*Published online 21 May 2009* *Nucleic Acids Research, 2009, Vol. 37, Web Server issue* ***W417–W421***
*doi:10.1093/nar/gkp329*

# d-Omix: a mixer of generic protein domain analysis tools

**Duangdao Wichadakul*, Somrak Numnark and Supawadee Ingsriswang**

National Center for Genetic Engineering and Biotechnology (BIOTEC) - Information Systems Laboratory (ISL), Pathumthani, Thailand

Received February 20, 2009; Revised April 17, 2009; Accepted April 20, 2009

## ABSTRACT

**Domain combination provides important clues to the roles of protein domains in protein function, interaction and evolution. We have developed a web server d-Omix (a Mixer of Protein Domain Analysis Tools) aiming as a unified platform to analyze, compare and visualize protein data sets in various aspects of protein domain combinations. With InterProScan files for protein sets of interest provided by users, the server incorporates four services for domain analyses. First, it constructs protein phylogenetic tree based on a distance matrix calculated from protein domain architectures (DAs), allowing the comparison with a sequence-based tree. Second, it calculates and visualizes the versatility, abundance and co-presence of protein domains via a domain graph. Third, it compares the similarity of proteins based on DA alignment. Fourth, it builds a putative protein network derived from domain–domain interactions from DOMINE. Users may select a variety of input data files and flexibly choose domain search tools (e.g. hmmpfam, superfamily) for a specific analysis. Results from the d-Omix could be interactively explored and exported into various formats such as SVG, JPG, BMP and CSV. Users with only protein sequences could prepare an InterProScan file using a service provided by the server as well. The d-Omix web server is freely available at http://www.biotec.or.th/isl/Domix.**

## INTRODUCTION

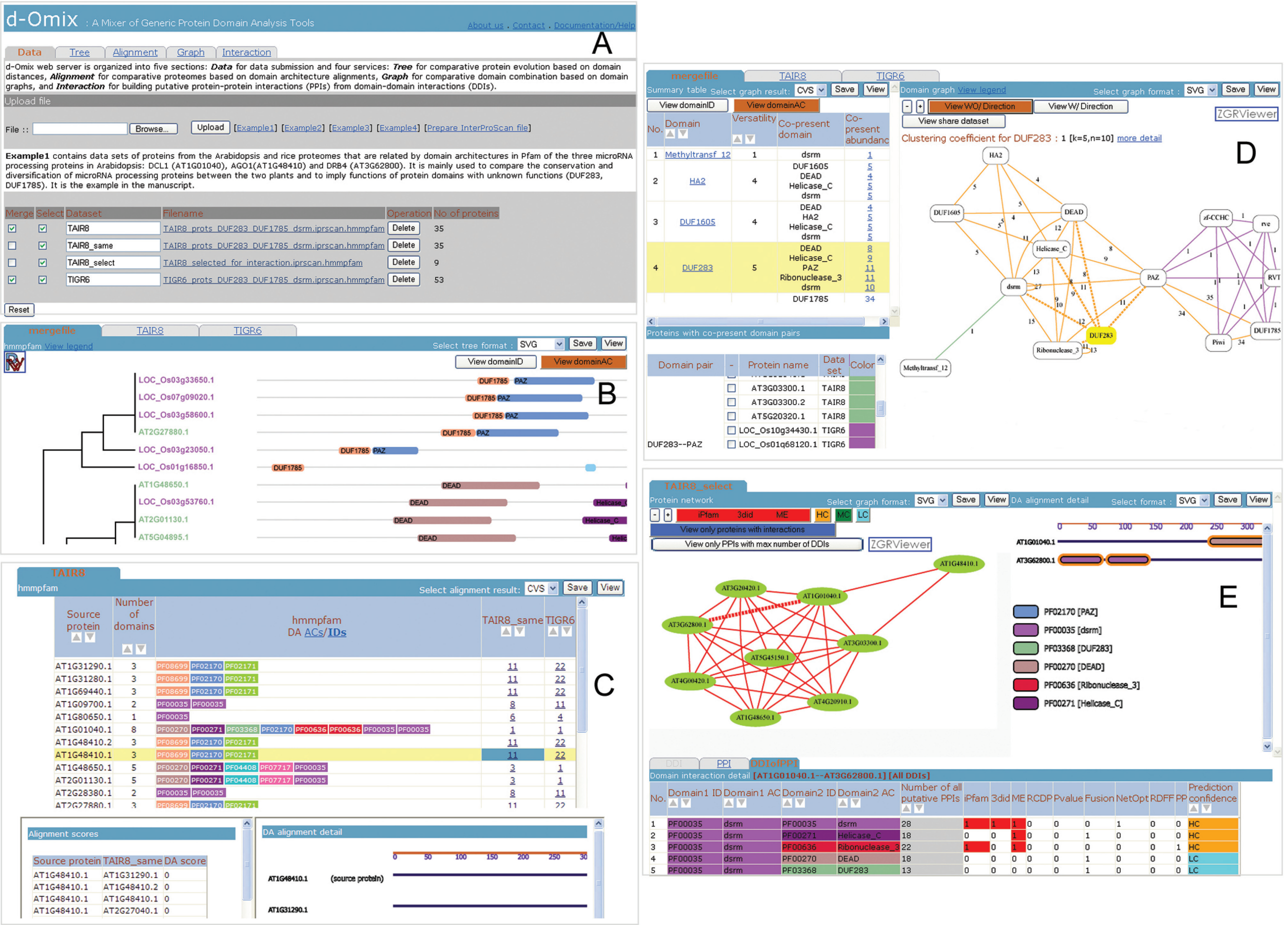
Protein domains are units of evolution (1,2). Different combinations of protein domains generate several types of modifications affecting protein functions. Addition or deletion of domains can modify substrate binding, increase or decrease catalytic activity, change the categorized reaction, cause loss of catalytic function, or regulate enzyme function (3). The comparison of protein domain combinations and architectures (DAs) will shed light on their related functions, possible annotations of unknown proteins and evolution. Domain combination has been analyzed for examining and predicting protein functions (3–6), protein cellular localization (7,8) and protein–protein interactions (PPIs), especially on domain fusion (9,10) and domain–domain interactions (DDIs) (11–14). To analyze and compare different domain combinations, a topology of co-occurring domains called domain graph was introduced (15). The highly connected nodes or versatile nodes in the graph characterize functional hubs in various cellular facets (15,16) and functional homogeneity (17). Domain distance was proposed to measure the similarity between two DAs for investigating protein evolution. The number of mismatched domains in the alignment relates to the number of evolutionary events (18) and proteins having the same DA tend to evolve from the same ancestor (19).

Several web servers concerning protein domain analyses and visualization are available. Among them are CDART (20), PDART (21), PfamAlyzer (22) and DAhunter (23), all of which mainly serve for homology search based on domain architectures. CADO (17) web server allows a user to query a domain graph and compare domain combinations among the organisms in their built-in database. TreeDomViewer (24) web server provides a visualization tool that incorporates protein domain information over a phylogenetic tree. PhyloDome (25) web server provides a quick visualization of lineage specific distribution of protein domains. In this article, we propose a new web server, d-Omix, which is distinct from previously developed servers in two aspects. First, it integrates various analyses of domain combinations into a unified and comparative platform. Second, all services except the building of putative protein network are applicable with various domain search tools.

## WEB SERVER IMPLEMENTATION

The d-Omix web server is organized into five sections: Data tab for data submission and four services including Tree tab for comparative protein evolution based on

*To whom correspondence should be addressed. Tel: +66 0 2564 7000, ext. 5536; Fax: +66 0 2564 6607; Email: duangdao.wic@biotec.or.th



d-Omix web server is organized into five sections: ***Data*** for data submission and four services: ***Tree*** for comparative protein evolution based on domain distances, ***Alignment*** for comparative proteomes based on domain architecture alignments, ***Graph*** for comparative domain combination based on domain graphs, and ***Interaction*** for building putative protein-protein interactions (PPIs) from domain-domain interactions (DDIs).

**Example1** contains data sets of proteins from the Arabidopsis and rice proteomes that are related by domain architectures in Pfam of the three microRNA processing proteins in Arabidopsis: DCL1 (AT1G01040), AGO1(AT1G48410) and DRB4 (AT3G62800). It is mainly used to compare the conservation and diversification of microRNA processing proteins between the two plants and to imply functions of protein domains with unknown functions (DUF283, DUF1785). It is the example in the manuscript.

| Merge | Select | Dataset | Filename | Operation | No of proteins |
|---|---|---|---|---|---|
| ☑ | ☑ | TAIR8 | TAIR8_prots_DUF283_DUF1785_dsrm.iprscan.hmmpfam | Delete | 35 |
| □ | ☑ | TAIR8_same | TAIR8_prots_DUF283_DUF1785_dsrm.iprscan.hmmpfam | Delete | 35 |
| □ | ☑ | TAIR8_select | TAIR8_selected_for_interaction.iprscan.hmmpfam | Delete | 9 |
| ☑ | ☑ | TIGR6 | TIGR6_prots_DUF283_DUF1785_dsrm.iprscan.hmmpfam | Delete | 53 |

| Source protein | Number of domains | hmmpfam DA ACs/IDs | TAIR8_same | TIGR6 |
|---|---|---|---|---|
| AT1G31290.1 | 3 | PF08699 PF02170 PF02171 | 11 | 22 |
| AT1G31280.1 | 3 | PF08699 PF02170 PF02171 | 11 | 22 |
| AT1G69440.1 | 3 | PF08699 PF02170 PF02171 | 11 | 22 |
| AT1G09700.1 | 2 | PF00035 PF00035 | 8 | 11 |
| AT1G80650.1 | 1 | PF00035 | 6 | 4 |
| AT1G01040.1 | 8 | PF00270 PF00271 PF03368 PF02170 PF00636 PF00636 PF00035 PF00035 | 1 | 1 |
| AT1G48410.2 | 3 | PF08699 PF02170 PF02171 | 11 | 22 |
| AT1G48410.1 | 3 | PF08699 PF02170 PF02171 | 11 | 22 |
| AT1G48650.1 | 5 | PF00270 PF00271 PF04408 PF07717 PF00035 | 3 | 1 |
| AT2G01130.1 | 5 | PF00270 PF00271 PF04408 PF07717 PF00035 | 3 | 1 |
| AT2G28380.1 | 2 | PF00035 PF00035 | 8 | 11 |
| AT2G27880.1 | 3 | PF08699 PF02170 PF02171 | 11 | 22 |

| Source protein | TAIR8_same | DA score |
|---|---|---|
| AT1G48410.1 | AT1G31290.1 | 0 |
| AT1G48410.1 | AT1G48410.2 | 0 |
| AT1G48410.1 | AT1G48410.1 | 0 |
| AT1G48410.1 | AT2G27040.1 | 0 |

| No. | Domain | Versatility | Co-present domain | Co-present abundanc |
|---|---|---|---|---|
| 1 | Methyltransf_12 | 1 | dsrm | 1 |
| 2 | HA2 | 4 | DUF1605<br>DEAD<br>Helicase_C<br>dsrm | 5<br>4<br>5<br>5 |
| 3 | DUF1605 | 4 | DEAD<br>HA2<br>Helicase_C<br>dsrm | 4<br>5<br>5<br>5 |
| 4 | DUF283 | 5 | DEAD<br>Helicase_C<br>PAZ<br>Ribonuclease_3<br>dsrm | 8<br>9<br>11<br>11<br>10 |
|  |  |  | DUF1785 | 34 |

| Domain pair | - | Protein name | Data set | Color |
|---|---|---|---|---|
|  | □ | AT3G03300.1 | TAIR8 | 🟩 |
|  | □ | AT3G03300.2 | TAIR8 | 🟩 |
|  | □ | AT5G20320.1 | TAIR8 | 🟩 |
|  | □ | LOC_Os10g34430.1 | TIGR6 | 🟪 |
| DUF283--PAZ | □ | LOC_Os01g68120.1 | TIGR6 | 🟪 |

Domain interaction detail [AT1G01040.1--AT3G62800.1] [All DDIs]

| No. | Domain1 ID | Domain1 AC | Domain2 ID | Domain2 AC | Number of all putative PPIs | iPfam | 3did | ME | RCDP | Pvalue | Fusion | NetOpt | RDFF | PP | Prediction confidence |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | PF00035 | dsrm | PF00035 | dsrm | 28 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | HC |
| 2 | PF00035 | dsrm | PF00271 | Helicase_C | 18 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | HC |
| 3 | PF00035 | dsrm | PF00636 | Ribonuclease_3 | 22 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | HC |
| 4 | PF00035 | dsrm | PF00270 | DEAD | 18 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | LC |
| 5 | PF00035 | dsrm | PF03368 | DUF283 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | LC |

**Figure 1.** Screenshots of the d-Omix web interface. (**A**) Data tab with *Example1* data sets. (**B**) The DA-based tree generated from the *mergefile* between TAIR8 and TIGR6 data sets in *Example1*. (**C**) The alignment results between TAIR8 and TAIR8_same data sets which are the same set of proteins from Arabidopsis and between Arabidopsis as source- (TAIR8) and rice as target- (TIGR6) data sets. (**D**) The domain graph built from the *mergefile* in *Example1*. The highlighted node in the domain graph corresponds with the highlighted row in the summary table on the left. Colors of the edges in the graph indicate different sources of protein sets. (**E**) A putative protein network of TAIR8_select data set with the detailed DDIs between DCL1 (AT1G01040.1) and DRB4 (AT3G62800.1) proteins.

domain distances; Graph tab for comparative domain combination based on domain graphs; Alignment tab for comparative proteomes based on domain architecture alignments; and Interaction tab for building a putative protein interaction network from DDIs.

## Data submission

The d-Omix web server requires an InterProScan (26) file in raw format as an input. Under Data tab, users may upload multiple files and merge some of them for the comparative analyses across protein sets (e.g. among pathways in the same organism or among organisms for the same pathway). Normally, InterProScan files generated from the proteomes of model organisms with genome sequences will be available (e.g. TAIR8_all.domains of *Arabidopsis thaliana* (Arabidopsis) from http://www.arabidopsis.org/, all.interpro of TIGR Rice release 6 from http://rice.Plantbiology.msu.edu/). Users with only protein sequences could also prepare the InterProScan file using feature 'Prepare InterProScan file'. Figure 1A shows Data tab with *Example1* data sets of proteins from the Arabidopsis and rice proteomes that are related by DAs to the three microRNA-processing proteins in Arabidopsis: DCL1 (AT1G01040), AGO1 (AT1G48410) and DRB4 (AT3G62800).

All services of d-Omix are composed of input data sets selected and/or merged from Data tab and eleven domain search tools incorporated with InterProScan. Users may choose only some data sets and domain search tools for a specific running. An analysis for a large data set will be batched. The results of all services will be presented as a series of tabs of chosen data sets for the highlighted search tool. Users may switch the representation between protein domain ID (e.g. PF03368) and AC (e.g. DUF283) if both are available in the input data files. The click on a protein domain will link to its corresponding online database.

## Comparative protein evolution

The Tree tab enables users to explore common ancestors, conservation, or linage-specific DAs among proteins.

It uses PHYLIP (27) to construct a phylogenetic tree for a selected protein set from a distance matrix of DA scores calculated from all pairs of proteins. CLUSTALW (28) is also incorporated to enable the building of alternative phylogenetic tree based on global sequence alignments. The DA-based tree complements the sequence-based tree. It reveals the closest neighbor for each domain architecture and efficiently categorizes multi-domain proteins that are distantly related or containing 'promiscuous domains' (18). Promiscuous domains such as PF00017 (SH2) and PF00400 (WD40) are small, versatile, typically repetitive and occurring in proteins with a variety of functions (9). Users may compare trees generated from different domain search tools (e.g. hmmpfam, hmmsmart, etc.) or distance matrixes (e.g. DA-based, sequence-based). Proteins with the same or similar DAs will be clustered together. Colors of proteins in trees indicate their source data sets. Users may export trees into SVG, JPG, BMP or NEWICK format and edit the tree using PhyloWidget (29). The DA-based tree built from the *mergefile* data set in *Example1* (Figure 1B) reveals the conservation of Dicer and Argonaute proteins between the Arabidopsis and rice. The clustered sets are categorized by their detailed DAs that might be caused by domain insertion/deletion, suggesting possible functional modifications. In addition, it suggests specific co-occurrences of the PAZ (PF02170), DUF1785 (PF08699) and Piwi (PF02171) domains in the cluster of Argonaute proteins and the PAZ and DUF283 domains in the Dicer proteins.

### Comparative proteome

The Alignment tab enables users to compare the similarity and explore the diversification of proteins based on domain architectures within and across data sets. It calculates DA scores for all pairs of proteins between source- and target- data sets. It is analogous to BLAST with DA based comparison. Users may limit the alignment results using the DA score and hit limit; the lower DA score represents the more similar DAs. The alignment results are summarized in a table, where each row shows a protein name with its DA from the source data set and the number of proteins hit with satisfying DA scores from each target data set. The number of hits suggests DA conservation, proteins with redundant or related functions, and possible annotations for unknown proteins. To explore the alignments in detail, users may click for further information on the hit number. Figure 1C shows the alignment results within the same set of proteins from Arabidopsis and between Arabidopsis as source- and rice as target- data sets. Results with the exact matched DA (DA score = 0) show that most Arabidopsis proteins hit some rice proteins with the same DA. There are 11 and 22 proteins respectively in Arabidopsis and rice having exactly the same DA as of AGO1 (AT1G48410) protein in Arabidopsis.

### Comparative domain combination

The Graph tab builds domain graphs (15) that enable users to (i) investigate the versatility and abundance of protein domains and domain pairs, (ii) explore the modularity of protein domains based on clustering coefficient (30) and (iii) compare shared and specific domain pairs across data sets. The results include a summary table with sortable versatility and abundance of all domains occurring in protein sequences of a selected data set and domain search tool. The click on a domain in the summary table will highlight its corresponding node and neighbors (co-present domains) in the domain graph on the right with the clustering coefficient. Domains in a small cluster with clustering coefficient close to 1 tend to have high functional homogeneity (17). The number of neighbors of a domain in the graph represents versatility of the domain. Most versatile domains tend to be functional centers in different biological aspects (15,16). The click on a co-present abundance number in the summary table or on an edge label in the graph will provide its corresponding protein list for the domain pair. An arrowed edge in a domain graph with direction indicates the presence of both domains in a consecutive order from N- to C- terminals. Users may save domain graphs into SVG, JPG, BMP, or DOT format and further explore a large graph using ZGRViewer (31) with smooth zoomable features. The domain graph built from the *mergefile* in *Example1* is shown in Figure 1D. The functions of DUF283 domain and its neighbors (e.g. PAZ, dsrm) tend to be homogeneous. This corresponds with the previous report that DUF283 domain contains a double-stranded RNA-binding fold and involves in siRNA/miRNA selection (32). The co-presence of DUF1785, Piwi and PAZ domains in both Arabidopsis and rice proteins suggests their related functions in RNA silencing of AGO1.

### Building putative protein interaction

The Interaction tab allows users to investigate possible PPIs for an input protein set. It builds a putative protein interaction network based on DDIs from DOMINE (33). Each edge between a putative PPI represents an existing DDI between the two proteins, where its color denotes the DDI confidence level from DOMINE. Users may filter the network based on these confidence levels. The DA alignment detail on the right shows the DAs of all participating proteins in the network on the left. The click on a PPI in the network will limit the DA alignment detail on the right to the DAs of the two proteins of the PPI. The click on a domain with the DDIs between the two proteins will highlight the domain and its interacting partners. The DDI tab lists all source DDIs of the current putative protein interaction network. Users may filter the network to focus on a specific domain of interest and its DDIs. All PPIs in the protein network are listed under PPI tab and interactively updated according to filtering conditions. The more number of DDIs with high confidence level between a protein pair suggests a higher chance of protein interaction. All DDIs of a PPI will be shown under DDIofPPI tab when the number of all DDIs is clicked. Similar to domain graphs, users may send the current protein network to ZGRViewer for smooth zoomable features. Figure 1E shows a putative protein network of selected proteins from Arabidopsis (TAIR8_select data set in

*Example1*) where each protein is the representative of a cluster or group of proteins with the same DA resulted from the DA-based tree and DA alignment. The possible PPI between DCL1 (AT1G01040.1) and DRB4 (AT3G62800.1) proteins come from five DDIs from DOMINE where three of them show high DDI confidence level. DRB4 and HYL1 (AT1G09700.1) have been reported to interact with DCL4 (AT5G20320.1) and DCL1, respectively (34). DCL4 has the same DA as of DCL1 while HYL1 has the same DA as of DRB4. While a putative PPI might not have the exact participating partners, it suggests and/or narrows down possible partners and their related domains for the interaction.

## METHODS

### DA score

DA score measures the similarity between two protein sequences based on the alignment of their DAs. With one protein as source and the other as target, their protein domain units from N-terminal to C-terminal will be orderly compared and scored with the following function.

$$
\begin{aligned}
&\mathrm{DA}_{\mathrm{P1-P2,Pfam}} && \\
&= 0; && \text{if } \mathrm{MDR}_{\mathrm{Pfam}} = 1 \\
&= 1; && \text{if } \mathrm{MDR}_{\mathrm{Pfam}} = 0 \\
&= 1 - \mathrm{MDR}_{\mathrm{Pfam}} && \\
&\quad + \{\mathrm{Gap}/[\mathrm{len(longer\_DA)} * 10]\}; && \text{otherwise}
\end{aligned}
$$

We define the Matched Domain Ratio$_{\mathrm{Pfam}}$ (MDR$_{\mathrm{Pfam}}$) as the number of matched Pfam domains, in conserved order between proteins P1 and P2 over the total number of Pfam domains in the longer DA defined as len(longer_DA). Gap is the number of the inserted gaps during the alignment of the two DAs. The number 10 in the above equation is introduced to make gap penalties small. The small gap penalties are necessary to be included to avoid sporadic gaps in long repeating regions (18). A DA score between two proteins is calculated for individual sources of protein domains (e.g. Pfam, SUPERFAMILY); the lower the DA score, the more similar DAs between the two proteins. DA score is fundamental for both tree and alignment services.

### Domain graph

A domain graph is an undirected graph where each vertex represents a protein domain and an edge between two domains indicates the co-presence of the two domains on at least a protein sequence (15). The d-Omix builds a domain graph according to this definition and extends it with direction. A domain graph is drawn using GraphViz (35) via PHP GraphViz extension.

## ACKNOWLEDGEMENTS

We thank Dr Anan Jongkaewwattana and Dr Sithichoke Tangphatsornruang for editing the manuscript. We also thank Wuttichai Mhuantong for preparing figure in this article, and Eakasit Pacharawongsakda and Sunai Yokwai for their technical assistances including external software installation.

## FUNDING

Cluster Program Management Office (CPMO) [grant FC0033 B21 (SPA B1-1)]; National Science and Technology Development Agency (NSTDA), Thailand. Funding for open access charge: Cluster Program Management Office (FC0033 B21).

*Conflict of interest statement*. None declared.

## REFERENCES

1. Orengo,C.A., Michie,A.D., Jones,S., Jones,D.T., Swindells,M.B. and Thornton,J.M. (1997) CATH – a hierarchic classification of protein domain structures. *Structure*, **5**, 1093–1108.
2. Murzin,A.G., Brenner,S.E., Hubbard,T. and Chothia,C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536.
3. Bashton,M. and Chothia,C. (2007) The generation of new protein functions by the combination of domains. *Structure*, **15**, 85–99.
4. Bashton,M. and Chothia,C. (2002) The geometry of domain combination in proteins. *J. Mol. Biol.*, **315**, 927.
5. Hegyi,H. and Gerstein,M. (2001) Annotation transfer for genomics: measuring functional divergence in multi-domain proteins. *Genome Res.*, **11**, 1632–1640.
6. Forslund,K. and Sonnhammer,E.L.L. (2008) Predicting protein function from domain content. *Bioinformatics*, **24**, 1681–1687.
7. Mott,R., Schultz,J., Bork,P. and Ponting,C.P. (2002) Predicting protein cellular localization using a domain projection method. *Genome Res.*, **12**, 1168–1174.
8. Bork,P., Hofmann,K., Bucher,P., Neuwald,A.F., Altschul,S.F. and Koonin,E.V. (1997) A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. *FASEB J.*, **11**, 68–76.
9. 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.
10. Enright,A.J., Iliopoulos,I., Kyrpides,N.C. and Ouzounis,C.A. (1999) Protein interaction maps for complete genomes based on gene fusion events. *Nature*, **402**, 86–90.
11. Han,D.-S., Kim,H.-S., Jang,W.-H., Lee,S.-D. and Suh,J.-K. (2004) PreSPI: a domain combination based prediction system for protein-protein interaction. *Nucleic Acids Res.*, **32**, 6312–6320.
12. Chen,X.-W. and Liu,M. (2005) Prediction of protein-protein interactions using random decision forest framework. *Bioinformatics*, **21**, 4394–4400.
13. Kim,I., Liu,Y. and Zhao,H. (2007) Bayesian methods for predicting interacting protein pairs using domain Information. *Biometrics*, **63**, 824–833.
14. Huang,C., Morcos,F., Kanaan,S.P., Wuchty,S., Chen,D.Z. and Izaguirre,J.A. (2007) Predicting protein-protein interactions from protein domains using a set cover approach. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, **4**, 78–87.
15. Wuchty,S. (2001) Scale-free behavior in protein domain networks. *Mol. Biol. Evol.*, **18**, 1694–1702.
16. Apic,G., Gough,J. and Teichmann,S.A. (2001) Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J. Mol. Biol.*, **310**, 311.
17. Ye,Y. and Godzik,A. (2004) Comparative analysis of protein domain organization. *Genome Res.*, **14**, 343–353.
18. Bjorklund,A.K., Ekman,D., Light,S., Frey-Skott,J. and Elofsson,A. (2005) Domain rearrangements in protein evolution. *J. Mol. Biol.*, **353**, 911.
19. Vogel,C., Bashton,M., Kerrison,N.D., Chothia,C. and Teichmann,S.A. (2004) Structure, function and evolution of multidomain proteins. *Curr. Opin. Struct. Biol.*, **14**, 208.

20. Geer,L.Y., Domrachev,M., Lipman,D.J. and Bryant,S.H. (2002) CDART: protein homology by domain architecture. *Genome Res.*, **12**, 1619–1623.
21. Lin,K., Zhu,L. and Zhang,D.-Y. (2006) An initial strategy for comparing proteins at the domain architecture level. *Bioinformatics*, **22**, 2081–2086.
22. Hollich,V. and Sonnhammer,E.L.L. (2007) PfamAlyzer: domain-centric homology search. *Bioinformatics*, **23**, 3382–3383.
23. Lee,B. and Lee,D. (2008) DAhunter: a web-based server that identifies homologous proteins by comparing domain architecture. *Nucleic Acids Res.*, **36**, W60–W64.
24. Alako,B.T.F., Rainey,D., Nijveen,H. and Leunissen,J.A.M. (2006) TreeDomViewer: a tool for the visualization of phylogeny and protein domain structure. *Nucleic Acids Res.*, **34**, W104–W109.
25. Novatchkova,M., Wildpaner,M., Schweizer,D. and Eisenhaber,F. (2005) PhyloDome–visualization of taxonomic distributions of domains occurring in eukaryote protein sequence sets. *Nucleic Acids Res.*, **33**, W121–W125.
26. Quevillon,E., Silventoinen,V., Pillai,S., Harte,N., Mulder,N., Apweiler,R. and Lopez,R. (2005) InterProScan: protein domains identifier. *Nucleic Acids Res.*, **33**, W116–W120.
27. Felsenstein,J. (2005) *PHYLIP (Phylogeny Inference Package) version 3.67*. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
28. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
29. Jordan,G.E. and Piel,W.H. (2008) PhyloWidget: web-based visualizations for the tree of life. *Bioinformatics*, **24**, 1641–1642.
30. Watts,D.J. and Strogatz,S.H. (1998) Collective dynamics of ‘small-world’ networks. *Nature*, **393**, 440–442.
31. Pietriga,E. (2005) A toolkit for addressing HCI issues in visual language environments. *IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pp. 145–152.
32. Dlakic,M. (2006) DUF283 domain of Dicer proteins has a double-stranded RNA-binding fold. *Bioinformatics*, **22**, 2711–2714.
33. Raghavachari,B., Tasneem,A., Przytycka,T.M. and Jothi,R. (2008) DOMINE: a database of protein domain interactions. *Nucleic Acids Res.*, **36**, D656–D661.
34. Hiraguri,A., Itoh,R., Kondo,N., Nomura,Y., Aizawa,D., Murai,Y., Koiwa,H., Seki,M., Shinozaki,K. and Fukuhara,T. (2005) Specific interactions between Dicer-like proteins and HYL1/DRB- family dsRNA-binding proteins in Arabidopsis thaliana. *Plant Mol. Biol.*, **57**, 173–188.
35. Gansner,E.R. and North,S.C. (1999) An open graph visualization system and its applications to software engineering. *Softw.–Pract. Exper.*, **30**, 1203–1233.