $\langle$A, B, C, D, . . . $\rangle$ are paired with putative homologs occurring in the same order in the second path $\langle$a, b, c, d, . . . $\rangle$ (Fig. 1*a*). Evolutionary variations and experimental errors in pathway structure are accommodated by allowing "gaps" and "mismatches" (see also ref. 14). A gap occurs when a protein interaction in one path skips over a protein in the other, whereas a mismatch occurs when aligned proteins do not share sequence similarity. Because of space limitations, only abbreviated methods are given in the following sections; full descriptions are available in *Supporting Materials and Methods* and Figs. 5 and 6, which are published as supporting information on the PNAS web site, www.pnas.org.¶

Global Alignment and Scoring. To perform the alignment, the two networks are combined into a global alignment graph (Fig. 1*b*) in which each vertex represents a pair of proteins (one from each network) having at least weak sequence similarity (BLAST E value $\leq 10^{-2}$) and each edge represents a conserved interaction, gap, or mismatch. A path through this graph represents a pathway alignment between the two networks. We formulate a log probability score $S(P)$ that decomposes over the vertices v and edges e of a path P through the global alignment graph

Abbreviation: MAPK, mitogen-activated protein kinase.

‡To whom correspondence should be sent at the present address: University of California at San Diego, Department of Bioengineering, 9500 Gilman Drive, La Jolla, CA 92093. E-mail: trey@bioeng.ucsd.edu.

§The term "pathway" has been used broadly within various molecular biological contexts to refer to biochemical reaction chains, signal transduction cascades, gene regulatory systems, or other sequences of biomolecular events. Here a pathway refers to a sequence of protein–protein interactions forming a connected path in the network.

¶We have also explored methods for identifying conserved subnetworks as opposed to linear paths (see Fig. 7, which is published as supporting information on the PNAS web site); choosing which approach is most desirable remains an open problem and depends on issues of computational efficiency and whether protein complexes or sequential pathways such as signal transduction or regulatory cascades are of highest interest.

www.pnas.org/cgi/doi/10.1073/pnas.1534710100

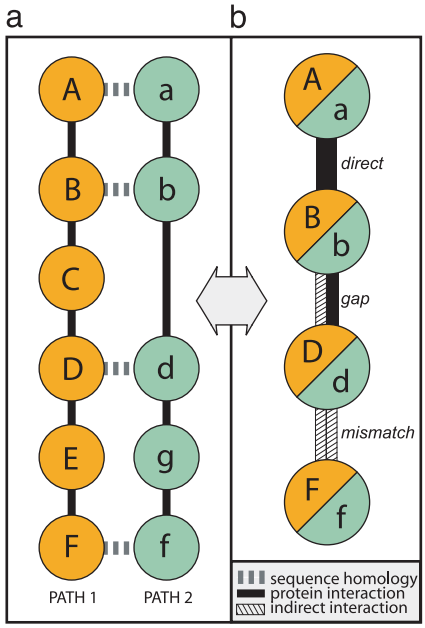

Fig. 1. Example pathway alignment and merged representation. (*a*) Vertical solid lines indicate direct protein–protein interactions within a single pathway, and horizontal dotted lines link proteins with significant sequence similarity (BLAST E value $\leq E_{\text{cutoff}}$). An interaction in one pathway may skip over a protein in the other (protein *C*), introducing a ''gap.'' Proteins at a particular position that are dissimilar in sequence (E value $> E_{\text{cutoff}}$, proteins *E* and *g*) introduce a ''mismatch.'' The same protein pair may not occur more than once per pathway, and neither gaps nor mismatches may occur consecutively. (*b*) Pathways are combined as a global alignment graph in which each node represents a homologous protein pair and links represent protein interaction relationships of three types: direct interaction, gap (one interaction is indirect), and mismatch (both interactions are indirect).

$$S(P) = \sum_{v \in P} \log_{10} \frac{p(v)}{p_{\text{random}}} + \sum_{e \in P} \log_{10} \frac{q(e)}{q_{\text{random}}},$$

where $p(v)$ is the probability of true homology within the protein pair represented by v, given its pairwise protein sequence similarity expressed as a BLAST E value, and $q(e)$ is the probability that the protein–protein interactions represented by e are real, i.e., not false-positive errors. The background probabilities p_{random} and q_{random} are the expected values of $p(v)$ and $q(e)$ over all vertices and edges in the global alignment graph. Protein sequence alignments and associated E values were computed by using BLAST 2.0 (17) with parameters $b = 0$, $e = 1 \times 10^6$, $f =$ ''C;S'', and $v = 6 \times 10^5$. Unalignable proteins were assigned a maximum E value of 5.

Optimal Pathway Alignments and Significance. For acyclic graphs, the highest-scoring path of length L can be found in linear time by using a procedure based on dynamic programming as described in *Supporting Materials and Methods*. Because the global alignment graph may contain cycles, we first generate a sufficient number, $5L!$, of acyclic subgraphs by random removal of edges from the global alignment graph and then aggregate the results of running dynamic programming on each.

Because conserved regions of the network could be highly interconnected (e.g., a conserved protein complex), it was sometimes possible to identify a large number of distinct paths involving the same small set of proteins. Rather than enumerate each of these, we used PATHBLAST in consecutive stages. For each stage k, we recorded the set of 50 highest-scoring pathway alignments (with average score $\langle S_k \rangle$) and then removed their vertices and edges from the alignment graph before the next stage. The p value of each stage was assessed by comparing $\langle S_k \rangle$ to the distribution of average scores $\langle S_1 \rangle$ observed over 100 random global alignment graphs (constructed as per Table 1) and assigned to every conserved network region resulting from that stage (Figs. 2 and 3). The p values for pathway queries (Fig. 4) were computed individually, not in stages, by comparing each pathway-alignment score to the best scores achieved over 100 random alignment graphs involving the query and target (yeast) network.

Software Availability. PATHBLAST is available at www.pathblast.org.

Results

Yeast vs. Bacteria: Orthologous Pathways Between the Networks of Two Species. We first performed a global alignment between the protein–protein interaction networks of yeast (*S. cerevisiae*) and bacteria (*H. pylori*). To construct the yeast network, we downloaded the 14,489 interactions among 4,688 yeast proteins present in the Database of Interacting Proteins (18) as of November 2002. These interactions represented a pooled collection of several data sets derived through systematic coimmunoprecipitation and two-hybrid studies. The *H. pylori* network was also obtained from the Database of Interacting Proteins and represented a single two-hybrid study identifying 1,465 interactions among 732 proteins (6). Protein sequences for both species were obtained from the Protein Information Resource (19).

Table 1 compares the bacterial/yeast global alignment graph to those that resulted if the protein networks were randomized by permuting the protein names. Both the graph size and the best pathway-alignment scores were significantly larger for real than for random data, suggesting that the two species shared conserved interaction pathways. Surprisingly, conservation of direct interaction pairs between the yeast and bacterial networks was rare (only 7 direct edges vs. 2.5 ± 1.9 in random data, probably due to low coverage or quality of interactions). However, the use

Table 1. Combining protein networks as a global alignment graph

	Vertices (homologs)	Edges				CPU, min	Score	
		Total	Direct	Gap	Mismatch		Best*	Best 50†
Yeast vs. *H. pylori* ($E_{\text{cutoff}} = 10^{-2}$)	829	2,036	7	260	1,769	0.38	8.1	7.5
Random: mean ± SD		509.0 ± 128.0	2.5 ± 1.9	68.8 ± 23.8	437.7 ± 110.3	0.4 ± 0.02	6.1 ± 0.8	4.8 ± 0.7
Yeast vs. yeast ($E_{\text{cutoff}} = 10^{-10}$)	5,593	1,389	1,389	N/A	N/A	7.08	11.9	11.0
Random: mean ± SD		62.3 ± 29.4	62.3 ± 29.4	N/A	N/A	6.9 ± 0.2	−4.1 ± 9.5	−15.3 ± 6.5

Protein interaction networks drawn from either yeast and *H. pylori* or yeast and yeast were merged as a global alignment graph: The resulting numbers of vertices and edges are given along with the CPU time required for merging. Also shown is the best pathway alignment score and the average of the best 50 scores achieved in the graph for paths of four vertices. Alignment graphs were compared to random graphs constructed by permuting the protein names on each network before merging (mean ± SD for 100 permutations). N/A, not applicable.

*$p = 0.006$ and 0.05 for yeast vs. *H. pylori* and yeast vs. yeast, respectively.

†$p = 1.7 \times 10^{-5}$ and 2.4×10^{-5} for yeast vs. *H. pylori* and yeast vs. yeast, respectively.

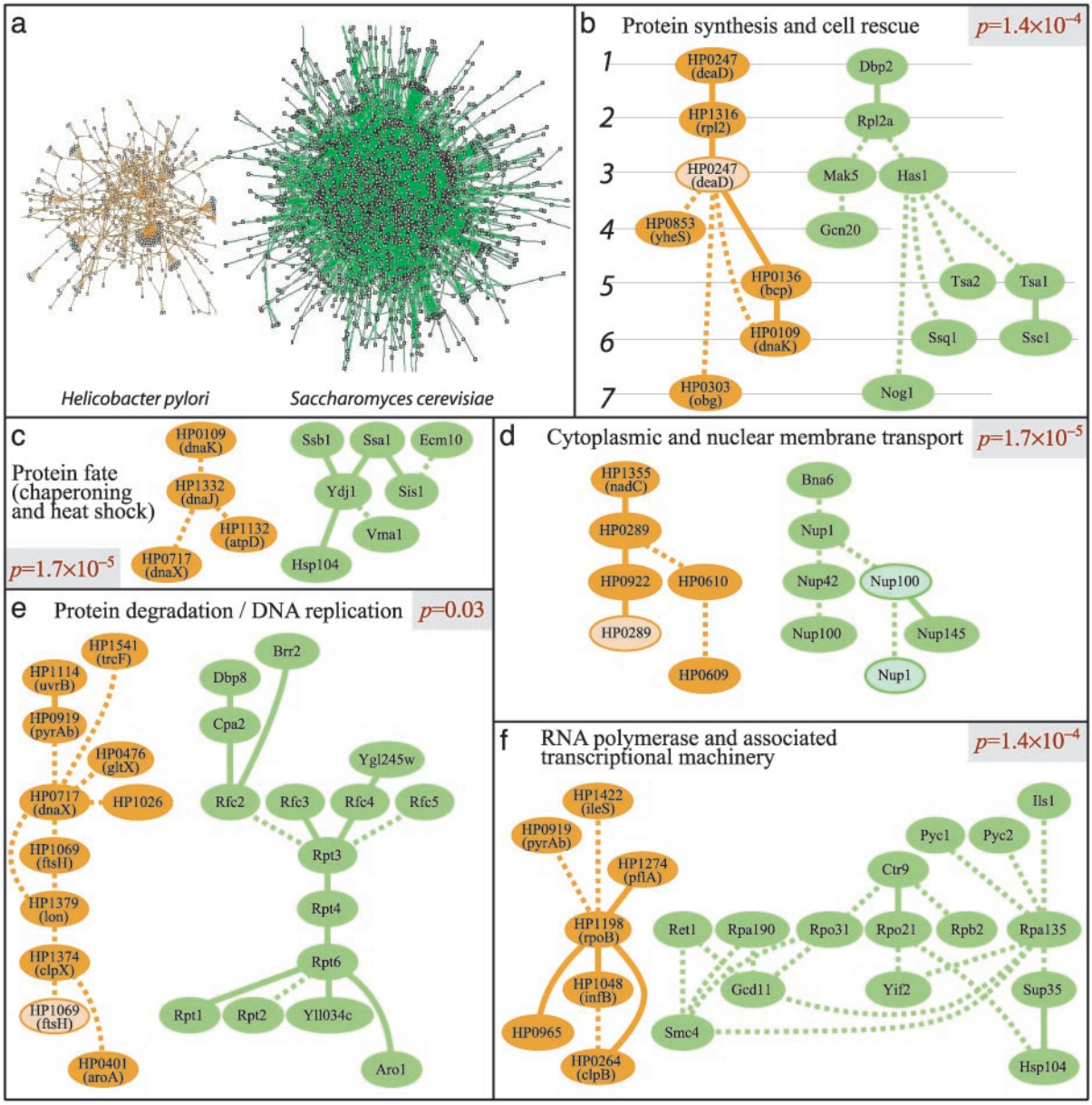

Fig. 2. Top-scoring pathway alignments between bacteria and yeast. (*a*) The protein–protein interaction networks of *H. pylori* (orange network) and *S. cerevisiae* (green network) were globally aligned to reveal conserved network regions (*b–f*). Proteins with above-threshold sequence similarity are placed on the same row of the pathway alignment (e.g., deaD and Dbp2 in row 1 of *b*). Direct protein interactions appear as solid links, and gaps or mismatches are dotted. Proteins recurring within a region due to multiple sequence homologies (e.g., deaD in *b*) are lighter in color. *P* values were computed by using random graphs as described in *Methods*.

of pathway gaps and mismatches allowed us to detect larger regions of the network that were generally conserved even when direct interactions were not.

We analyzed the global alignment graph to select the 150 highest-scoring pathway alignments of length four (four proteins per path), corresponding to a level of significance of $p \leq 0.05$ vs. random networks. By combining all overlapping pathway alignments, we found that each of the 150 fell into one of five connected network regions shown in Fig. 2 *b–f*. For instance, Fig. 2*b* involved the union of six paths: two were yeast Dbp2-Rpl2A-Mak5-Gcn20 (vs. *H. pylori* deaD-rpl2-deaD-yheS) and Rpl2A-Has1-Tsa1-Sse1 (vs. rpl2-deaD-bcp-dnaK). A total of 4.1% and 1.2% of proteins in the *H. pylori* and *S. cerevisiae* protein networks were included in a high-scoring pathway alignment.

As validation that pathway alignments corresponded to specific conserved cellular functions, we found that network regions were significantly enriched for particular protein functional categories from the Munich Information Center for Protein Sequences (http://mips.gsf.de) *S. cerevisiae* and the Institute for Genomic Research (www.tigr.org) *H. pylori* databases at a level of $p \leq 0.005$ using the hypergeometric test. Functions associated with each region included protein synthesis and cell rescue (Fig. 2*b*), protein fate and targeting (Fig. 2*c*), cell envelope and nuclear transport (Fig. 2*d*), proteolytic degradation (Fig. 2*e*), and rRNA transcription (Fig. 2*f*); further details are provided in Fig. 8, which is published as supporting information on the PNAS web site.

Yeast vs. Yeast: Paralogous Pathways Within the Network of a Single Species. In addition to identifying homologous features between the protein networks of yeast and bacteria, we also searched

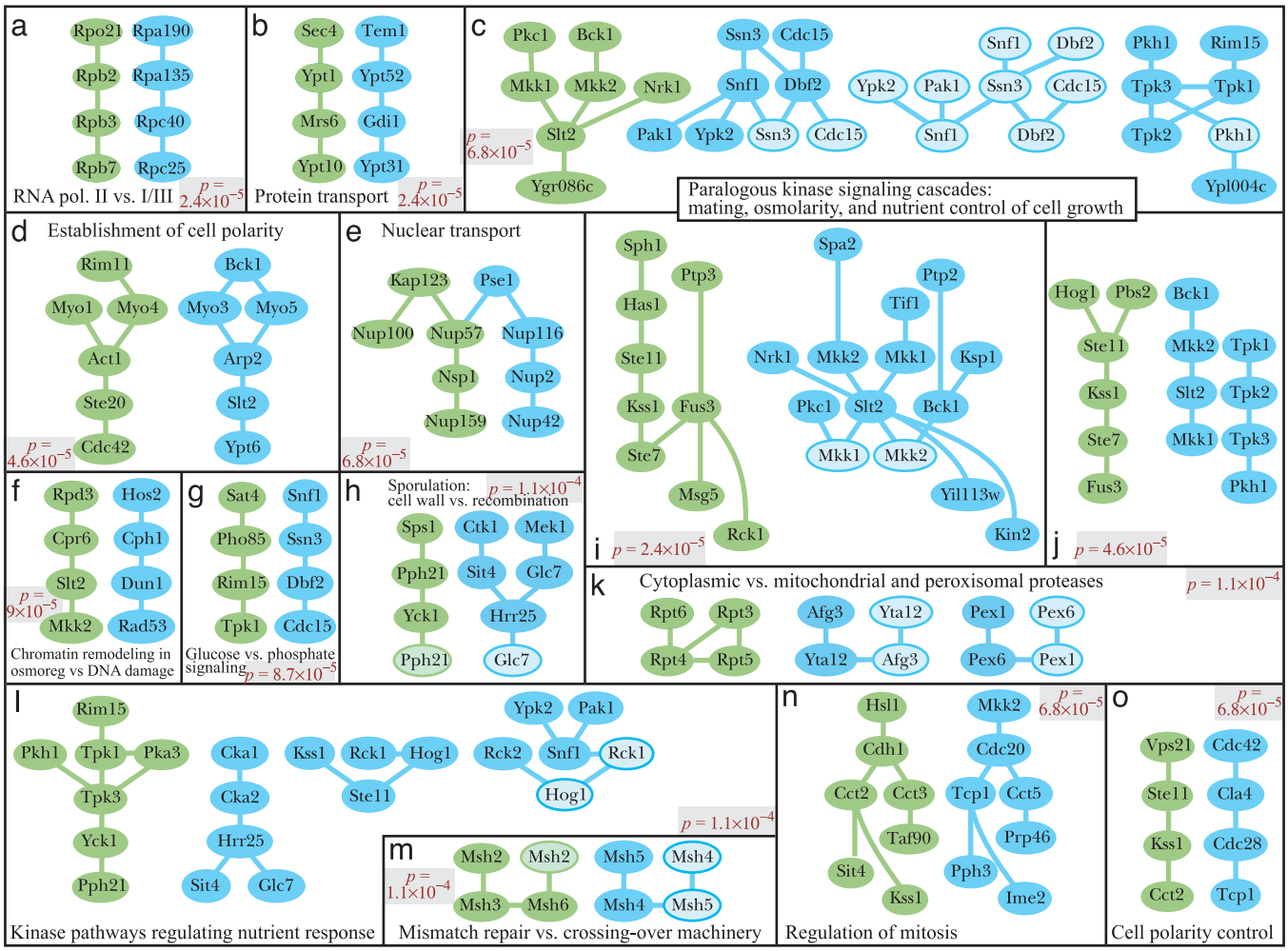

Fig. 3. Paralogous pathways within yeast. To find pathways conserved within yeast, the protein interaction network from *S. cerevisiae* (green pathways) was compared against a copy of itself (blue pathways). High-scoring pathway alignments are displayed as described for Fig. 2, with cellular functions indicated.

within each network individually to identify its potentially paralogous pathways; that is, pathways with proteins and interactions that have been duplicated one or more times in the course of evolution. Such an approach is akin to performing an "all vs. all" BLAST of sequences encoded by a single genome to elucidate gene families. To explore this procedure in the context of yeast, we constructed a global alignment graph by merging the yeast protein interaction network with an identical copy of itself. Because the resulting graph was potentially much larger than for bacteria/yeast (see Table 1), we could afford to be more restrictive: Vertices were defined as protein pairs with BLAST E values $\leq 10^{-10}$, with only direct edges permitted (no gaps or mismatches). To ensure that pathway alignments occurred between two distinct network regions and to avoid aligning a path with its exact copy, proteins were not allowed to pair with themselves or their network neighbors.

We analyzed the yeast/yeast alignment graph to obtain the 300 highest-scoring pathway alignments of length four, corresponding to a level of significance of $p \leq 0.0001$. These are shown in Fig. 3 *a–o*, with overlapping pathways grouped into connected network regions as shown in Fig. 2. Several regions involve alignments between protein complexes known to be distinct (i.e., noninteracting) but homologous in function, confirming that the approach is capable of identifying paralogous network structures. For example, Fig. 3*a* shows four subunits of the RNA polymerase II complex (green path) aligned against those of the RNA polymerase I and III complexes (blue path), with polymerase I and III linked via the shared Rpc40 subunit. Fig. 3*k* shows an alignment among three AAA heteromeric complexes with separate subcellular localizations: subunits of the cytosolic 26S proteasome (Rpt3-6), the mitochondrial AAA protease complex (Afg3 and Yta12), and the Pex1/6 complex, thought to function in protein disassembly before peroxisomal import (20). Likewise, although the aligned complexes shown in Fig. 3*m* both have DNA-binding activity, they act in two distinct processes: Msh2/3/6 is involved in mismatch repair during meiosis and vegetative growth, whereas Msh4/5 facilitates crossing over during homologous recombination and is specific to meiosis (21).

Interrogating the Protein Network with Pathway Queries. Finally, although an entire network-vs.-network comparison was invaluable for cataloging all the homologous pathways between and within organisms, we also queried a single protein network with specific pathways of interest. Use of PATHBLAST in this mode is similar to using BLAST to interrogate a sequence database with a short nucleotide or amino acid sequence query. As a test of this approach, we queried the *S. cerevisiae* protein network with a classic MAPK pathway associated with the filamentation response, consisting of a MAPK (Ste11), a MAPK kinase (Ste7), and a MAPK kinase kinase (Kss1). MAPK pathways transmit incoming signals to the nucleus through activation cascades in which each kinase phosphorylates the next one downstream.

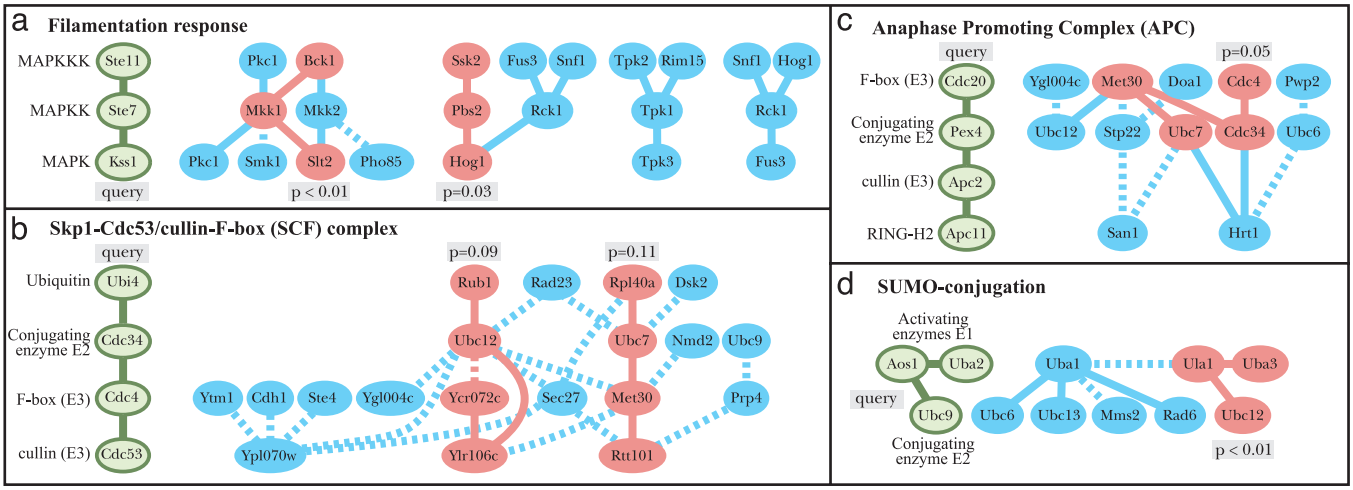

Fig. 4. Querying the yeast network with specific pathways. Pathway queries are shown for the kinase cascade involved in filamentation (*a*) and three ubiquitin-like conjugating systems: the SCF complex (*b*), the APC (*c*), and the SUMO complex (*d*). The display follows the conventions of Figs. 2 and 3, with the highest-scoring alignments indicated in red. Queries of this type may be submitted online at www.pathblast.org.

PATHBLAST identified two other well known MAPK pathways as the highest-scoring hits (the low- and high-osmolarity response pathways Bck1-Mkk1-Slt2 and Ssk2-Pbs2-Hog1), indicating that the algorithm was sufficiently sensitive and specific to identify known paralogous pathways.

We repeated this strategy to search for new components of the cellular ubiquitin and ubiquitin-like conjugation machinery. Ubiquitin targets proteins for degradation by the proteasome and modifies different sets of proteins through distinct pathways, some of which are unknown (22). Two well characterized ubiquitin conjugation systems were used as queries: the Skp1-Cdc53/cullin-F-box (SCF) complex (Fig. 4*b*) and the anaphase-promoting complex (APC; Fig. 4*c*); a third query was based on SUMO, a ubiquitin-like protein that is ligated to protein substrates but does not induce protein degradation (23) (Fig. 4*d*). Several of these queries aligned with components of a paralogous Rub1-conjugating complex (24) (Rub1-Ubc12-Ula1-Uba3; Fig. 4 *b* and *d*) and suggested new F-box and cullin-like proteins (Ycr072c and Ylr106c) for the Rub1 ligation system. Also identified were a putative ubiquitin-conjugation pathway involving Rpl40a, a fusion protein with ubiquitin at its N terminus (25), and several additional pathways involved in lower-scoring alignments. Thus, short pathway-based queries using PATHBLAST are capable of identifying both known and potentially novel paralogous pathways within an organism.

Discussion

We highlight several broad insights made possible by the pathway-alignment approach; further analyses of each network region are provided in Fig. 8. Most straightforwardly, we used pathways from a well studied network (*S. cerevisiae*) to shed light on their aligned counterparts from a less well characterized one (*H. pylori*). For instance, although the function of HP1026 is unknown, its interaction with DNA polymerase (dnaX) and its position in the pathway alignment opposite yeast replication factor C (Rfc2/3/4) suggest that HP1026 is involved in DNA replication in close association with polymerase (Fig. 2*e*; see also Fig. 7*d*). In another example (Fig. 2*d*), the hypothetical protein HP0609 is adjacent to HP0610 and HP0289, which localize to the bacterial outer membrane (26), and opposite yeast Nup1, which localizes to the nuclear pore. This suggests that HP0609 is also membrane-specific and that the bacterial pathway shares homology with the yeast nuclear pore complex.

H. pylori proteins can also shed light on yeast protein function. YLL034C encodes a yeast protein of the AAA family (ATPases associated with various cellular activities) with an undetermined role in ribosome biogenesis (27). A pathway alignment in Fig. 2*e* provides evidence that this protein functions in proteolysis (perhaps to promote ribosome assembly) via its direct interaction with a proteasome 26S subunit (Rpt6), its alignment opposite HP1069 (identified as a protease by Clusters of Orthologous Groups analysis) (28), and its position parallel to Rpt1 and Rpt2, two other proteasome subunits.

A second major insight of our analysis was that pathway alignments often linked two or more pathways or cellular processes not previously known to associate. For example, the network region in Fig. 2*e* contains yeast proteins associated with either DNA polymerase (Rfc2/3/4/5) or the 19S proteasome regulatory cap (Rpt1/2/3/4/6) and provides evidence from both bacteria and yeast that these complexes associate *in vivo*. Consistent with this view, recent evidence suggests that the 19S complex, in addition to its established role in protein degradation, is involved in nucleotide excision repair (29) and can be recruited to promoters during transcription elongation (30).

Other network regions representing multiple functions appear in Fig. 2*f*, linking RNA polymerase (rpoB vs. six yeast genes) with proteins involved in translation (infB, Gcd11, Yif2, Sup35, ileS, and Ils1); Fig. 2*e*, linking RNA helicases involved in nucleic acid processing (uvrB and Dbp8) with arginine biosynthetic enzymes (pyrAb and Cpa2); and Fig. 2*b*, linking RNA helicases involved in translation and ribosomal RNA assembly (Dbp2, Has1, Nog1, Rpl2a, Mak5, and Gcn20) and antioxidative mechanisms (Ssq1, Tsa1, Tsa2, and Sse1). The reasons for many of these associations are unclear, but conservation of interactions across the networks of both species suggests that the associations are functionally significant. Because members of the RNA helicase family are numerous, often poorly characterized, and participate in diverse and sometimes multiple functions (31), pathway alignment may be particularly useful for placing individual RNA helicases into specific functional contexts and for suggesting cases in which they facilitate crosstalk among different cellular processes.

Interestingly, a single bacterial RNA helicase (deaD) in Fig. 2*b* occupies the same pathway position, and perhaps functional role, as three different helicases in yeast (Dbp2, Mak5, and Has1). This observation underscores a third broad insight: Single pathways in bacteria frequently correspond to multiple pathways in yeast, consistent with the current model that yeast have

undergone one or more whole-genome duplications relative to bacteria (32). Other examples include dnaX vs. Rfc2/3/4/5 (Fig. 2*e*) and rpoB vs. six ribosomal subunits in yeast (Fig. 2*f*).

As a fourth insight, we found that proteins within high-scoring pathway alignments did not necessarily pair with their best sequence matches in the other pathway. In Fig. 2, 22% of yeast proteins (13/59) are paired with bacterial proteins other than their best BLAST *H. pylori* match; this proportion is somewhat higher (30/40) when bacterial proteins are compared with yeast (Table 2, which is published as supporting information on the PNAS web site). For example, the best sequence match of *H. pylori* bcp is yeast Dot5, not Tsa1, its pair in the Fig. 2*b* pathway alignment ($E = 8 \times 10^{-15}$ vs. 9×10^{-5}, respectively); similarly, the best match of yeast Tsa1 is *H. pylori* tsaA ($E = 5 \times 10^{-40}$). However, Dot5 and tsaA do not interact with members of the pathway in Fig. 2*b* (e.g., Has1, Sse1, dnaK, or deaD), so that bcp and Tsa1 seem to be the true functional orthologs despite their weaker sequence similarity. We must consider that if a protein with multiple functions in one organism undergoes several rounds of duplication and specialization in the other, it may have different orthologs in different pathway contexts.

Finally, protein kinase pathways appeared as a pervasive feature of the yeast/yeast alignments (see Fig. 3 *c*, *g*, *i*, *j*, *l*, and *o* among others). Because a major challenge in biology is to discover how information is transmitted through compartmentalized modules consisting of protein kinases, phosphatases, and their substrates (33), pathway alignments may provide natural contexts for recognizing new kinase roles and modes of crosstalk. For instance, although the aligned signal transduction pathways Bck1-Mkk2-Slt2 and Rim15-Tpk1/2/3-Pkh1 (Fig. 3*c*) have well established roles in regulation of osmolarity (34) and in nutrient sensing (35), their alignment includes interactions with the kinases Ygr086c and Ypl004c, implicating them as additional signaling factors.

One difficulty with parsing the yeast/yeast network alignment into distinct kinase cascades was that the >120 yeast kinases all shared moderate (≈30%) protein sequence identity such that kinase pathways could shift or invert relative to each other with only a marginal impact on the pathway-alignment score. One way to address this problem might be to augment the score with additional and/or more stringent matching criteria such as protein structural similarity.

Conclusions

We have only begun to explore how global protein network comparisons, whether between species, within species, or using pathway queries, can impact the study of evolution beyond what is gained from analysis of either sequences or interactions alone. By enabling systematic discovery of conserved network structures, pathway alignment is a powerful tool for predicting protein functions, revealing signaling crosstalk, and distinguishing true orthologs from among multiple candidates between species. Moreover, observation of similar interaction paths in two networks increases the likelihood that these paths are biologically significant, such that pathway alignments provide implicit verification for the protein interactions they contain. As the flood of genomic and postgenomic data propels biological research at a variety of informational levels, methods such as these may form the foundations for understanding evolution not merely in terms of conserved sequences but as an array of basic functional modules.

We acknowledge G. Fink, R. Shamir, and O. Ozier for helpful discussions and comments on the manuscript. B.P.K. and D.E.R. were funded by National Cancer Institute Grant 1R01CA97061-01. R.S. was supported by a Fulbright grant. B.R.S. is funded in part by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund. T.I. is the David Baltimore Whitehead Fellow and was funded in part by Pfizer.

1. Mann, M., Hendrickson, R. C. & Pandey, A. (2001) *Annu. Rev. Biochem.* **70,** 437–473.
2. Ito, T., Chiba, T. & Yoshida, M. (2001) *Trends Biotechnol.* **19,** S23–S27.
3. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., *et al.* (2000) *Nature* **403,** 623–627.
4. Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., *et al.* (2002) *Nature* **415,** 180–183.
5. Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., *et al.* (2002) *Nature* **415,** 141–147.
6. Rain, J. C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schachter, V., *et al.* (2001) *Nature* **409,** 211–215.
7. Boulton, S. J., Gartner, A., Reboul, J., Vaglio, P., Dyson, N., Hill, D. E. & Vidal, M. (2002) *Science* **295,** 127–131.
8. von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S. & Bork, P. (2002) *Nature* **417,** 399–403.
9. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 4285–4288.
10. Matthews, L. R., Vaglio, P., Reboul, J., Ge, H., Davis, B. P., Garrels, J., Vincent, S. & Vidal, M. (2001) *Genome Res.* **11,** 2120–2126.
11. Forst, C. V. & Schulten, K. (2001) *J. Mol. Evol.* **52,** 471–489.
12. Dandekar, T., Schuster, S., Snel, B., Huynen, M. & Bork, P. (1999) *Biochem. J.* **343,** 115–124.
13. Tohsato, Y., Matsuda, H. & Hashimoto, A. (2000) *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **8,** 376–383.
14. Ogata, H., Fujibuchi, W., Goto, S. & Kanehisa, M. (2000) *Nucleic Acids Res.* **28,** 4021–4028.
15. Loots, G. G., Ovcharenko, I., Pachter, L., Dubchak, I. & Rubin, E. M. (2002) *Genome Res.* **12,** 832–839.
16. Nakaya, A., Goto, S. & Kanehisa, M. (2001) *Genome Inform. Ser. Workshop Genome Inform.* **12,** 44–53.
17. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
18. Xenarios, I., Salwinski, L., Duan, X. J., Higney, P., Kim, S. M. & Eisenberg, D. (2002) *Nucleic Acids Res.* **30,** 303–305.
19. Wu, C. H., Huang, H., Arminski, L., Castro-Alvear, J., Chen, Y., Hu, Z. Z., Ledley, R. S., Lewis, K. C., Mewes, H. W., Orcutt, B. C., *et al.* (2002) *Nucleic Acids Res.* **30,** 35–37.
20. Gould, S. J. & Collins, C. S. (2002) *Nat. Rev. Mol. Cell Biol.* **3,** 382–389.
21. Pochart, P., Woltering, D. & Hollingsworth, N. M. (1997) *J. Biol. Chem.* **272,** 30345–30349.
22. Patton, E. E., Willems, A. R. & Tyers, M. (1998) *Trends Genet.* **14,** 236–243.
23. Johnson, E. S., Schwienhorst, I., Dohmen, R. J. & Blobel, G. (1997) *EMBO J.* **16,** 5509–5519.
24. Liakopoulos, D., Doenges, G., Matuschewski, K. & Jentsch, S. (1998) *EMBO J.* **17,** 2208–2214.
25. Ozkaynak, E., Finley, D., Solomon, M. J. & Varshavsky, A. (1987) *EMBO J.* **6,** 1429–1439.
26. Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., *et al.* (1997) *Nature* **388,** 539–547.
27. Gadal, O., Strauss, D., Braspenning, J., Hoepfner, D., Petfalski, E., Philippsen, P., Tollervey, D. & Hurt, E. (2001) *EMBO J.* **20,** 3695–3704.
28. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. (2000) *Nucleic Acids Res.* **28,** 33–36.
29. Gillette, T. G., Huang, W., Russell, S. J., Reed, S. H., Johnston, S. A. & Friedberg, E. C. (2001) *Genes Dev.* **15,** 1528–1539.
30. Ferdous, A., Gonzalez, F., Sun, L., Kodadek, T. & Johnston, S. A. (2001) *Mol. Cell* **7,** 981–991.
31. de la Cruz, J., Kressler, D. & Linder, P. (1999) *Trends Biochem. Sci.* **24,** 192–198.
32. Wolfe, K. H. & Shields, D. C. (1997) *Nature* **387,** 708–713.
33. Smith, F. D. & Scott, J. D. (2002) *Curr. Biol.* **12,** R32–R40.
34. Banuett, F. (1998) *Microbiol. Mol. Biol. Rev.* **62,** 249–274.
35. Thevelein, J. M. & de Winde, J. H. (1999) *Mol. Microbiol.* **33,** 904–918.