

## 188. A notion before projecting biological networks across species: have the genes co-occurred through evolution?

Gang Fang<sup>1,2\*</sup> ([gang.fang@yale.edu](mailto:gang.fang@yale.edu)), Doreen Ware,<sup>4,5</sup> Sergei Maslov,<sup>6</sup> and Mark Gerstein<sup>1,2,3</sup>

<sup>1</sup>Department of Molecular Biophysics and Biochemistry, <sup>2</sup>Program in Computational Biology and Bioinformatics, <sup>3</sup>Department of Computer Science, Yale University, New Haven, CT; <sup>4</sup>Cold Spring Harbor Laboratory, NY; <sup>5</sup>United States Department of Agriculture-Agricultural Research Service; <sup>6</sup>Brookhaven National Laboratory, NY

**Project Goals: This university-KBase project is to develop a set of confidence values for cross-species biological network and modules projection.**

KBase and other resources have stored many pre-computed biological networks and gene clusters as functional modules. This changes the way of biological database search from one of querying individual gene's function to examining relationships in a group of genes in networks. Biological network query is particularly useful for studying genes from less investigated species, and often the first step of such studies is to look for clues in database by checking how their orthologs are connected in various functional modules in other species. The reliability of cross-species projection of network information, namely the connections (edges) between genes (nodes) and node attributes, is dependent on many factors. Among which, one factor is fundamental: co-occurrence of genes on phylogenetic tree. If the genes were randomly distributed in phylogeny, projection of their biological connections from one species to another would be less reliable. Therefore, we suggest a set of confidence values for projected modules. First, we propose an evolutionary profile  $\{x_0, x_1, \dots, x_n\}$  describing the conservation of module among  $n$  species, where  $x_i$  is percentage of member genes present in species  $i$ . With this profile, correlation between this module and phylogenetic tree of the  $n$  species is calculated. This is the  $K$  statistic [1]. Significant  $K$  suggests this module is not randomly distributed on phylogenetic tree. In addition, MPD and MNTD  $Z$ -scores are calculated [2]. The two values describe how the module preserves through evolution. Significant MPD  $Z$ -score means the module supports phylogenetic tree level clustering, i.e. significant association between phylogeny co-occurrence and pre-computed biological relationships of the member genes, so that projection of this module is reliable. Significant MNTD  $Z$ -score also supports projected biological relationships between genes. It is different from MPD  $Z$ -score in that it suggests the group of genes subject to co-horizontal transfer or very species-specific. Several study cases of these confidence scores are reported [3] and presented in our poster. A further note is that our strategy is different from the workflow where co-occurrence clusters of input genes is first calculated followed by examining overlaps between co-occurred gene clusters and pre-computed modules. That kind of workflow may introduce extra issues due to inconsistent clustering thresholds etc. Instead, our suggested confidence values leverage the pre-computed functional modules and cluster input genes into groups and then evaluate these modules based on phylogeny of member genes. The projected modules could be further sorted according to their confidence values and provide guidance for subsequent experimental work.

### References:

1. Blomberg, S. P., T. Garland, Jr., and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57:717- 745.
2. Webb, C.O., Ackerly, D.D., and Kembel, S.W. 2008. Phylocom: software for the analysis of phylogenetic community structure and trait evolution. Version 4.0.1.

3. Fang, G., Passalacqua, K.D, Hocking, J., Montero Llopis, P., Gerstein, M., Bergman, N. H., and Jacobs-Wagner, C. 2013. Transcriptomic and phylogenetic analysis of a bacterial cell cycle reveals strong associations between gene co- expression and evolution. *BMC Genomics* 14:450