# Protein interaction networks – now and then

Thomas Weinmaier[1], Dmitrij Turaev[1], Yuliang Jin[2], Hernan Makse[2], Thomas Rattei[1]

[1] Department of Computational Systems Biology, Ecology Center, University of Vienna, 1090 Vienna, Austria
[2] Levich Institute and Department of Physics, City College of New York, New York, NY 10031, USA.

Protein-protein interactions (PPIs) are the foundation of biological functions in all organisms. Therefore, protein interaction networks (PINs) formed by PPIs have been studied extensively for many years. Researchers aim to unscramble the modular organization of the networks and identify the relationship between modules and their biological functions. One approach to better understand modular organization of today's networks is to investigate the structure of ancient networks and their evolution. Since ancient PPIs are experimentally unaccessible, ancestral PINs have to be reconstructed bioinformatically. However, the accuracy of the reconstructed ancestral PINs cannot be verified and there are so far no established means to reconstruct the PINs computationally. Therefore we suggest to combine two concepts which are well-established in other fields of applications for the reconstruction of ancestral PINs introducing a minimal number of assumptions:

(1) Clusters of orthologous groups (COGs) consist of proteins which have evolved from one last common ancestor (LCA) via speciation and duplication events. COGs are used in genomics for the functional annotation of proteins and the analysis of their evolution. The selection of organisms used for the calculation of a COG defines its LCA. Thus, COGs can be used to predict representatives of ancestral proteins at a given time point. The eggNOG database [1] contains COGs calculated relative to different LCAs which means relative to different evolutionary time points. (2) The maximum parsimony approach is used in phylogeny to estimate phylogenetic trees based on observed data. Using maximum parsimony, the most likely phylogenetic tree is the one which requires the smallest number of changes relative to a known species tree. The parameters of the reconstruction are defined in a cost-function for all possible changes in the tree. In the same way, maximum parsimony can be applied to estimate which proteins in the ancestral network really interacted based on today's networks. Therefore the most likely ancestral PIN is the one requiring the smallest number of changes compared to the present-day PIN using a cost-function for the gain and loss of a PPI.

In our analysis we use present-day PINs extracted from the STRING database [2] and map the proteins to their corresponding COGs in the eggNOG database. Using the evolutionary time points defined in eggNOG we reconstruct ancestral networks at these evolutionary levels. Applying maximum parsimony with different parameter sets for the costs of gain and loss of a PPI we reconstruct different networks. We compare the network properties of the different networks and investigate the dependency of the network properties on the chosen parameters.

First results for yeast show that the network properties differ significantly for different parameter sets. We discuss how these results can be used to estimate reasonable parameter ranges for the reconstruction of ancestral PINs.

# References

1: Muller J, Szklarczyk D, Julien P, Letunic I, Roth A, Kuhn M, Powell S, von Mering C, Doerks T, Jensen LJ, Bork P., eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations., Nucleic Acids Res. (Database issue), 2010, D190-195
2: Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C, The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored, Nucleic Acids Res, , D561-8

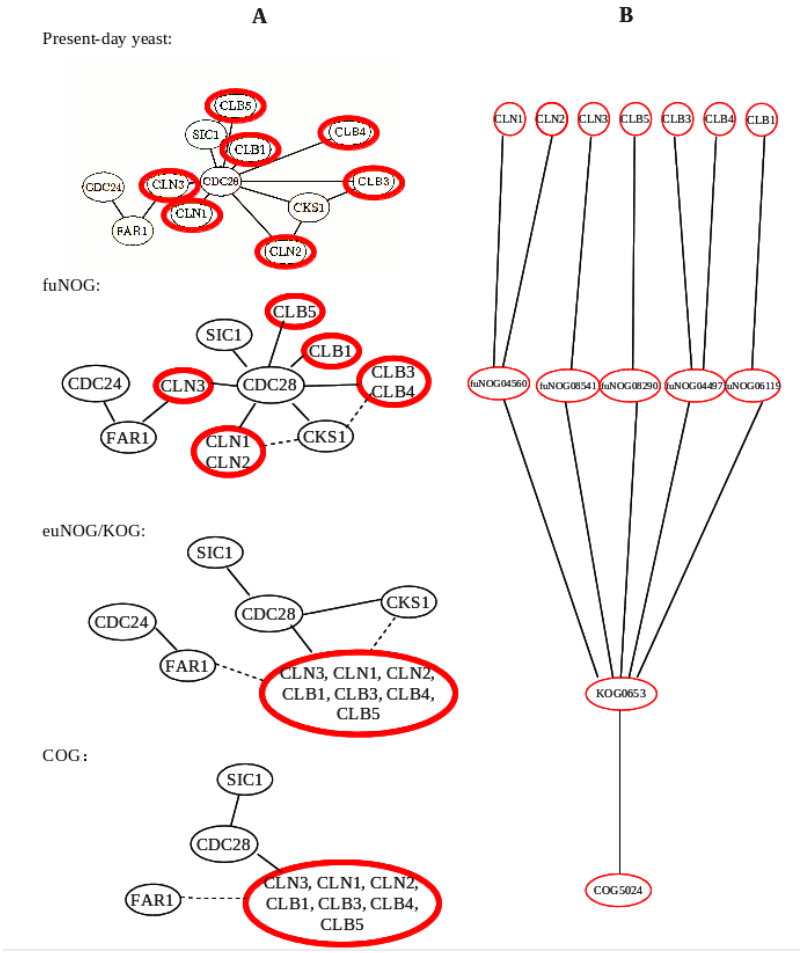


Figure 1: An exemplary subnet of the yeast PIN showing the interactions of CDC28. A: PPI network on the different evolutionary levels described in the eggNOG database when the proteins are mapped on the eggNOG database, B: a schematic phylogenetic tree of the proteins highlighted in red in A and their corresponding COGs in eggNOG.