PPISearch: a web server for searching homologous protein–protein interactions across multiple species

Chun-Chen Chen[1], Chun-Yu Lin[1], Yu-Shu Lo[1] and Jinn-Moon Yang[1,2,3,]*

[1]Institute of Bioinformatics and Systems Biology, [2]Department of Biological Science and Technology and [3]Core Facility for Structural Bioinformatics, National Chiao Tung University, Hsinchu, 30050, Taiwan

ABSTRACT

As an increasing number of reliable protein–protein interactions (PPIs) become available and high-throughput experimental methods provide systematic identification of PPIs, there is a growing need for fast and accurate methods for discovering homologous PPIs of a newly determined PPI. PPISearch is a web server that rapidly identifies homologous PPIs (called PPI family) and infers transferability of interacting domains and functions of a query protein pair. This server first identifies two homologous families of the query, respectively, by using BLASTP to scan an annotated PPIs database (290 137 PPIs in 576 species), which is a collection of five public databases. We determined homologous PPIs from protein pairs of homologous families when these protein pairs were in the annotated database and have significant joint sequence similarity ($E \leq 10^{-40}$) with the query. Using these homologous PPIs across multiple species, this sever infers the conserved domain–domain pairs (Pfam and InterPro domains) and function pairs (Gene Ontology annotations). Our results demonstrate that the transferability of conserved domain-domain pairs between homologous PPIs and query pairs is 88% using 103 762 PPI queries, and the transferability of conserved function pairs is 69% based on 106 997 PPI queries. The PPISearch server should be useful for searching homologous PPIs and PPI families across multiple species. The PPISearch server is available through the website at http://gemdock.life.nctu.edu.tw/ ppisearch/.

INTRODUCTION

Interactions between proteins are critical to most biological processes. To identify and characterize protein–protein interactions (PPIs) and their networks, many high-throughput experimental approaches, such as yeast two-hybrid screening, mass spectroscopy and tandem affinity purification and computational methods [phylogenetic profiles (1), known 3D complexes (2) and interologs (3)] have been proposed (4). Some PPI databases, such as IntAct (5), BioGRID (6), DIP (7), MIPS (8) and MINT (9), have accumulated PPIs submitted by biologists, and those from mining literature, high-throughput experiments and other data sources. As these interaction databases continue growing in size, they become increasingly useful for analysis of newly identified interactions.

The discovery of sequence homologs to a known protein often provides clues for understanding the function of a newly sequenced gene. As an increasing number of reliable PPIs become available, identifying homologous PPIs should be useful to understand a newly determined PPI. Recently, several PPI databases (e.g. IntAct and BioGRID) allow users to input one or a pair of proteins or gene names to acquire the PPIs associated with the query protein(s). Few computational methods (10,11) applied homologous interactions to assess the reliability of PPIs.

To address this issue, we proposed the PPISearch server for searching homologous PPIs across multiple species and annotating the query protein pair. According to our knowledge, PPISearch is the first public server that identifies homologous PPIs from annotated PPI databases and infers transferability of interacting domains and functions between homologous PPIs and the query. PPISearch is an easy-to-use web server that allows users to input a pair of protein sequences. Then, this server finds homologous PPIs in multiple species from five public databases

*To whom correspondence should be addressed. Tel: 886 3 571212 56942; Fax: 886 3 5729288; Email: moon@faculty.nctu.edu.tw
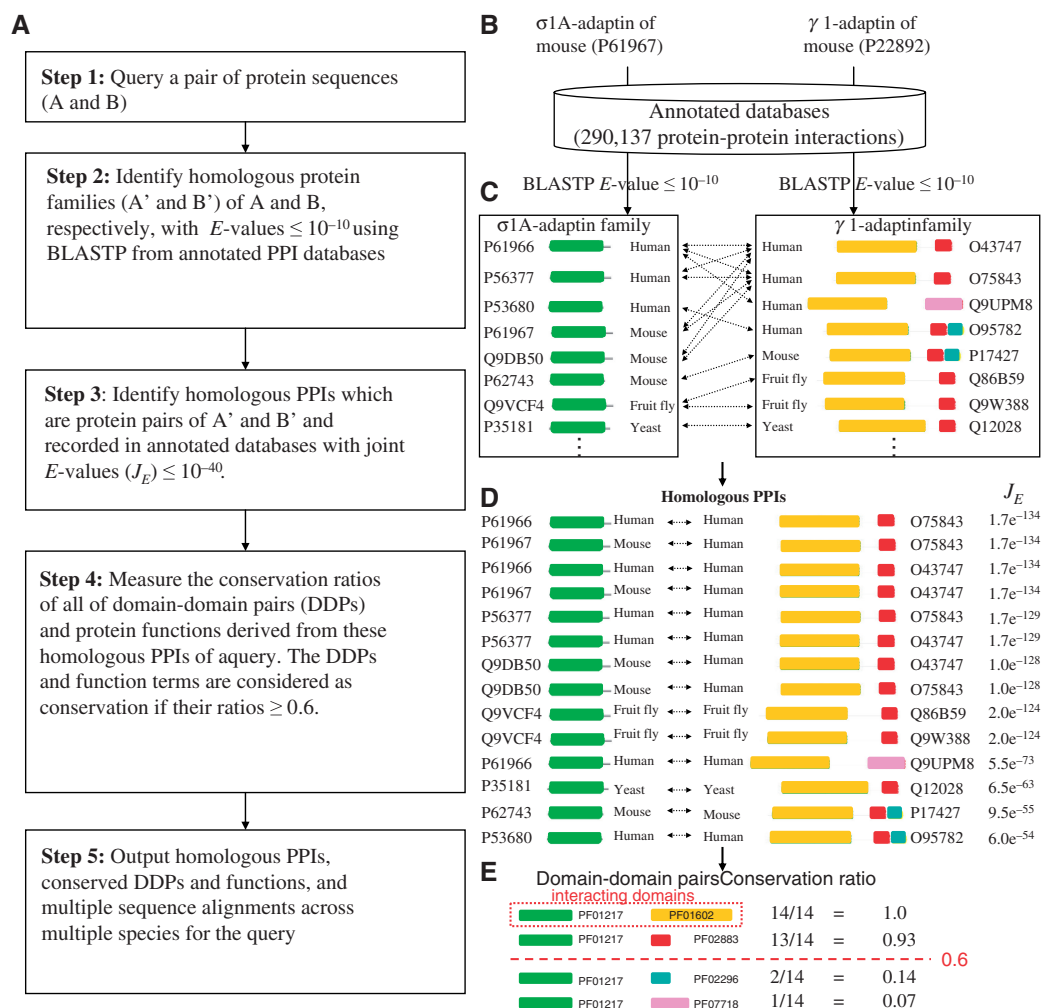
Figure 1. Overview of the PPISearch server for homologous protein–protein interaction search and conservation analysis using proteins σ1A-adaptin and γ1-adaptin as the query. (**A**) The main procedure. (**B**) Identify homologs of σ1A-adaptin and γ1-adaptin using BLASTP to scan the annotated PPI databases. (**C**) The homologous families of σ1A-adaptin and γ1-adaptin with E-values $\leq 10^{-10}$. (**D**) Homologous PPIs of the query. (**E**) Conservation ratios of domain-domain pairs derived from homologous PPIs.

(IntAct, MIPS, DIP, MINT and BioGRID) and annotates the query. Our results demonstrate that this server achieves high agreements on interacting domain–domain pairs and function pairs between query protein pairs and their respective homologous PPIs.

METHOD AND IMPLEMENTATION

Figure 1 shows the details of the PPISearch server to search homologous PPIs of a query protein pair (A and B) by the following steps (Figure 1A). This server first identifies the homologous families (A′ and B′) of A and B, respectively, with $E \leq 10^{-10}$ by using BLASTP to scan the annotated PPI databases (Figure 1B and C). All protein pairs of A′ and B′ are considered candidates of homologous PPIs. We selected homologous PPIs from these candidates, which are recorded in the annotated databases, and have significant joint sequence similarity ($E \leq 10^{-40}$) between candidates and the query (Figure 1D). Then, we measure the conservation ratios

of domain-domain pairs [DDPs; Pfam (12) and InterPro (13) domains] and protein functions [Gene Ontology annotations (14)] derived from these homologous PPIs of the query (Figure 1E). This server provides conserved DDPs and protein functions for annotating the query. Finally, this server provides homologous PPIs in multiple species; conservations and GO annotations of protein functions; conservations and annotations of DDPs; and the best-matched protein pair of the query.

Homologous protein–protein interaction

The concept of homologous PPI is the core of the PPISearch server to identify the PPI family and measure DDPs and functional conservations of a query protein pair (A and B). We define a homologous PPI as follows: (1) homologs of A and B are proteins with significant sequence similarity BLASTP E-values $\leq 10^{-10}$ (3,15); (2) significant joint sequence similarity (joint E-value $J_E \leq 10^{-40}$) between two pairs, i.e. (A, A$_1$′) and (B, B$_1$′), of the query protein pair (A and B) and their respective

homologs (A_1' and B_1') recorded in annotated PPI databases. This work followed previous studies (3,15) to define joint sequence similarity as

$$J_E = \sqrt{E_A} \times E_B \qquad \qquad 1$$

where E_A is the *E*-value of proteins A and A_1'; and E_B is the *E*-value of proteins B and B_1'. Here, $J_E \leq 10^{-40}$ is considered a significant similarity according to statistical analysis of 290 137 annotated PPIs and 6597 orthologous PPI families collected from the PORC database (16).

Annotations of homologous PPI

A query protein pair and its homologous PPIs, significant both in sequence and joint sequence similarity, can be considered a PPI family. The concept of PPI families is similar to that of protein sequence family (12,13) and protein structure family (17). We believe that PPI families can be applied widely in biological investigations. Here, we assume that the members of a PPI family are conserved on specific functions and in interacting domain(s). Using these conservations of a PPI family, our server can be used to annotate the protein functions and DDPs of a query protein pair.

Transferability of domain–domain pairs. A query protein pair and its homologous PPIs often show conserve interacting DDPs. To measure the occurence of each DDP in a PPI family, we define the conservation ratio (CRD_p) of a DDP_p in homologous PPIs of a query protein pair *i* as

$$CRD_p = \frac{\text{Number of homologous PPIs with a domain pair } p}{\text{Number of homologous PPIs of query } i}$$
$$2$$

Figure 1D and E show an example to calculate the CRD values of four DDPs. In addition, to evaluate the transferability of DDPs between a query and its homologous PPIs statistically, this study defines the shared ratio (SRD) of DDPs using CRD_p and 103 762 annotated PPIs as query protein pairs. The SRD of DDPs against different ratio *c* is given as

$$SRD = \frac{\sum_{i \in Q} d_i(CRD_p \geq c)}{\sum_{i \in Q} D_i(CRD_p \geq c)} \qquad 3$$

where *Q* is a set of annotated PPIs in databases (here, the total number of PPIs in *Q* is 103 762); *i* is a query protein pair; $d_i(CRD_p \geq c)$ is the number of DDPs with CRD_p values exceeding *c*; and these DDPs are shared by the query *i* and its homologous PPIs. $D_i(CRD_p \geq c)$ is the total number of the DDPs with $CRD_p \geq c$, where DDPs are derived from homologous PPIs of the query *i*. Here, this work used a statistical approach to determine the threshold *c* (here, *c* = 0.6) of CRD_p to yield reliable DDP annotations with an acceptable level of D_i. Please note that CRD_p and SRD are computed from a query protein pair and a set of queries, respectively.

Transferability of molecular function. The members of a PPI family often have similar molecular functions. PPISearch uses the molecular function (MF) terms of Gene Ontology (14) to annotate the functions of a query protein pair. The conservation ratio (CRF_m) of an MF term pair (MFP) *m* in homologous PPIs of a query *i* is utilized to measure the agreement and is defined as

$$CRF_m = \frac{\text{Number of homologous PPIs with a GO MF term pair } m}{\text{Number of homologous PPIs of query } i}$$
$$4$$

Additionally, the shared ratio of MFPs (SRF), which is statistically derived from 106 997 annotated queries, is utilized to estimate the transferability of conserved function pairs shared by the query and its homologous PPIs. The SRF against different ratio *k* is defined as

$$SRF = \frac{\sum_{i \in Q} f_i(CRF_m \geq k)}{\sum_{i \in Q} F_i(CRF_m \geq k)} \qquad 5$$

where *Q* is a set of annotated PPIs in databases; *i* is a query protein pair; $f_i(CRF_m \geq k)$ is the number of MFPs with CRF_m values exceeding *k* and these MFPs are shared by the query *i* and its homologous PPIs; and $F_i(CRF_m \geq k)$ is the total number of MFPs with $CRF_m \geq k$, where MFPs are derived from homologous PPIs of the query *i*. Here, *k* is set to 0.6.
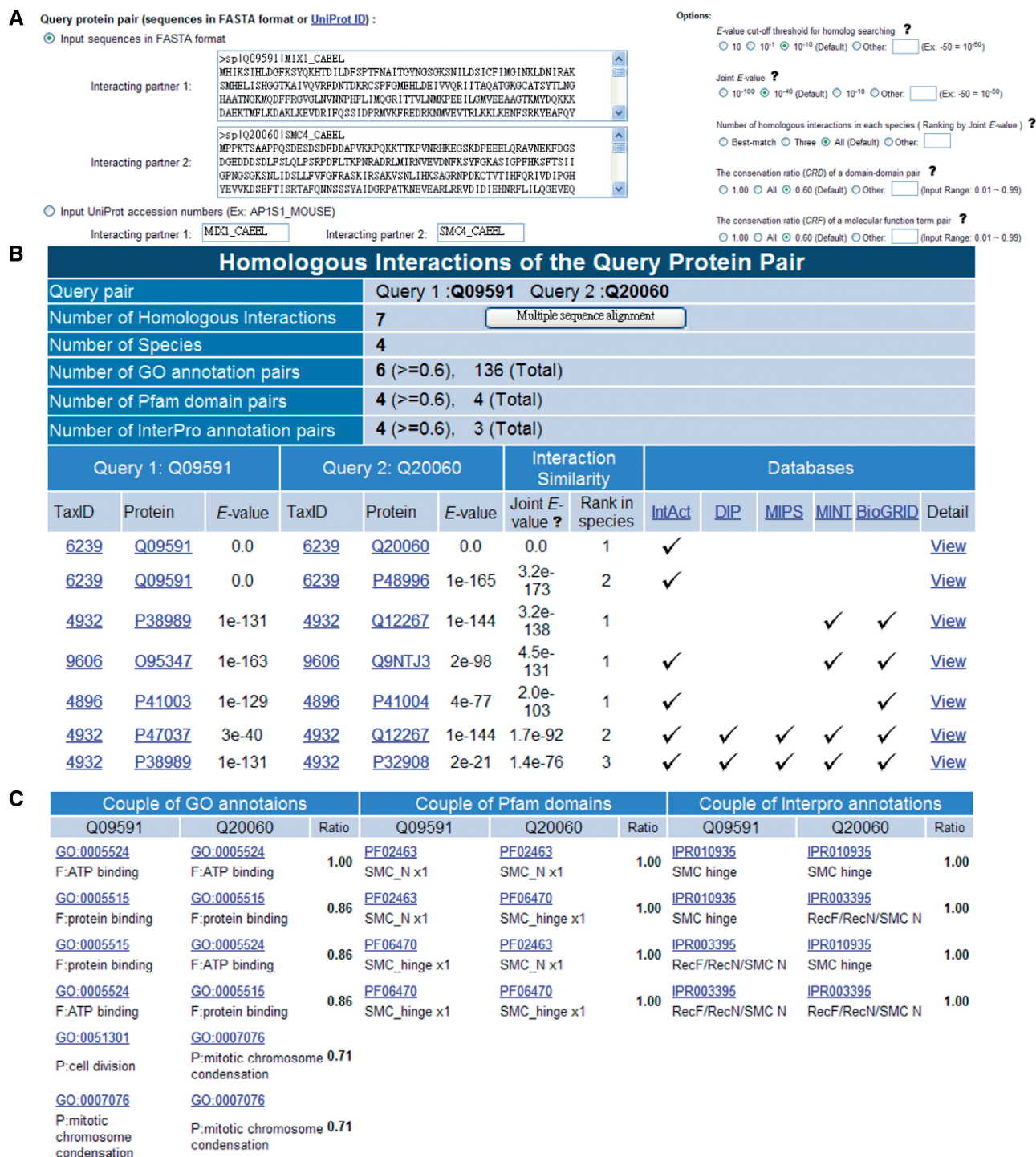
INPUT, OUTPUT AND OPTIONS

The PPISearch is an easy-to-use web server (Figure 2). Users input a pair of protein sequences in FASTA format or UniProt ID, and choose *E*-value thresholds for homologs and for homologous PPIs (Figure 2A). In addition, users can assign the CRD and CRF thresholds, specific species and the number of homologous PPIs in a species.

Typically, the PPISearch server yields homologous PPIs within 20 s when sequence length is ≤350 (Figure 2B). This server identifies homologous PPIs in multiple species; conservations and GO annotations of protein functions; conservations and annotations of DDPs; and the best-matched protein pairs of the query (Figure 2C). Additionally, the PPISearch server provides multiple sequence alignments of homologous PPIs and indicates the conserved residues based on amino acid types. For each homologous PPI, this server shows the alignments and experimental annotations (e.g. interaction types, experimental methods, gene names and GO terms).

Example analysis

σ1A-adaptin and γ1-adaptin. Figure 1C and D show search results using σ1A-adaptin (UniProt accession number: P61967) and γ1-adaptin (P22892) of *Mus musculus* as the query. These two proteins are components of the heterotetrameric adaptor protein complex 1 (AP-1), which medicates clathrin-coated vesicle transport from the *trans*-Golgi network to endosome (18). According to the crystal structure (PDB code 1W63) (19), this protein pair is a physical interaction, but it is not recorded in the annotated PPI database. For this query, the PPISearch server identifies 14 homologous PPIs, a PPI family, from four species (human, mouse, fruit fly and yeast). This PPI family has

A Query protein pair (sequences in FASTA format or UniProt ID) :

◉ Input sequences in FASTA format

Interacting partner 1:
```
>sp|Q09591|MIX1_CAEEL
MHIKSIHLDGFKSYQKHTDILDFSPTFNAITGYNGSGKSNILDSICFIMGINKLDNIRAK
SMHELISHGGTKAIVQVRFDNTDKRCSPFGMEHLDEIVVQRIITAQATGKGCATSYTLNG
HAATNGKMQDFFRGVGLNVNNPHFLIMQGRITTVLNMKPEEILGMVEEAAGTKMYDQKKK
DAEKTMFLKDAKLKEVDRIFQSSIDPRMVKFREDRKNMVEVTRLKKLKENFSRKYEAFQY
```

Interacting partner 2:
```
>sp|Q20060|SMC4_CAEEL
MPPKTSAAPPQSDESDSDFDDAPVKKPQKKTTKPVNRHKEGSKDPEEELQRAVNEKFDGS
DGEDDDSDLFSLQLPSRPDFLTKPNRADRLMIRNVEVDNFKSYFGKASIGPFHKSFTSII
GPNGSGKSNLIDSLLFVFGFRASKIRSAKVSNLIHKSAGRNPDKCTVTIHFQRIVDIPGH
YEVVKDSEFTISRTAFQNNSSSYAIDGRPATKNEVEARLRRVDIDIEHNRFLILQGEVEQ
```

◯ Input UniProt accession numbers (Ex: AP1S1_MOUSE)

Interacting partner 1: MIX1_CAEEL   Interacting partner 2: SMC4_CAEEL

Options:

E-value cut-off threshold for homolog searching ?
◯ 10 ◯ 10⁻¹ ◉ 10⁻¹⁰ (Default) ◯ Other: ___ (Ex: -50 = 10⁻⁵⁰)

Joint E-value ?
◯ 10⁻¹⁰⁰ ◉ 10⁻⁴⁰ (Default) ◯ 10⁻¹⁰ ◯ Other: ___ (Ex: -50 = 10⁻⁵⁰)

Number of homologous interactions in each species (Ranking by Joint E-value) ?
◯ Best-match ◯ Three ◉ All (Default) ◯ Other: ___

The conservation ratio (CRD) of a domain-domain pair ?
◯ 1.00 ◯ All ◉ 0.60 (Default) ◯ Other: ___ (Input Range: 0.01 ~ 0.99)

The conservation ratio (CRF) of a molecular function term pair ?
◯ 1.00 ◯ All ◉ 0.60 (Default) ◯ Other: ___ (Input Range: 0.01 ~ 0.99)

B

Homologous Interactions of the Query Protein Pair

Query pair	Query 1 :Q09591 Query 2 :Q20060
Number of Homologous Interactions	7 [Multiple sequence alignment]
Number of Species	4
Number of GO annotation pairs	6 (>=0.6), 136 (Total)
Number of Pfam domain pairs	4 (>=0.6), 4 (Total)
Number of InterPro annotation pairs	4 (>=0.6), 3 (Total)

Query 1: Q09591			Query 2: Q20060			Interaction Similarity		Databases					
TaxID	Protein	E-value	TaxID	Protein	E-value	Joint E-value ?	Rank in species	IntAct	DIP	MIPS	MINT	BioGRID	Detail
6239	Q09591	0.0	6239	Q20060	0.0	0.0	1	✓					View
6239	Q09591	0.0	6239	P48996	1e-165	3.2e-173	2	✓					View
4932	P38989	1e-131	4932	Q12267	1e-144	3.2e-138	1				✓	✓	View
9606	O95347	1e-163	9606	Q9NTJ3	2e-98	4.5e-131	1	✓			✓	✓	View
4896	P41003	1e-129	4896	P41004	4e-77	2.0e-103	1	✓				✓	View
4932	P47037	3e-40	4932	Q12267	1e-144	1.7e-92	2	✓	✓	✓	✓	✓	View
4932	P38989	1e-131	4932	P32908	2e-21	1.4e-76	3	✓	✓	✓	✓	✓	View

C

Couple of GO annotaions			Couple of Pfam domains			Couple of Interpro annotations		
Q09591	Q20060	Ratio	Q09591	Q20060	Ratio	Q09591	Q20060	Ratio
GO:0005524 F:ATP binding	GO:0005524 F:ATP binding	1.00	PF02463 SMC_N x1	PF02463 SMC_N x1	1.00	IPR010935 SMC hinge	IPR010935 SMC hinge	1.00
GO:0005515 F:protein binding	GO:0005515 F:protein binding	0.86	PF02463 SMC_N x1	PF06470 SMC_hinge x1	1.00	IPR010935 SMC hinge	IPR003395 RecF/RecN/SMC N	1.00
GO:0005515 F:protein binding	GO:0005524 F:ATP binding	0.86	PF06470 SMC_hinge x1	PF02463 SMC_N x1	1.00	IPR003395 RecF/RecN/SMC N	IPR010935 SMC hinge	1.00
GO:0005524 F:ATP binding	GO:0005515 F:protein binding	0.86	PF06470 SMC_hinge x1	PF06470 SMC_hinge x1	1.00	IPR003395 RecF/RecN/SMC N	IPR003395 RecF/RecN/SMC N	1.00
GO:0051301 P:cell division	GO:0007076 P:mitotic chromosome condensation	0.71						
GO:0007076 P:mitotic chromosome condensation	GO:0007076 P:mitotic chromosome condensation	0.71						

Figure 2. The PPISearch server search results using proteins MIX-1 and SMC-4 of *Caenorhabditis elegans* as the query. (**A**) The user interface for assignments of query protein sequences and *E*-value thresholds of homologs and homologous PPIs. (**B**) Homologous PPIs of MIX-1−SMC-4 in multiple species and public databases. (**C**) Conserved protein functions (GO terms) and domain-domain pairs (Pfam and InterPro) of homologous PPIs with a conservation ratio ≥ 0.6.

four DDPs (Figure 1E)—PF01217-PF01602 (CRD is 1.0), PF01217-PF02883 (0.93), PF01217-PF02296 (0.14) and PF01217-PF07718 (0.07). Two DDPs (PF01217-PF01602 and PF01217-PF02883) with highest CRD ratios are the domain compositions of the query and PF01217-PF01602 is the interacting domains (19).

This server allows users to choose the J_E threshold of homologous PPIs. For example, when J_E is set to 10^{-100} (default value is 10^{-40}), the number of homologous PPIs decreases from 14 to 10 by filtering out the last four PPIs (Figure 1D). These 10 homologous PPIs consistently include the two DDPs PF01217-PF01602 and
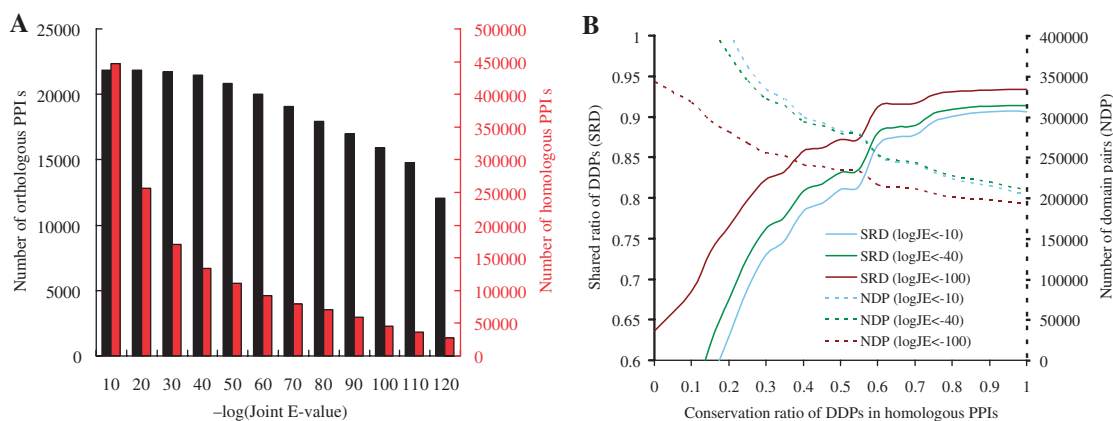
Figure 3. Evaluations of the PPISearch server. (**A**) The relationships between joint *E*-value J_E and the numbers of orthologous PPIs (black) and homologous PPIs (red) derived from 290 137 annotated PPIs. (**B**) The relationships between conservation ratios of DDPs with shared ratios of DDPs and with the number (dotted lines) of DDPs derived from 103 762 PPI families. The shared ratio of DDPs is 0.88 and the number of DDPs is 252 728 when the conservation ratio is 0.6 and joint *E*-value is 10^{-40} (green lines).

PF01217-PF02883, each with a CRD = 1.0. Furthermore, users can choose the best match or number of homologous PPIs in a species. In this manner, the PPISearch server is able to select the primary homologous PPIs of each species for specific applications, such as evolutionary analysis of essential proteins.

MIX-1 and SMC-4. Mitotic chromosome and X-chromosome-associated protein (MIX-1, Q09591) and structural maintenance of chromosomes protein 4 (SMC-4, Q20060) of *Caenorhabditis elegans* are members of SMC protein family, and are required for mitotic chromosome segregation (20). Both MIX-1 and SMC-4 are essential components in forming the condensin complex for interphase chromatin to convert into mitotic-like condense chromosomes (20,21). Using *C. elegans* MIX-1 and SMC-4 as the query protein pair and J_E is set to 10^{-40}, the PPISearch server found seven homologous interactions from annotated PPI databases (Figure 2B). These seven homologous PPIs are consistently SMC–SMC protein interactions, including SMC-2−SMC-4, SMC-3−SMC-4 and SMC-2−SMC-1, in four species. Among these homologous PPIs, two PPIs, Q95347-Q9NTJ3 (*Homo sapiens*) and P38989-Q12267 (*Saccharomyces cerevisiae*), are orthologous interactions of the query MIX-1−SMC-4 (16).

These seven homologous PPIs of MIX-1 and SMC-4 include 136 GO term pairs. Among these GO terms, the CRF ratios of four GO MF term pairs and two GO BP term pairs exceed 0.6 (Figure 2C). These six GO term pairs are consistent with the term-pair combinations of MIX-1 and SMC-4. For example, MIX-1 and SMC-4 have the same two GO MF annotations, protein binding (GO:0005515) and ATP-binding (GO:0005524). Additionally, these seven homologous PPIs contain four DDPs with CRD ratios of 1.0. These four DDPs, PF02463-PF02463, PF06470-PF02463, PF02463-PF06470 and PF06470-PF06470, are recorded in *i*Pfam (12) and are consistent with the query pair. The hinge–hinge interaction (PF02463-PF02463) is experimentally proved, and is conserved in the eukaryotic

SMC-2–SMC-4 heterodimer (22). These analytical results reveal that the PPISearch server is able to identify homologous PPIs that share conserved DDPs and MFPs with the query.

RESULTS

To evaluate the usefulness of the PPISearch server for the discovery of homologous PPIs and for the annotations of a query protein pair, we selected two query protein sets, termed HOM and ORT. To search homologous PPIs, HOM and ORT are used to assess PPISearch performance and to determine the threshold of joint *E*-value J_E [Equation (1)] (Figure 3A). In addition, the HOM set was applied to infer the relations between conservation ratios [CRD and CRF defined in Equations (2) and (4)] and the transferability of DDPs and MFPs, respectively, between a query and its homologous PPIs (Figure 3B and Supplementary Figure S1). The HOM set includes all 290 137 PPIs and the ORT set has 6597 orthologous PPI families (14 571 PPIs) derived from the annotated PPI database and PORC orthology database (16).

HOM and ORT were used to assess the PPISearch server in identifying homologous PPIs and orthologous PPIs, respectively, by searching the annotated PPI database (290 137 PPIs with 54 422 proteins). Figure 3A shows the relationships between joint *E*-value J_E and number of orthologous PPIs (black) and homologous PPIs (red). The orthologous PPIs often have the same functions and domains. When $J_E \leq 10^{-40}$, the number of orthologous PPIs decreases significantly; conversely, the number of homologous PPIs decreases more gradually than that at $J_E \geq 10^{-40}$. This result shows that the proposed method is able to identify 98.2% orthologous PPIs with a reasonable number of homologous PPIs when $J_E \leq 10^{-40}$.

To evaluate the transferability of DDPs and MFPs between a query and its homologous PPIs, we used the SRD [Equation (3)] and SRF [Equation (5)]. The HOM set is used to evaluate the utility of the PPISearch server in annotating the query protein pair. By excluding proteins

without domain annotations from the query set, 103 762 PPIs are used to evaluate the transferability (SRD) of conserved DDPs between these query PPIs and their respective homologous PPIs (Figure 3B). The transferability (SRF) of conserved functions between the 106 997 PPIs and their homologous PPIs is assessed by excluding proteins without molecular function terms of GO from the original query set (Supplementary Figure S1).

Figure 3B shows the relationship between conservation ratios (CRD) of DDPs and the SRD ratios. The SRD ratio increases significantly (solid lines) when the CRD increases and CRD $\leq$ 0.6. Conversely, the number of DDPs derived from 103 762 PPI families decreases (dotted lines) as CRD increases. If the CRD is set to 0.6 and the joint E-value is set to 10^{-40} (green lines), the SRD is 0.88 and the number of DDPs is 252 728. This result demonstrates that members of a PPI family derived by PPISearch reliably share DDPs (or interacting domains). Additionally, similar results were obtained for transferability of conserved functions between homologous PPIs and the query (Supplementary Figure S1). The members of a PPI family have similar molecular functions, and SRF ratios are highly correlated with conservation ratios (CRF) of MFPs. When the CRF is 0.6 and the joint E-value is 10^{-40} (green lines), the SRF is 0.69 and the number of MFPs is 454 251.

These results reveal that the PPISearch server achieves a high SRD with a reasonable number of DDPs when the joint E-value is set to 10^{-40}. In summary, these experimental results demonstrate that this server achieves high agreement on DDPs and MFPs between the query and their respective homologous PPIs.

CONCLUSIONS

This study demonstrates the utility and feasibility of the PPISearch server in identifying homologous PPIs and inferring conserved DDPs and MFPs from PPI families. By allowing users to input a pair of protein sequences, PPISearch is the first server that can identify homologous PPIs from annotated PPI databases and infer transferability of interacting domains and functions between homologous PPIs and a query. Our experimental results demonstrate that the query protein pair and its homologous PPIs achieve high agreement on conserved DDPs and MFPs. We believe that PPISearch is a fast homologous PPIs search server and is able to provide valuable annotations for a newly determined PPI.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. 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. USA*, **96**, 4285–4288.
2. Chen,Y.-C., Lo,Y.-S., Hsu,W.-C. and Yang,J.-M. (2007) 3D-partner: a web server to infer interacting partners and binding models. *Nucleic Acids Res.*, W561–W567.
3. Yu,H.Y., Luscombe,N.M., Lu,H.X., Zhu,X.W., Xia,Y., Han,J.D.J., Bertin,N., Chung,S., Vidal,M. and Gerstein,M. (2004) Annotation transfer between genomes: Protein-protein interologs and protein-DNA regulogs. *Gen. Res.*, **14**, 1107–1118.
4. Shoemaker,B.A. and Panchenko,A.R. (2007) Deciphering protein-protein interactions. Part I. Experimental techniques and databases. *PLoS Comput. Biol.*, **3**, 337–344.
5. Kerrien,S., Alam-Faruque,Y., Aranda,B., Bancarz,I., Bridge,A., Derow,C., Dimmer,E., Feuermann,M., Friedrichsen,A., Huntley,R. et al. (2007) IntAct – open source resource for molecular interaction data. *Nucleic Acids Res.*, **35**, D561–D565.
6. Stark,C., Breitkreutz,B.J., Reguly,T., Boucher,L., Breitkreutz,A. and Tyers,M. (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.*, **34**, D535–D539.
7. Salwinski,L., Miller,C.S., Smith,A.J., Pettit,F.K., Bowie,J.U. and Eisenberg,D. (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.*, **32**, D449–D451.
8. Mewes,H.W., Dietmann,S., Frishman,D., Gregory,R., Mannhaupt,G., Mayer,K.F.X., Munsterkotter,M., Ruepp,A., Spannagl,M., Stuempflen,V. et al. (2008) MIPS: analysis and annotation of genome information in 2007. *Nucleic Acids Res.*, **36**, D196–D201.
9. Chatr-Aryamontri,A., Ceol,A., Palazzi,L.M., Nardelli,G., Schneider,M.V., Castagnoli,L. and Cesareni,G. (2007) MINT: the molecular INTeraction database. *Nucleic Acids Res.*, **35**, D572–D574.
10. Patil,A. and Nakamura,H. (2005) Filtering high-throughput protein-protein interaction data using a combination of genomic features. *BMC Bioinformatics*, **6**, 100–112.
11. Saeed,R. and Deane,C. (2008) An assessment of the uses of homologous interactions. *Bioinformatics*, **24**, 689–695.
12. Finn,R.D., Tate,J., Mistry,J., Coggill,P.C., Sammut,S.J., Hotz,H.R., Ceric,G., Forslund,K., Eddy,S.R., Sonnhammer,E.L.L. et al. (2008) The Pfam protein families database. *Nucleic Acids Res.*, **36**, D281–D288.
13. Hunter,S., Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Binns,D., Bork,P., Das,U., Daugherty,L., Duquenne,L. et al. (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res.*, **37**, D211–D215.
14. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. et al. (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–29.
15. Matthews,L.R., Vaglio,P., Reboul,J., Ge,H., Davis,B.P., Garrels,J., Vincent,S. and Vidal,M. (2001) Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs". *Gen. Res.*, **11**, 2120–2126.
16. Kersey,P., Bower,L., Morris,L., Horne,A., Petryszak,R., Kanz,C., Kanapin,A., Das,U., Michoud,K., Phan,I. et al. (2005) Integr8 and genome reviews: integrated views of complete genomes and proteomes. *Nucleic Acids Res.*, **33**, D297–D302.
17. Andreeva,A., Howorth,D., Brenner,S.E., Hubbard,T.J., Chothia,C. and Murzin,A.G. (2004) SCOP database in 2004: refinements

integrate structure and sequence family data. *Nucleic Acids Res.*, **32**, D226–D229.

18. Bonifacino,J.S. and Traub,L.M. (2003) Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Ann. Rev. Biochem.*, **72**, 395–447.

19. Heldwein,E.E., Macia,E., Jing,W., Yin,H.L., Kirchhausen,T. and Harrison,S.C. (2004) Crystal structure of the clathrin adaptor protein 1 core. *Proc. Natl Acad. Sci. USA*, **101**, 14108–14113.

20. Lieb,J.D., Albrecht,M.R., Chuang,P.T. and Meyer,B.J. (1998) MIX-1: an essential component of the C. elegans mitotic machinery executes x chromosome dosage compensation. *Cell*, **92**, 265–277.

21. Hagstrom,K.A., Holmes,V.F., Cozzarelli,N.R. and Meyer,B.J. (2002) C. elegans condensin promotes mitotic chromosome architecture, centromere organization, and sister chromatid segregation during mitosis and meiosis. *Genes Dev.*, **16**, 729–742.

22. Hirano,M. and Hirano,T. (2002) Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA. *EMBO J.*, **21**, 5733–5744.