BMC Bioinformatics

Poster presentation

InteroPORC: an automated tool to predict highly conserved protein interaction networks

Magali Michaut*[1], Samuel Kerrien[2], Luisa Montecchi-Palazzi[2], Corinne Cassier-Chauvat[1], Franck Chauvat[1], Jean-Christophe Aude[1], Pierre Legrain[1] and Henning Hermjakob[2]

Address: [1]IBITECS, CEA, Gif sur Yvette, F-91191, France and [2]PANDA, EMBL-EBI, Cambridge, CB10 1SD, UK

Email: Magali Michaut* - michaut.bioinfo@gmail.com

* Corresponding author

from Fourth International Society for Computational Biology (ISCB) Student Council Symposium
Toronto, Canada. 18 July 2008

Published: 30 October 2008

BMC Bioinformatics 2008, **9**(Suppl 10):P1 doi:10.1186/1471-2105-9-S10-P1

This abstract is available from: http://www.biomedcentral.com/1471-2105/9/S10/P1

Motivation

Protein-protein interaction networks provide insights into the relationships between the proteins of an organism thereby contributing to a better understanding of cellular processes. Nevertheless, large-scale interaction networks are available for only a few model organisms but lack for most species. Thus, the interolog concept is useful to transfer interactions onto a target species. The idea is to combine known interactions from a source species with orthology relationships between source and target species (see Figure 1). Such transfers have already been done for a limited number of species. However, no software or standard method was available for that purpose so far. That is the reason why we decided to develop such a prediction tool.

Methods

We defined a new inference process, called InteroPorc, combining source interactions with clusters of orthologous proteins. The method is indeed based on the PORC data (Putative ORthologous Cluster) provided by Integr8. The Integr8 database systematically provides all sequenced genomes and their corresponding proteomes (currently 655 organisms). Consequently, these orthologous clusters are of paramount interest since they contain all sequenced organisms. The inference process consisted of two steps. First, we abstracted protein interactions onto orthologous cluster links. For a given source interaction, if both proteins belonged to a cluster, we constructed a link between these two clusters. In the second step, we projected these cluster links onto a specific target species. Practically, for a given link, if both clusters contained a protein from the target species, we predicted an interaction between these proteins.

Results

We applied our automated prediction tool to the cyanobacteria *Synechocystis*. It enabled us to predict a new network of 1,463 protein-protein interactions when less than 200 interactions were experimentally annotated in the databases. In the same way, we predicted for instance 13,469 interactions for the rat.

Availability

This open-source application can either be run online through a web interface or downloaded at http://biodev.extra.cea.fr/interoporc/. To run the tool online, we have collected source interactions from the three manually curated databases IntAct, MINT and DIP. The user just has to indicate the taxonomy identifier of the species he/she is interested in. Running online usually takes two minutes. It is also possible to download the tool for standalone use to get more flexibility. For example, the source interaction dataset can be changed to use only highly rel-

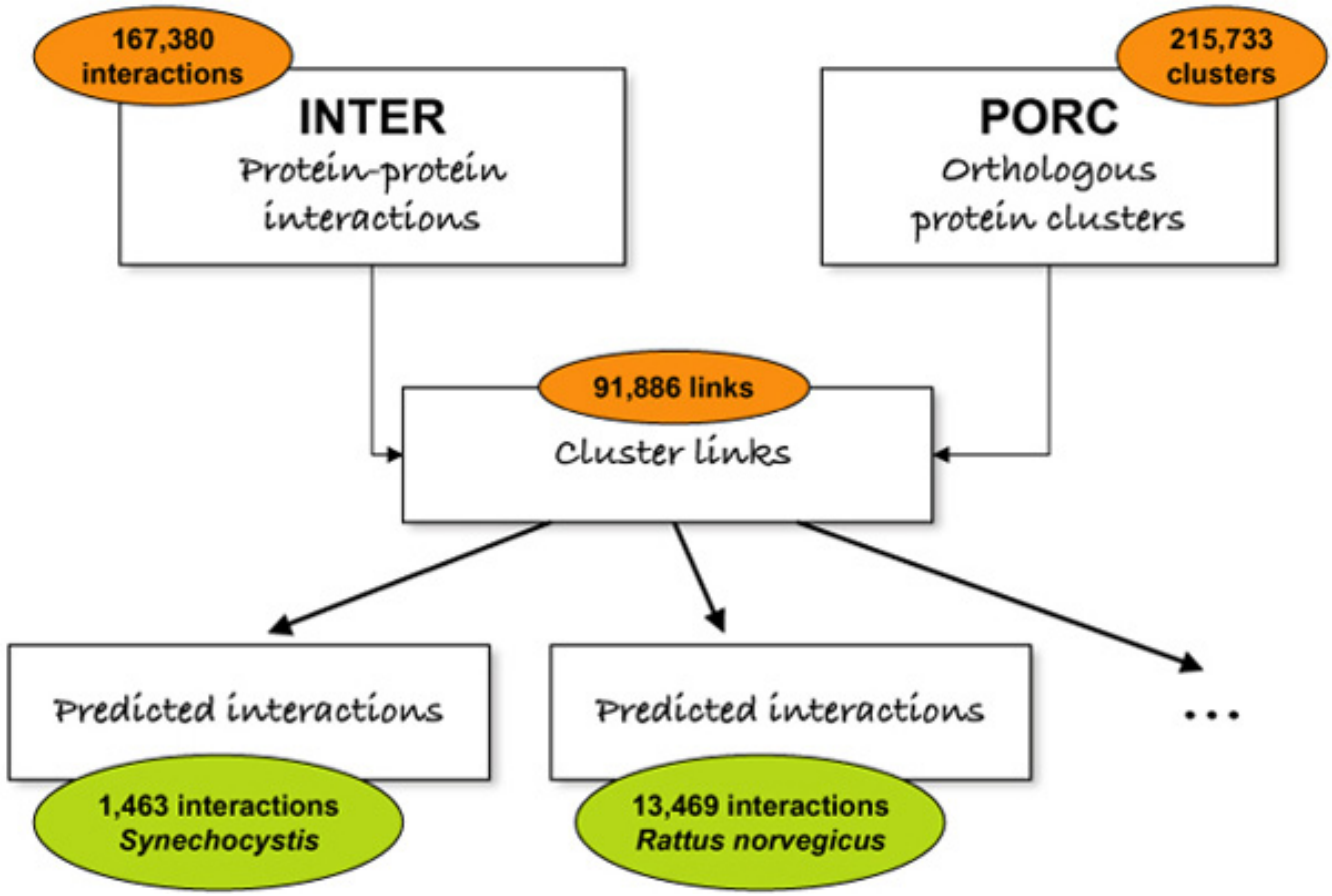

Figure 1
The InteroPorc inference process. The inference process consists of two steps. First, we abstract protein interactions onto cluster links. Then, we project these links onto a specific target species.

evant source interactions or private datasets. Moreover, this application can be run on all platforms since it has been developed in Java.

Conclusion

This tool is highly interesting to quickly get a raw picture of the protein interaction network of any sequenced organism. Moreover, it should greatly facilitate comparative studies since it provides a common method to predict protein interaction networks for lots of species in an automatic way. Finally, it is noteworthy that the method has been implemented separately from the interaction data used. Since the quality of the interactions is still a problem to be addressed, it is of great importance to be able to choose which interactions one would like to transfer.