Identification of Potential Interaction Networks Using Sequence-Based Searches for Conserved Protein-Protein Interactions or "Interologs"

Lisa R. Matthews,[1,2,3] Philippe Vaglio,[1,3] Jérôme Reboul,[1,3] Hui Ge,[1] Brian P. Davis,[2] James Garrels,[2] Sylvie Vincent,[1] and Marc Vidal[1,4]

Protein interaction maps have provided insight into the relationships among the predicted proteins of model organisms for which a genome sequence is available. These maps have been useful in generating potential interaction networks, which have confirmed the existence of known complexes and pathways and have suggested the existence of new complexes and or crosstalk between previously unlinked pathways. However, the generation of such maps is costly and labor intensive. Here, we investigate the extent to which a protein interaction map generated in one species can be used to predict interactions in another species.

Protein interaction maps have provided insight into the relationships among the predicted proteins of model organisms for which a genome sequence is available (Froment-Racine et al. 1997, 2000; Flajolet et al. 2000; Ito et al. 2000, 2001; McCraith et al. 2000; Walhout et al. 2000; Rain et al. 2001). Those maps have been useful particularly in generating potential interaction networks. Such networks have confirmed the existence of known complexes and pathways and have suggested the existence of new complexes and or crosstalk between previously unlinked pathways (Froment-Racine et al. 1997, 2000; Uetz et al. 2000). While the knowledge gained from proteome-wide interaction maps is highly valuable, the generation of such maps is costly and labor intensive. Here, we investigate the extent to which a protein interaction map generated in one species can be used to predict interactions in another species. Systematic BLAST searches for pairs of potential orthologs of known interacting protein partners in *Saccharomyces cerevisiae* have been performed to identify potentially conserved interactions, or "interologs", in *Caenorhabditis elegans*. Starting from a large number of published yeast two-hybrid interactions between yeast proteins, searches for candidate interologs identified networks of potential physical interactions among *C. elegans* proteins. At least 16% of protein interactions in these networks could be detected in a yeast two-hybrid system, suggesting that these interactions are indeed conserved. In addition, many true interologs were amenable to reverse two-hybrid selections (Vidal 1997). Thus, it should be possible to generate reagents such as interaction-defective alleles and/or interaction-dissociating compounds for the further experimental characterization of potential interologs. The above observations suggest that sequence searches for potential interologs and other "comparative proteomics" strategies performed using the protein interaction maps of model organisms will be useful for drug screening programs in parasites, pathogens, and humans.

Present address: [1]Dana-Farber Cancer Institute and Department of Genetics, Harvard Medical School, Boston 02115, Massachusetts, USA; [2]Proteome Inc., Beverly 01915, Massachusetts, USA.
[3]These authors contributed equally to this work.
[4]Corresponding author.
E-MAIL marc_vidal@dfci.harvard.edu; FAX (617) 632-2425.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.205301.

The conceptual translation of complete genome sequences into predicted proteomes and the use of this information to provide a first approximation of functional interactions underlying molecular complexes and regulatory pathways are among the most important challenges of the postgenomic era. One general approach is to use in silico methods to compare the proteomes from two or more organisms and predict interactions based upon presumed or demonstrated protein properties. For example, the Rosetta Stone method predicts functional interactions based upon the observation that pairs of functionally related proteins in one organism are sometimes expressed as single-fusion polypeptides in other organisms (Marcotte et al. 1999). Similarly, phylogenetic profiling is based upon the presumption that groups of functionally interacting proteins are selected together during the course of evolution and may have a tendency to coexist in the same proteomes (Pellegrini et al. 1999).

The approach described here presumes that large numbers of physically interacting proteins in one organism have "coevolved" so that their respective orthologs in other organisms interact as well. This notion of conserved interactions, or "interologs" (Walhout et al. 2000a), is substantiated by the observation that many interactions in signal transduction pathways or molecular machines are conserved between different species. For example, ~7% of interactions currently available in a *C. elegans* protein interaction database have interologs that already have been described in the literature (Walhout et al. 2000a). In addition, searches for potential interologs already have been used to identify the function of several genes. For example, a functional ortholog of the human retinoblastoma protein was found in *Drosophila* using an interolog-based approach (Du et al. 1996). Here, we investigate the extent to which large-scale systematic searches for interologs may be used to identify potential networks of interaction and discuss methods that could be used in studying such potential networks and/or generate therapeutic agents.

RESULTS AND DISCUSSION

To determine the extent to which a protein interaction map generated in *S. cerevisiae* can be used to predict interactions in *C. elegans*, we utilized two large-scale, two-hybrid interaction maps of yeast proteins, which together contain 1195 interac-

11:2120–2126 ©2001 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/01 $5.00; www.genome.org

tions (Ito et al. 2000; Uetz et al. 2000). We identified potential *C. elegans* interologs by comparing the yeast proteins involved in these interactions to all predicted worm proteins as described in Methods. From 1195 yeast interactions, 257 potential worm interologs involving 282 proteins were identified (see below). Two hundred sixteen worm pairs (corresponding to a total of 276 worm open reading frames [ORFs]) were tested further. A sample of 71 of these potential interologs (corresponding to 72 yeast interactions) was used both to recapitulate the original yeast interactions and to experimentally test the corresponding worm protein interactions using the two-hybrid system (Table 1). The yeast and worm ORFs were cloned into both DNA binding (DB) and activation domain (AD) two-hybrid vectors and scored for interaction in both possible orientations (DB-X/AD-Y and DB-Y/AD-X) (see Methods). Each step of this analysis, beginning with the amplification of ORFs from the original clones, was performed in duplicate on two separate days. Protein pairs conferring at least one of the four yeast two-hybrid phenotypes (Vidal 1997) in both experiments were scored as positive interactions. Of the 72 potential yeast interactions tested, 26% (19) exhibited a detectable interaction in our version of the two-hybrid system (Table 1). It is likely that this number reflects differences in the conditions of yeast two-hybrid systems such as yeast strains, reporter genes, and the procedures for scoring interactions. Of these 19 interactions, six (31%) also scored positive with *C. elegans* proteins (Table 1). In addition, one more worm interaction was detected although the corresponding yeast interaction was undetected. Finally, we tested the remaining 145 worm potential interologs for their ability to mediate a two-hybrid interaction. After combining the two experiments, 35/216 potential interolog pairs (16%) exhibited a detectable interaction (Fig. 1A–D, Fig. 2, and Table 2).

This suggests that, using this approach, the minimal proportion of true interologs that can be detected between two species that are evolutionarily distant by about 900 million years is between 16% and 31%. By comparison, an average of five interactors per bait typically is obtained using a worm AD-cDNA library representing ~19,000 genes (2.6×10^{-4}) (Walhout et al. 2000a; Davy et al. 2001). Thus, the frequency of detection of interactions through searches for potential interologs is between 600- and 1100-fold higher than that obtained through conventional two-hybrid screens using random libraries. It should be noted that this approach allows direct testing of interactions that might otherwise be difficult to test in a random two-hybrid screen because of the biased representation of cDNA libraries toward highly expressed genes. In addition, the direct testing of an interaction between two proteins can be performed in 96 well setting and therefore is easily amenable to automation.

The E-value reported by BLAST when searching a database is a measure of the likelihood that the observed similarity could have occurred by chance. Thus, biologists can use this value to infer the homology likelihood. The BLAST E-value between the potential orthologous protein pairs tested in this analysis ranged between 10^{-10} and 10^{-151}. One could expect that a potential interolog obtained with two low E-values is more likely to interact. Interestingly, no detectable correlation was found between the likelihood of homology (*E*-value) and the likelihood of an interaction being conserved between these two organisms (Fig. 1E). However, it is established that the three-dimensional structure of two proteins can be conserved despite considerable primary sequence divergence (Friedberg et al. 2000). In addition, it is possible that two

Table 1. Summary of Yeast Two-Hybrid Results

yeast pairs		positive	*C. elegans* interolog		positive
SMT3	UBC9	x	K12C11.2	F29B9.6	x
LSM4	LSM1		F32A5.7	F40F8.9	
LSM4	CKB2		F32A5.7	T01G9.6A	
LSM2	LSM1		T10G3.6	F40F8.9	
LSM2	DHH1		T10G3.6	C07H6.5	
TEM1	SMX3	x	C39F7.4	ZK652.1	
YPT1	PEP12		C39F7.4	F56A8.7	
TEM1	HMO1		C39F7.4	T20B12.8	
YPT31	PEP12		F53G12.1	F56A8.7	
GCN3	GCN3	x	ZK1098.4	ZK1098.4	x
CKA2	CKB2	x	B0205.7	T01G9.6A	x
CDC7	DHH1		C34G6.5	C07H6.5	
RPT3	RPT4		F23F12.6	F23F1.8	
RPT3	RPT3		F23F12.6	F23F12.6	
RPT3	RPT5	x	F23F12.6	F56H1.4	
VTI1	PEP12		Y57G11C.4	F56A8.7	
YLR432W	GCN3		T22D1.3	ZK1098.4	
YRB1	NAP1		F59A2.1	D2096.8	
CPA1	CKB2		D2085.1	T01G9.6A	
LSM4	LSM7		F32A5.7	ZK593.7	x
SKP1	MET30	x	F46A9.5	F55B12.3	x
SKP1	RUB1		F46A9.5	F45H11.2	
SKP1	BDF1		F46A9.5	F13C5.2	
SKP1	SGT1		F46A9.5	D1054.3	
LSM2	LSM5		T10G3.6	F28F8.3	
LSM2	LSM7		T10G3.6	ZK593.7	
LSM2	SMD2		T10G3.6	C52E4.3	
TEM1	ECI1	x	C39F7.4	R06F6.9	
TEM1	DMC1		C39F7.4	Y43C5A.6	
TEM1	LAC1		C39F7.4	C09G4.1	
TEM1	MNS1		C39F7.4	ZC410.3	
TEM1	SOR1	x	C39F7.4	R04B5.5	
TEM1	KIN3		C39F7.4	F19H6.1	
TIF35	GNA1		F22B5.2	B0024.12	
SNF4	GAL83	x	F55F3.1	Y111B2C.H	
CDC7	YOR006C		C34G6.5	F52C12.2	
RPT3	YGR232W	x	F23F12.6	F40G9.1	
LSM5	LSM7	x	F28F8.3	ZK593.7	x
GDH1	LSM1	x	ZK829.4	F40F8.9	
RAD51	UBC9		Y43C5A.6	F29B9.6	
RAD51	YMR233W		Y43C5A.6	T24G10.2	
CLB2	CKS1		T06E6.2	C09G4.3	
CLB3	CKS1	x	**T06E6.2**	**C09G4.3**	
CLB2	MUS81		T06E6.2	C43E11.2	
CLB2	FPR1		T06E6.2	Y18D10A.19C	
CLB2	YHR035W		T06E6.2	Y113G7A.3	
SRP1	CAR1		F32E10.4	T21F4.1	
SRP1	ICL1		F32E10.4	C05E4.9	
SRP1	PHO13		F32E10.4	F44E7.2	
SRP1	SOR1	x	F32E10.4	R04B5.5	
SRP1	YMR226C		F32E10.4	F28H7.2	
SRP1	MET17		F32E10.4	ZK1127.10	
SRP1	TSA1		F32E10.4	F09E5.2	
FZF1	CKB2		F53F8.1	T01G9.6A	
CLB1	CKS1		H31G24.4	C09G4.3	
YMR269W	GCN3		F49H12.5	ZK1098.4	
YDL246C	SOR1	x	R04B5.5	R04B5.5	x
HTA1	NAP1	x	C50F4.13	D2096.8	
PPA2	GCN3		C47E12.4	ZK1098.4	
GPD2	GNA1		K11H3.1	B0024.12	
PRP46	PEP12	x	D1054.15	F56A8.7	
PRE10	GNA1		ZK945.2	B0024.12	
SNF4	COQ5		Y111B2C.H	ZK652.9	
SNF4	YDL214C		Y111B2C.H	R02C2.1	
SNF4	YMR291W	x	Y111B2C.H	K12C11.4	
RPS8B	CKS1		F42C5.8	C09G4.3	
RPS8B	GNA1		F42C5.8	B0024.12	
PPG1	HSC82		Y75B8A.30	C47E8.5	
PPG1	PPT1		Y75B8A.30	Y39B6B.FF	
PPG1	CPR6		Y75B8A.30	Y75B12B.5	
NAB2	GCN3		T21B6.3	ZK1098.4	
UBC9	UBC9	x	Y59A8C.A	F29B9.6	

Summary of yeast two-hybrid results for a random set of 72 yeast interactions from two large-scale two-hybrid interaction maps of yeast proteins (Ito et al. 2000; Uetz et al. 2000) and of their corresponding *C. elegans* potential interologs. Interactions that were detected in both organisms are highlighted blue. Interactions detected only in yeast are highlighted green. One interaction was detected in *C. elegans* although undetected in yeast, and is highlighted orange.

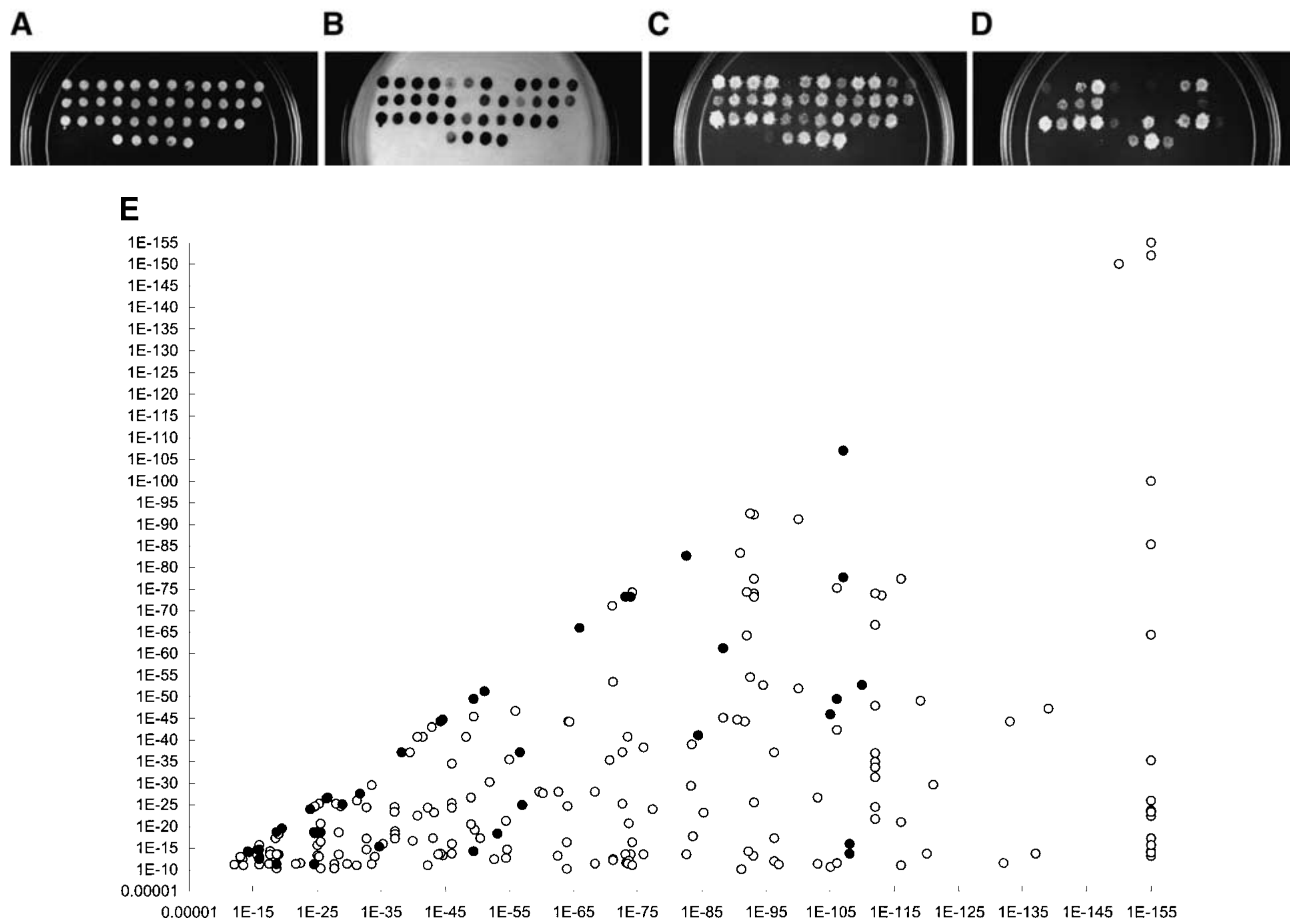

Figure 1 Experimentally verified interactions between *Saccharomyces cerevisiae* and *Caenorhabditis elegans*. (*A–D* Yeast diploid cells expressing each of 35 *C. elegans* potential interologs. Pairs are arranged in the order described in Table 2. The five patches at the bottom are controls (negative control on the left side and controls of increasing interaction strength towards the right side). See Vidal (1997) for a detailed description of these controls. (*B*) β-Galactosidase assay to detect the expression of *GAL1::lacZ*. (*C*) Growth assay on SC-Leu-Trp-His, +20 mM 3AT plates to detect the expression of *GAL1::HIS3*. (*D*) Growth assay on SC-Leu-Trp-Ura plates to detect the expression of *SPAL10::URA3*. (*E*) Conservation of interactions. Each *C. elegans* protein pair tested was plotted according to two *E*-values. The first *E*-value corresponds to the conservation between the X (from yeast) and X′ (from *C. elegans*) proteins while the second *E*-value corresponds to the conservation between the Y (from yeast) and Y′ (from *C. elegans*) proteins. The smaller of the two *E*-values was plotted on the *X*-axis and the greater on the *Y*-axis. The *C. elegans* protein pairs that tested positive in the two-hybrid system are labeled in black.

interacting proteins may have coevolved such that only discrete interacting domains were conserved.

The data described above suggest that the approach of sequence-based searches for candidate interologs can be used globally to identify potential networks of interactions. However, such networks only can be considered as biological hypotheses. Hence, we investigated methods to generate reagents that can be used to study potential interaction networks identified by interolog searches. The reverse two-hybrid system provides a genetic selection that allows the rapid identification of *cis*-acting mutations or *trans*-acting molecules that dissociate potential interactions (Vidal 1997). The two-hybrid *SPAL10::URA3* inducible reporter gene (Vidal et al. 1996) confers sensitivity to 5-Fluoroorotic acid (5-FOA). The dissociation of the yeast two-hybrid interaction confers a selective advantage allowing screens for dissociating compounds or for mutations that prevent the normal association of a protein pair using positive selection. Such reagents can be used back in vivo to characterize the role of the corresponding protein-protein interactions (Endoh et al. 2001).

To test the degree to which the reverse two-hybrid system can be applied to our network of identified interologs, we determined the percentage of the interactions described above that could be counter-selected on media containing 5-FOA (Vidal 1997). Starting from the 35 true worm interologs described above, 77% (27/35) of *C. elegans* interactions were detected as 5-FOA sensitive (Fig. 3). Because the reverse two-hybrid system can be automated (Endoh et al. 2001), it is possible that relatively large numbers of yeast two-hybrid interactions that emerge from interolog searches could indeed be tested back in the relevant biological settings.

This work suggests that interaction maps from one species may be useful in predicting interactions in another species and may provide insight into the function of otherwise uncharacterized proteins. In addition, the identification of an interolog provides additional support for the validity of the initial interaction found in the "reference" species. This may be most meaningful if the only evidence for the original interaction comes, itself, from a high-throughput experiment. When the function of one of the proteins in the starting spe-

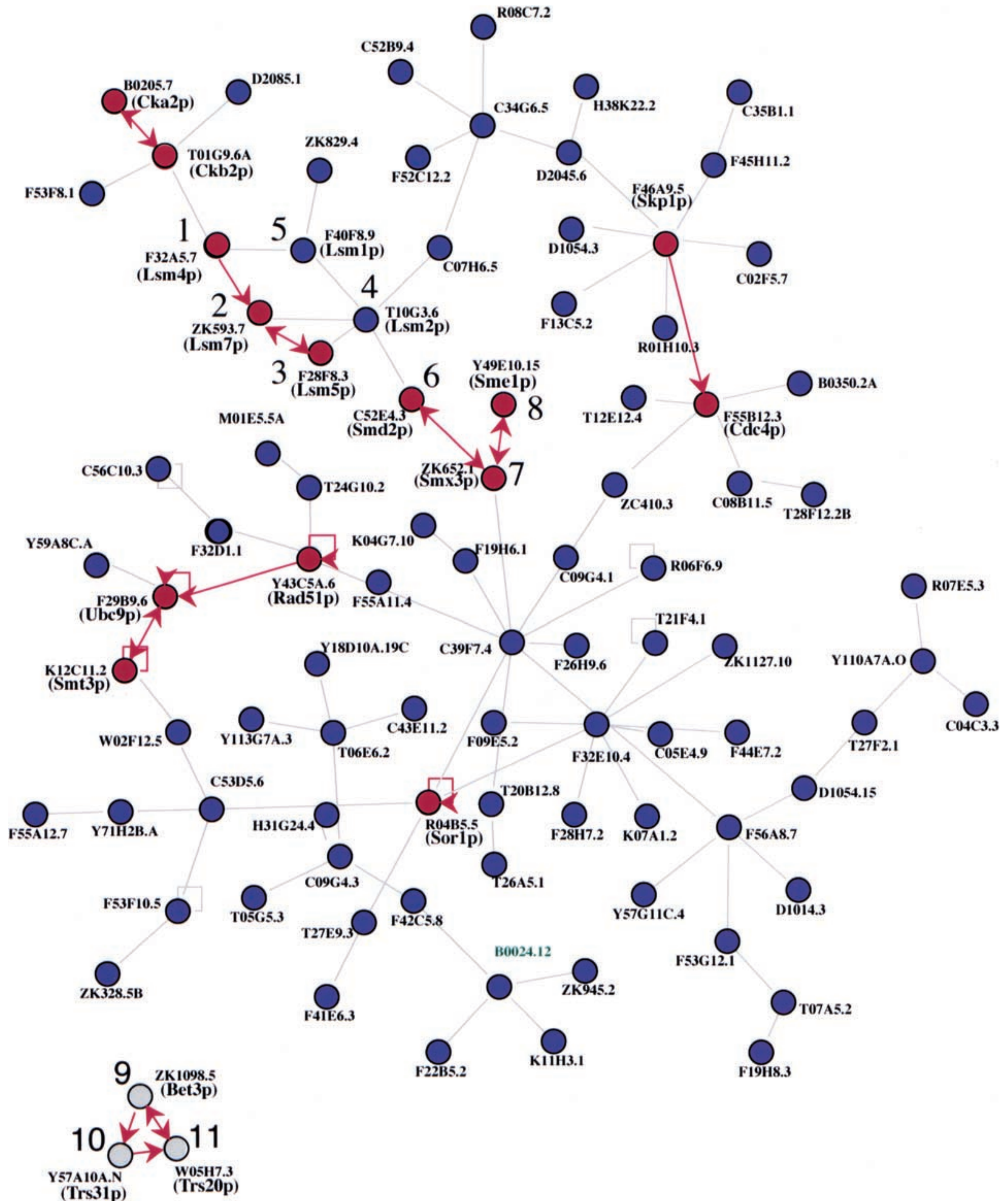

Figure 2 A predicted network of *Caenorhabditis elegans* protein interactions involving a subset of the 216 pairs of protein interactions tested in our analysis. Corresponding yeast proteins names are shown in parentheses. Nodes represent proteins and edges represent an interaction between two yeast proteins. Interactions also detected in *C. elegans* are represented by red arrows. Interactions detected in both orientations are represented by double-headed red arrows. Interactions involving proteins 1–8 have been implicated in RNA processing in yeast and the interaction cluster (Nos. 9–11) involves three proteins whose potential yeast orthologs have been found in TRAPP complexes.

Table 2. Summary of Yeast Two-Hybrid Results

		Protein 1					Protein 2				
		S. cerevisiae	*C.elegans*	E value	Extent of alignment in shortest protein	% identity	*S. cerevisiae*	*C.elegans*	E value	Extent of alignment in shortest protein	% identity
RNA metabolism	sNRNP complex	SMX3	ZK652.1	2.00E-19	87%	48.9%	SMD2	C52E4.3	8.00E-27	99%	49.2%
	sNRNP complex	SME1	Y49E10.15	2.00E-19	94%	48.5%	SMX3	ZK652.1	2.00E-19	85%	48.2%
	sNRNP complex	LSM4	F32A5.7	4.00E-13	84%	21.6%	LSM7	ZK593.7	1.00E-16	85%	36.0%
	sNRNP complex	LSM5	F28F8.3	2.00E-15	96%	37.4%	LSM7	ZK593.7	1.00E-16	87%	36.0%
	Polyadenylation factor I complex	YTH1	F11A10.3	2.00E-36	96%	7.2%	FIP1	F32D1.9	5.00E-16	93%	13.8%
Vesicular transport	TRAPP complex	TRS20	W05H7.3	2.00E-20	57%	32.0%	BET3	ZK1098.5	3.00E-26	96%	32.8%
	TRAPP complex	TRS31	Y57A10A.n	5.00E-12	57%	16.5%	TRS20	W05H7.3	2.00E-20	93%	32.0%
	TRAPP complex	TRS31	Y57A10A.n	5.00E-12	79%	16.5%	BET3	ZK1098.5	3.00E-26	94%	32.8%
	vesicular transport	VPS27	C07G1.5	1.00E-30	96%	19.3%	YHL002W	C34G6.7	8.00E-27	96%	24.1%
	Recombinosome	RAD51	Y43C5A.6	2.00E-84	95%	46.0%	RAD51	Y43C5A.6	2.00E-84	96%	46.0%
Protein metabolism	Recombinosome/protein degradation	RAD51	Y43C5A.6	1.00E-106	96%	46.0%	UBC9	F29B9.6	3.00E-51	96%	50.0%
	protein degradation	UBC9	F29B9.6	3.00E-51	85%	50.0%	UBC9	F29B9.6	3.00E-51	96%	50.0%
	protein degradation	SMT3	K12C11.2	6.00E-15	85%	36.3%	UBC9	F29B9.6	3.00E-51	85%	50.0%
	protein degradation	SMT3	K12C11.2	6.00E-15	74%	36.3%	SMT3	K12C11.2	6.00E-15	95%	36.3%
	protein degradation	LAP4	F01F1.9	2.00E-46	96%	23.6%	LAP4	F01F1.9	2.00E-46	96%	23.6%
	protein degradation	NPL4	F59E12.5	3.00E-86	87%	28.0%	UFD1	F19B6.2	9.00E-43	95%	33.6%
	Proteosome19s and 26s proteins	RPN11	K07D4.3	1.00E-107	100%	61.9%	RPN8	R12E2.3	2.00E-79	96%	38.7%
	Protein modification/conjugation	ULA1	C26E6.8	5.00E-20	99%	25.3%	UBA3	F11H8.1	6.00E-55	97%	27.2%
	Farnesyltransferase complex	RAM1	F23B12.6	9.00E-59	95%	33.5%	RAM2	R02D3.5	1.00E-26	82%	26.2%
	GIM complex (protein folding)	GIM5	R151.9	9.00E-15	83%	28.0%	YKE2	F21C3.5	4.00E-15	94%	30.2%
	eIF2B complex	GCN3	ZK1098.4	5.00E-46	63%	35.4%	GCN3	ZK1098.4	5.00E-46	77%	35.4%
	SCF-Cdc4p-Sic1p complex	SKP1	F46A9.5	7.00E-39	93%	40.1%	CDC4	F55B12.3	2.00E-58	98%	21.1%
Protein kinase	Protein kinase CKII complex	CKA2	B0205.7	1.00E-110	95%	51.3%	CKB2	T01G9.6a	2.00E-54	88%	39.0%
	SNF1 protein kinase complex	SNF1	T01C8.1	1.00E-108	78%	35.3%	GAL83	F55F3.1	2.00E-14	45%	20.7%
	Transcription factor complex	HAP3	W10D9.4	3.00E-27	99%	14.6%	HAP5	F23F1.1	2.00E-20	99%	8.9%
General metabolism	carbohydrate metabolism	SOR1	R04B5.5	7.00E-75	83%	41.2%	SOR1	R04B5.5	1.00E-75	83%	41.2%
	carbohydrate metabolism	YJR024C	ZC373.5	3.00E-28	88%	29.2%	YJR024C	ZC373.5	3.00E-28	88%	29.2%
	nucleotide metabolism	CDD1	C47D2.2	3.00E-21	84%	36.0%	CDD1	C47D2.2	3.00E-21	99%	36.0%
	nucleotide metabolism	PRS5	R151.2	1.00E-47	97%	32.1%	PRS2	R151.2	1.00E-105	97%	50.1%
	nucleotide metabolism	YLR328W	F26H9.4	1.00E-25	99%	17.9%	YLR328W	F26H9.4	1.00E-25	99%	17.9%
	nucleotide metabolism	YNK1	F25H2.5	6.00E-53	93%	61.7%	YNK1	F25H2.5	6.00E-53	92%	61.7%
	amino acid metabolism	PRO3	M153.1	2.00E-28	100%	27.4%	PRO3	M153.1	2.00E-28	100%	27.4%
	metabolism (?)	FOL2	F32G8.6	1.00E-67	95%	50.8%	FOL2	F32G8.6	1.00E-67	95%	50.8%
	unknown function	YJR072C	C34E10.2	5.00E-90	91%	40.0%	YLR243W	Y75B8A.14	6.00E-63	97%	41.7%
	unknown function	YDR267C	Y18D10A.9	2.00E-33	94%	29.3%	YHR122W	F45G2.10	3.00E-29	84%	29.3%

The yeast protein pairs involved in the 35 conserved interactions could be divided into several general functional groups of core biological processes as described in the Yeast Protein Database YPD (http://www.proteome.com): (1) RNA metabolism, (2) vesicular transport, (3) protein metabolism (synthesis, modification, degradation), and (4) general metabolism. The *C. elegans* and *S. cerevisiae* protein names, their corresponding BLAST *E*-values, and the extent of the alignment in the shortest protein for each pair is shown.

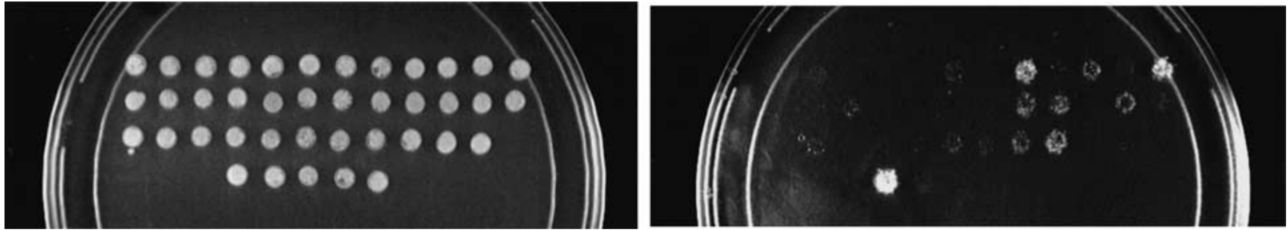

Figure 3 Assessment of the reverse two-hybrid system as a method to further study *Caenorhabditis elegans* protein interactions. Reverse two-hybrid system. Each yeast diploid expressing one of the 35 conserved interacting protein pairs from *C. elegans* was tested to identify those that are sensitive to 5-Fluoroorotic acid (5-FOA). Pairs are arranged in the order described in Table 2. Diploids were replica plated onto both SC-Leu-Trp (A) and SC -Leu, -Trp + 0.2% 5-FOA (B). Twenty-seven of thirty-five pairs (77%) were 5-FOA sensitive.

cies has not been characterized, a "guilt by association" annotation also can be applied with more confidence. One potential therapeutic application of this approach would be to identify interactions conserved between a well-characterized model organism and a related pathogen. The majority of the interactions conserved between the distantly related species *S. cerevisiae* and *C. elegans* seem to be involved primarily in core metabolic processes. However, it is reasonable to predict that certain interactions conserved between more closely related organisms such as *Drosophila* and other insects such as mosquitoes, or *C. elegans* and other nematodes such as *Ascaris lumbricoides*, would be specific to the very closely related species. These interactions could be subjected to the reverse two-hybrid analysis (Vidal 1997) to identify reagents that are capable of dissociating them (Young et al. 1998). Such reagents might be valuable candidates for therapeutic agents.

METHODS

Potential Ortholog Identification

To identify potential orthologs of yeast proteins in *C. elegans*, we used BLASTP to search a *C. elegans* database generated from the ORF predictions in version WS7 of ACeDB. Only matches with *E*-values lower than 10^{-10} were considered. Although other studies (Snel et al. 1999) have considered as potential orthologs two proteins that are each other's best match in their respective genomes, we decided to systematically select as potential orthologs the best *C. elegans* matches for the yeast proteins. Indeed, orthology can be a one-to-many or many-to-many relationship (Tatusov et al. 1997). Thus, in the few cases where the same *C. elegans* protein is the best match for two distinct yeast proteins, we consider this *C. elegans* protein to be the potential orthologs for both yeast proteins. We recognize that this procedure allowed us to evaluate only one potential *C. elegans* ortholog per yeast protein.

PCR and Cloning

Complete *C. elegans* ORFs that correspond to 282 potential *C. elegans* orthologs were PCR amplified from a mixed-stage cDNA library (Walhout et al. 2000b). The control *S. cerevisiae* ORFs were amplified from S288C genomic DNA. PCR products were generated using ORF-specific primers designed by the program OSP (Hillier and Green 1991) and tailed with the AttB1 (5′) and AttB2 (3′) Gateway recombinational cloning sequences (Walhout et al. 2000a,b; Reboul et al. 2001). *C. elegans* and yeast ORFs were cloned into the Gateway Entry vector (INVITROGEN) as described with several modifications (Reboul et al. 2001).

Each ORF was PCR amplified from its corresponding Entry clone for subsequent cloning into both Gal4p DNA binding domain-fusion (DB) and Gal4p activation domain-fusion (AD) yeast two-hybrid vectors pGBT9 (Clontech) and pACT2 (Endoh 2000), respectively. ORFs to be cloned into the DB-fusion vector were amplified from Entry clones using the primers DB-B1: 5′TAGTAACAAAGGTCAAAGACAGTTGACT-GTATCGTCGAGGTTGTACAAAAAAGCAGGCT-3′ and PGBT9.B2-TERM: 5′-AAATCAT AAATCATAAGAAATTCGC-CCGGAATTAGCTTGGTTGTACAAGAAAGCTGGGT-3′. ORFs to be cloned into the AD vector were amplified from Entry clones using the primers AD-B1: 5′-CTATTCGATGATGAAGA TACCCCACCAAACCCAAAAAAAGAGTTGTACAAAAA AGCAGGCT-3′ and pACT2.B2-Term: 5′-TGAAGTGAACTTG CGGGGTTTTTCAGTATCTACGATTCATTTGTACAAG AAAGCTGGGT-3′.

Yeast Transformation/Gap-Repair and Mating

Yeast transformation/gap-repair reactions were performed in 96-well plates as described elsewhere (Walhout and Vidal 2001) with several modifications. MaV103 (*MAT***a**) cells were transformed with the 2 µm AD-fusion vector pACT2-GFP and the corresponding AD-PCR fragments. MaV203 (*MAT*α) cells were transformed with the 2 µm DB-fusion vector pGBT9 and the corresponding DB-PCR fragments. Five microliters of each transformation mix were spotted in 96-spot format onto synthetic complete (SC) plates lacking leucine or tryptophane (SC-Leu or SC-Trp). Yeast carrying the AD-fusion constructs were mated to yeast cells carrying the DB-fusion constructs in "96-spot" format by replica plating both onto YEPD plates. To identify self-activator DB-fusion proteins, each transformant bearing a DB-fusion protein construct was mated to a transformant carrying an empty AD vector. Following 1 d of growth at 30°C, mating plates were replica plated to SC-Leu-Trp plates to select for diploids.

Yeast Two-Hybrid Analysis

Diploids were transferred to a 15-cm nylon filter on a YEPD plate for subsequent β-galactosidase assays and to an SC-Leu-Trp-Ura plate and an SC-Leu-Trp-His plate supplemented with 20 mM 3-aminotriazole (3AT) to assay for the *SPAL10::URA3* and *GAL1::HIS3* reporter gene activity, respectively. The identity of both hybrid proteins from diploids exhibiting a positive yeast two-hybrid phenotype was determined by sequencing the corresponding ORFs. The identity of all ORFs tested was verified by sequencing.

5-FOA Sensitivity

The two-hybrid inducible *SPAL10::URA3* reporter gene drives expression of the *URA3* gene, which is involved in uracil biosynthesis, and which also is able to catalyze the conversion of 5-FOA into a toxic compound. Thus, yeast diploids expressing interacting partners fused to AD and DB are able to grow on medium lacking uracil, but do not grow on medium containing 5-FOA. This allows for the screening of either dissociating compounds or interaction-defective alleles that will confer a

positive growth selection. The 5-FOA negative selection is, however, dependent on the strength of the two-hybrid interaction. We identified interactions that are 5-FOA sensitive and thus suitable for reverse two-hybrid screening, by replica plating diploids onto both SC-Leu-Trp and SC-Leu-Trp + 0.2% 5-FOA.

ACKNOWLEDGMENTS

We thank Marian Walhout and David Hill for helpful suggestions and critical reading of this manuscript. S.V. was supported by a fellowship from the Charles A. King Memorial Trust. This work was supported by grants 5R01HG01715–02 (NHGRI), P01CA80111–02 (NCI), 7 R33 CA81658–02 (NCI) and 232 (MGRI) awarded to M.V.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Davy, A., Bello, P., Thierry-Mieg, N., Vaglio, P., Hitti, J., Doucette-Stamm, L., Thierry-Mieg, D., Reboul, J., Boulton, S., Walhout, A.J.M., et al. 2001. A protein-protein interaction map of the *C. elegans* 26S proteasome. *EMBO Reports* **2:** 821–828.

Du, W., Vidal, M., Xie, J.-E., and Dyson, N. 1996. *RBF*, a novel RB-related gene that regulates E2F activity and interacts with *cyclin E* in *Drosophila*. *Genes & Dev.* **10:** 1206–1218.

Endoh, H., Walhout, A.J.M., and Vidal, M. 2000. GFP-reverse two-hybrid system: Application to the characterization of large numbers of potential protein-protein interactions. *Methods Enzymol.* **328:** 74–88.

Endoh, H., Vincent, S., Jacob, Y., Réal, E., Walhout, A.J.M., and Vidal, M. An integrated version of the reverse two-hybrid system for the "post-proteomic" era. *Methods Enzymol.* in press.

Flajolet, M., Rotondo, G., Daviet, L., Bergametti, F., Inchausp, G., Tiollais, P., Transy, C. and Legrain, P. 2000. A genomic approach of the hepatitis C virus generates a protein interaction map. *Gene.* **242:** 369–379

Friedberg, I., Kaplan, T., and Margalit, H. 2000. Evaluation of PSI-BLAST alignment accuracy in comparison to structural alignments. *Protein Sci.* **9:** 2278–2284.

Fromont-Racine, M., Rain, J.C., and Legrain, P. 1997. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genetics* **16:** 277–282.

Guthrie, C. 1994. The spliceosome is a dynamic ribonucleoprotein machine. *Harvey Lect.* **90:** 59–80.

Hillier, L. and Green, P. 1991. OSP: A computer program for choosing PCR and DNA sequencing primers. *PCR Methods Appl.* **1:** 124–128.

Ito, T., Tashiro, K., Muta, S., Ozawa, R., Chiba, T., Nishizawa, M., Yamamoto, K., Kuhara, S., and Sakaki, Y. 2000. Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between yeast proteins. *Proc. Natl. Acad. Sci.* **97:** 1143–1147.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., Sakaki, Y. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci.* **98:** 4569–4574.

Marcotte, E.M., Pellegrini, M., Ng, H.L., Rice, D.W., Yeates, T.O., and Eisenberg, D. 1999. Detecting protein function and protein-protein interactions from genome sequences. *Science* **285:** 751–753.

McCraith, S., Holtzman, T., Moss, B., and Fields S. 2000. Genome-wide analysis of vaccinia virus protein-protein interactions. *Proc. Natl. Acad. Sci.* **97:** 4879–4884.

Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D., and Yeates, T.O. 1999. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc. Natl. Acad. Sci.* **96:** 4285–4288.

Rain, J.C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schachter, V. 2001. The protein-protein interaction map of *Helicobacter pylori*. *Nature* **409:** 211–215.

Reboul, J., Vaglio, P., Tzellas, N., Thierry-Mieg, N., Moore, T., Jackson, C., Shin-i, T., Kohara, Y., Thierry-Mieg, D., Thierry-Mieg, J., et al. 2001. Open-reading-frame sequence tags (OSTs) support the existence of at least 17,300 genes in *C. elegans*. *Nat. Genet.* **27:** 332–336.

Snel, B., Bork, P., and Huynen, M.A. 1999. Genome phylogeny based on gene content. *Nat. Genet.* **21:** 108–110.

Tatusov, R.L., Koonin E.V., and Lipman, D.J.A. 1997. A genomic perspective on protein families. *Science* **278:** 631–637.

The *C. elegans* Genome Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* **282:** 2012–2018.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P.A., et al. 2000. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403:** 623–627.

Vidal, M. 1997. The reverse two-hybrid system. In *The Yeast Two-Hybrid System* (eds. Bartel, P. and Fields, S.), pp. 109–147. Oxford University Press, New York.

Vidal, M. and Endoh, H. 1999. Prospects for drug screening using the reverse two-hybrid system. *Trends in Biotechnology* **17:** 374–381.

Walhout, A.J.M., Sordella, R., Lu, X., Hartley, J.L., Temple, G.F., Brasch, M.A., Thierry-Mieg, N., and Vidal, M. 2000a. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287:** 116–122.

Walhout, A.J.M., Temple, G.F., Brasch, M.A., Hartley, J.L., Lorson, M.A., van den Heuvel, S., and Vidal, M. 2000b. Gateway recombinational cloning: Application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol.* **328:** 575–592.

Walhout, A.J.M. and Vidal, M. 2001. High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. *Methods* **24:** 297–306.

Young, K., Lin, S., Sun, L., Lee, E., Modi, M., Hellings, S., Husbands, M., Ozenberger, B., and Franco, R. 1998. Identification of a calcium channel modulator using a high-throughput yeast two-hybrid system. *Nature Biotechnology* **16:** 946–950.

Received July 12, 2001; accepted in revised form September 20, 2001.