Bioinformatics, 32(9), 2016, 1427–1429
doi: 10.1093/bioinformatics/btv765
Advance Access Publication Date: 5 January 2016
Applications Note

OXFORD

Systems biology

Path2PPI: an R package to predict protein–protein interaction networks for a set of proteins

Oliver Philipp[1,2], Heinz D. Osiewacz[2] and Ina Koch[1,*]

[1]Molecular Bioinformatics, Institute of Computer Science, Faculty of Computer Science and Mathematics & Cluster of Excellence 'Macromolecular Complexes', Goethe-University, Frankfurt am Main, Germany and [2]Molecular Developmental Biology, Institute of Molecular Biosciences, Faculty for Biosciences & Cluster of Excellence 'Macromolecular Complexes', Goethe-University, Frankfurt am Main, Germany

*To whom correspondence should be addressed.
Associate Editor: Ziv Bar-Joseph

Received on 19 August 2015; revised on 18 November 2015; accepted on 25 December 2015

Abstract

Summary: We introduce Path2PPI, a new R package to identify protein–protein interaction (PPI) networks for fully sequenced organisms for which nearly none PPI are known. Path2PPI predicts PPI networks based on sets of proteins from well-established model organisms, providing an intuitive visualization and usability. It can be used to combine and transfer information of a certain pathway or biological process from several reference organisms to one target organism.
Availability and implementation: Path2PPI is an open-source tool implemented in R. It can be obtained from the Bioconductor project: http://bioconductor.org/packages/Path2PPI/
Contact: ina.koch@bioinformatik.uni-frankfurt.de
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Plenty of databases exist which contain protein–protein interaction (PPI) data for various organisms (e.g. Chatr-aryamontri *et al.*, 2015; Franceschini *et al.*, 2013). For some well-established model organisms, species-specific data repositories are available (e.g. Güldener *et al.*, 2006; Prasad *et al.*, 2009) providing also PPI data. In contrast, for the majority of organisms such a comprehensive amount of PPI data is not available. Therefore, different approaches have been developed to predict PPIs. Some of these approaches aim to deduce new interactions from known PPIs by means of homology-based mapping based on sequence similarity. Other methods apply supervised learning to filter and score the predicted interactions, using additional biological data, e.g. functional annotation, co-expression and / or text-mining data (Yu *et al.*, 2010; for a recent review see Rao *et al.*, 2014). Approaches that predict PPIs based on sequence data or network topology information often provide only precomputed data (Franceschini *et al.*, 2013; Pesch and Zimmer, 2013) or predict interactions only for a set of predefined organisms (Deng *et al.*, 2013; Wiles *et al.*, 2010). Furthermore, most of the methods do not supply information about the underlying reference interactions. In the majority, only the scores are provided to validate a predicted interaction. Often it is necessary to easily access the entire underlying information about the predicted interactions since the interpretation and experimental validation is one of the most important steps after the prediction.

As we were interested in aging processes and interaction networks of age-related pathways in the fungal model organism *Podospora anserina* (Osiewacz *et al.*, 2013; Philipp *et al.*, 2013), we found only a few data repositories for interaction data. For example, the *KEGG* database (Kanehisa *et al.*, 2014) provides small subnetworks of some selected, mainly metabolic, pathways. Recently, also the *STRING* database (e.g. Franceschini *et al.*, 2013) involves some predicted interactions for *P.anserina*. Nevertheless, there was no satisfactory solution and no easy and fast way to directly gain knowledge about proteins and their interactions of certain biological processes, which are well established in some model organisms, but nearly unknown in the target organism. Homology-based tools which theoretically enable to predict or transfer interactions between species are mostly implemented for a set of predefined organisms or

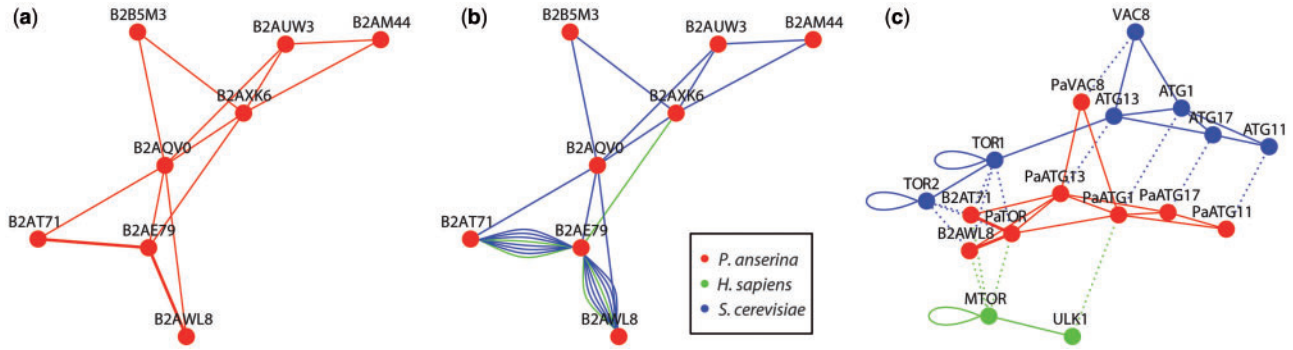

Fig. 1. The predicted PPI network of autophagy induction in *P.anserina* based on the corresponding PPIs in human and yeast. (**a**) The predicted PPI network (*normal* view). The edge thickness corresponds to the scores and the number of reference species showing the interaction. (**b**) *Detailed* view of the PPI network. Each edge is specifically colored (see legend), indicating in which reference species this interaction occurs. Multiple edges represent multiple findings of an interaction. That means, that e.g. the interaction of the proteins 'B2AT71' and 'B2AE79' was found six times in yeast (blue edges) and two times in human (green edges). (**c**) *Hybrid* network representation of the predicted PPI: The relevant parts of the PPI networks of the reference species are included together with the predicted PPI network of the target species. Interactions are depicted as solid lines in the respective color (see legend). Homologous relations of the reference proteins to those of the target species are drawn as dotted edges in the respective color of the target node

require pairs of proteins in the target organism to decide whether they may interact (Chen *et al.*, 2009; Murakami and Mizuguchi, 2014). Here, we report the implementation of PATH2PPI which helps finding proteins and interactions of certain pathways or biological processes in each fully sequenced organism without the need for predefinition of putative proteins and interactions.

2 Features

Using PATH2PPI, the user can choose up to seven of the most established model organisms (human, mouse, rat, yeast, *E.coli*, *C.elegans* and *D.melanogaster*). Based on sets of proteins from these reference species PATH2PPI uses the interaction repository *iRefIndex* (Razick *et al.*, 2008) to find the corresponding relevant interactions. We implemented a more flexible and comfortable search engine than provided by the *iRefR* package (Mora and Donaldson, 2011). Additionally, PATH2PPI requires results of *NCBI BLAST+* (Camacho *et al.*, 2009) searches of all reference species against the target species. Based on these data, PATH2PPI computes new interactions in the target species and scores them. The score is based on the degree of homology and the number of reference species which show the corresponding interaction. A major advantage of PATH2PPI is the easy access to the underlying reference interactions, i.e. all information provided by *iRefIndex*, e.g. source database, interaction type and reference publication. Based on the *igraph* package (Csardi and Nepusz, 2006) the computed PPI can directly be visualized in R (see Fig. 1).

3 Implementation

PATH2PPI can be obtained from the Bioconductor project (Huber *et al.*, 2015). It contains a comprehensive tutorial and for the case study, data files necessary to predict interactions of the induction step of autophagy in *P.anserina* by means of the corresponding PPIs in human and yeast. There are three types of visualization methods available, the *normal*, *detailed* and *hybrid* (Fig 1a–c). Additionally, detailed information about each interaction can be obtained. Results are provided as data frame or as *igraph* objects, enabling for subsequent analyses in R or in advanced analysis tools like *Cytoscape* (Cline *et al.*, 2007). Through the *S4* class architecture PATH2PPI can be easily extended by further prediction and validation algorithms. The example depicted in Figure 1, the prediction algorithm and all features of PATH2PPI are described in detail in the tutorial, see the supplement and the corresponding Bioconductor web site.

4 Conclusion

We introduced a new R package to predict PPI networks based on sets of proteins which may belong to a specific biological pathway, providing an intuitive visualization and usability. We implemented PATH2PPI to reveal putative proteins and interactions for a pathway or a biological process in organisms for which nearly none PPI information is available. The results can serve as starting points for further network modeling studies and experimental validations.

Acknowledgements

We thank Heiko Giese, Hendrik Schäfer and Tim Schäfer for testing and the reviewers for the valuable comments.

Funding

This work was partly funded by the LOEWE program of the State of Hesse (Germany), the 'Bundesministerium für Bildung und Forschung' (BMBF) and by the Prof. Dr. Dieter Platt Stiftung.

Conflict of Interest: none declared.

References

Camacho,C. *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinf.*, **10**, 421.

Chatr-aryamontri,A. *et al.* (2015) The BioGRID interaction database: 2015 update. *Nucleic Acids Res.*, **43**, D470–D4708.

Chen,C.C. *et al.* (2009) PPIsearch: a web server for searching homologous protein–protein interactions across multiple species. *Nucleic Acids Res.*, **37**, W369–W375.

Cline,M.S. *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.*, **2**, 2366–2382.

Csardi,G. and Nepusz,T. (2006) The igraph software package for complex network research. *Int. J.*, **1695**, 1–9.

Deng,Y. *et al*. (2013) ppiPre: predicting protein–protein interactions by combining heterogeneous features. *BMC Syst. Biol*., **7**, S8.

Franceschini,A. *et al*. (2013) STRING v9. 1: protein–protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*., **41**, D808–D815.

Güldener,U. *et al*. (2006) MPact: the MIPS protein interaction resource on yeast. *Nucleic Acids Res*., **34**, D436–D441.

Huber,W. *et al*. (2015) Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods*, **12**, 115–121.

Kanehisa,M. *et al*. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res*., **42**, D199–D205.

Mora,A. and Donaldson,I.M. (2011) iRefR: an R package to manipulate the iRefIndex consolidated protein interaction database. *BMC Bioinf*., **12**, 455.

Murakami,Y. and Mizuguchi,K. (2014) Homology-based prediction of interactions between proteins using averaged one-dependence estimators. *BMC Bioinf*., **15**, 213.

Osiewacz,H.D. *et al*. (2013) Assessing organismal aging in the filamentous fungus *Podospora anserina*. *Methods Mol. Biol*., **965**, 439–462.

Pesch,R. and Zimmer,R. (2013) Complementing the eukaryotic protein interactome. *PLoS ONE*, **8**, e66635.

Philipp,O. *et al*. (2013) A genome-wide longitudinal transcriptome analysis of the aging model *Podospora anserina*. *PLoS ONE*, **8**, e83109.

Prasad,T.K. *et al*. (2009) Human protein reference database 2009 update. *Nucleic Acids Res*., **37**, D767–D772.

Rao,V.S. *et al*. (2014) Protein–protein interaction detection: Methods and analysis. *Int. J. Proteomics*, **2014**, 147648.

Razick,S. *et al*. (2008) iRefIndex: a consolidated protein interaction database with provenance. *BMC Bioinf*., **9**, 405.

Wiles,A.M. *et al*. (2010) Building and analyzing protein interactome networks by cross-species comparisons. *BMC Syst. Biol*., **4**, 36.

Yu,J. *et al*. (2010) Simple sequence-based kernels do not predict protein–protein interactions. *Bioinformatics*, **26**, 2610–2614.