2011 IEEE International Conference on Bioinformatics and Biomedicine

Detecting Pathway Cross-talks by Analyzing Conserved Functional Modules across Multiple Phenotype-Expressing Organisms

Kevin Wilson[§], Andrea M. Rocha[‡], Kanchana Padmanabhan[*†], Kuangyu Wang[*†], Zhengzhang Chen[*†], Ye Jin [*†], James R. Mihelcic[‡] and Nagiza F. Samatova[*†¶]
[*]*North Carolina State University, Raleigh, NC 27695*
[†]*Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, TN 37831*
[‡]*University of South Florida, Tampa, FL 33620*
[§]*RTI International, Durham, NC 27709*
[¶]*Corressponding author - Samatova@csc.ncsu.edu*

***Abstract*—Biological systems are organized hierarchically, starting from the protein level and expanding to pathway or even higher levels. Understanding interactions at lower levels (proteins interactions) in the hierarchy will help us understand interactions at higher levels (pathway cross-talks). Identifying cross-talks that are related to the expression of a particular-phenotype will be of interest to genetic engineers, because it will provide information on how different cellular subsystems could work together to express a phenotype. Current research has typically focused on identifying genotype-phenotype associations or pathway-phenotype associations. In contrast, we developed a method to identify phenotype-related pathway cross-talks by obtaining conserved groups of interacting proteins (functional modules). By applying our method to two groups of hydrogen producing organisms (light fermentation and dark fermentation), we have shown that our method effectively unearths known pathway cross-talks that are important to hydrogen production.**

***Keywords*-protein functional module; phenotype-expressing organism; pathway; cross-talk;**

I. Introduction

Proteins, such as enzymes, often work together to achieve a particular function. A metabolic pathway is a series of chemical reactions catalyzed by enzymes. Different metabolic pathways may cross-talk (interact) with each other for purposes, such as regulation or compensation. For example, in *Anabaena (Nostoc) sp. PCC7120*, cross-talks between nitrogen, iron, and central metabolism have been observed at regulatory level [1], [2]. Nitrogen metabolism cross-talks with iron uptake pathway in a way that the nitrogen regulator, NtcA, is able to alter the expression of the iron uptake protein, FurA [3]. Interaction between the two proteins is important for maintaining iron homeostasis, which is essential for nitrogen-fixation in organisms such as *Anabaena*. Moreover, in a study by Lopez-Gollomon *et al.* [2], NtcA and FurA were shown to co-regulate the expression of several genes involved in nitrogen metabolism and photosynthesis, which demonstrates how interrelated many metabolic pathways are within microorganisms. Understanding of cross-talks in metabolic networks is particularly important when engineering metabolic pathways for enhanced expression of a trait or desired end-product for industrial use (e.g., ethanol and hydrogen).

II. Approach

A. Overview

To identify cross-talks that contribute to the expression of a specific phenotype, we include multiple phenotype-expressing organisms in our study based on the assumption that phenotype-related cross-talks are likely to be conserved across organisms with the same phenotype.

The phenotype-related cross-talks can be identified by analyzing groups of interacting proteins (functional module) present across multiple phenotype-expressing organismal networks. There are several ways that a functional module is typically modeled, the most common being the clique and cluster models. Cliques are completely connected subgraphs and hence, using clique as a model might not allow us to capture some subtle cross-talking mechanisms that may exist. For example, a cross-talk mechanism where only few proteins from each pathway interact while the rest of the proteins have no interaction will not form a clique. Thus, in this paper we use a cluster to model the conserved functional module. The only restriction we place is that the conserved cluster of proteins must form a connected component. A disconnected set of proteins may not be interacting at all and likely do not cross-talk.

Another factor to be considered is that all phenotype-expressing organisms may not use the same cross talking mechanisms and so it is important to capture signals that may only be present in a particular subset of the organisms.

Thus our method enumerates all the conserved connected components present across all or a subset of the given set of phenotype-expressing organisms. These components are further analyzed for potential cross-talk mechanisms. However, directly comparing the organismal networks to identify these conserved modules might not be tractable. Hence, we

978-0-7695-4574-5/11 $26.00 © 2011 IEEE
DOI 10.1109/BIBM.2011.35

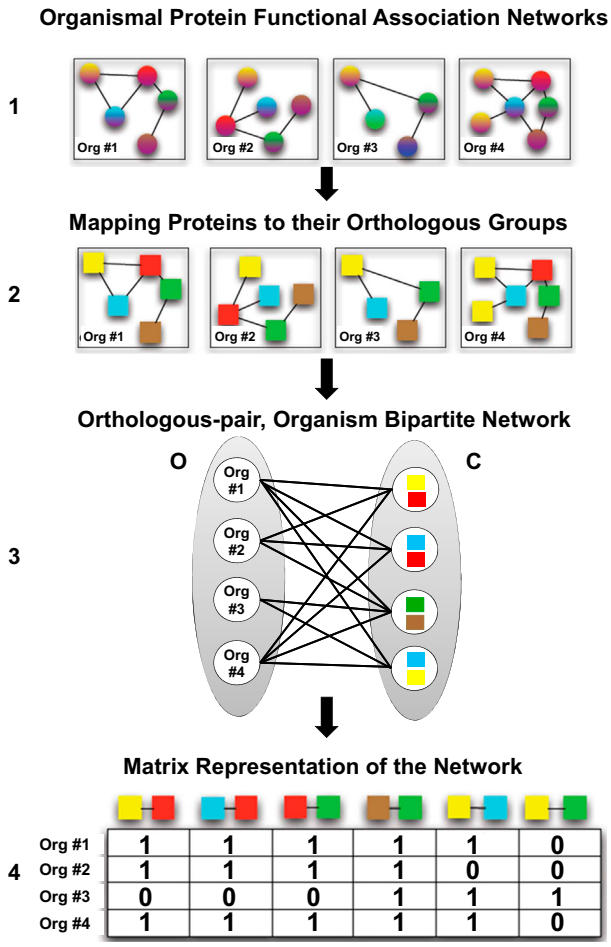

Figure 1. Building the orthologous group-pair, organism bipartite network

propose a new network model called the *orthologous group-pair, organism bipartite network* to capture the information present in all of the organismal networks combined. Using this network we enumerate all the conserved connected components. We also perform statistical analysis to select only those components with significant connectivity.

B. Orthologous Group-pair, Organism Bipartite Network

In order to identify the conserved functional modules across a given set of organismal protein functional association networks, we need a representation that would help us enumerate these modules efficiently. The organismal protein functional association network is obtained from STRING database [4], each node is a protein and a pair of proteins are connected by an edge if there is some evidence of their functional association. Some examples of the evidences considered in STRING are gene fusion, co-occurance on the same operon, co-expression etc. In this paper we propose the *orthologous group-pair, organism bipartite network* that combines the information present in all of the individual organismal protein functional association networks into one single network (Figure 1).

As a first step to constructing this network, we need some kind of transformation that would help us understand the commonalty and differences among the networks. One such transformation is replacing all proteins in all of the organismal networks with their corresponding orthologous group IDs (Figure 1.2). The most common representation used in biology is the manually curated *Clusters of Orthologous Groups (COGs)* [5].

In the second step, we construct two sets, O and C (Figure 1.3). In C, each element is a pair (x, y), where both x and y are *COGs*. In O, each element represents an organism. These two sets become the two partites of the graph.

As a final step, we construct the orthologous group-pair, organism bipartite network (Figure 1.3), $N = (O, C, E)$. An edge $(a, b) \in E$, where $a \in O$ and $b = (u, v) \in C$ exists if and only if the COG pair (u, v) is functionally associated in organism a, i.e., in the organismal protein functional association network $A = (V(A), E(A))$ corressponding to organism a, $\exists x, y \in V(A) : x$ and y belong to orthologous cluster groups u and v, respectively, and $(x, y) \in E(A)$. Since in this paper we make use of COGs, the network N will henceforth be referred to as the ***COG-pair, organism bipartite network***.

C. Network Representation and Preprocessing

The COG-pair, organism bipartite network, N is represented using an adjacency matrix for the purpose of identifying the conserved functional modules (Figure 1.4). The organisms are the row-headers and each column header is a COG-pair. A matrix cell has a 1, if the corresponding organism (row-header) and the COG pair (column-header) are connected by an edge in network N. This matrix is typically sparse.

D. Identifying Conserved Functional Modules

1) Obtaining the Conserved COG Clusters: As a first step to identifying the modules, we identify sets of COG edges that are conserved across two or more organisms. These sets can be represented as bicliques (Figure 2.B) in the COG-pair, organism bipartite network. To avoid enumerating the same information more than once, we only enumerate the maximal bicliques (Figure 2.C).

Definition 2.1: Given a bipartite graph $N = (O, C, E)$, a subgraph $S = (O', C', E')$ of N is a biclique if $\forall a \in O'$ and $b \in C'$, $(a, b) \in E'$.

Definition 2.2: A biclique S of N is also maximal if there is no supergraph S' of S that forms a biclique in N.

The problem of identifying maximal bicliques using the binary matrix representation translates to identifying the maximal biclusters (Figure 2.B) in the matrix. Although any biclustering technique that works on binary matrices would suffice, we chose Prelic *et al.'s* Bimax biclustering algorithm

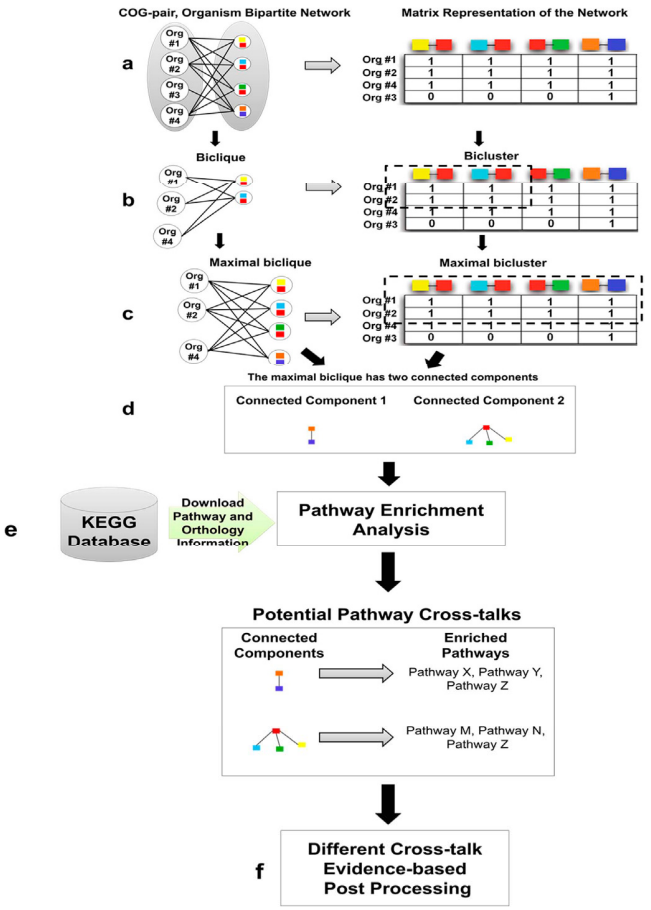

Figure 2. Method overview

[6]. There are two reasons for this choice: (1) Bimax performs on par with the best biclustering techniques [6], and (2) It has also been shown that Bimax is able to output all the optimal (maximal) biclusters in the given binary matrix [6]. The algorithm uses a divide-and-conquer approach to enumerate maximal bicliques (maximal biclusters) in the COG-pair, organism bipartite network (Figure 2.C).

2) Identifying the Connected Components in each Conserved Cluster: In each maximal biclique $S = (O^{'}, C^{'}, E^{'})$ enumerated in the previous section, $C^{'}$ represents the set of COG-COG edges conserved across the set of organisms $O^{'}$. However, we cannot consider $C^{'}$ as a functional module as is. A functional module is a connected subgraph of an organismal network as opposed to a set of edges and there is no gurantee that $C^{'}$ is connected. Hence, for each edge set $C^{'}$ of each maximal biclique $S^{'}$, we enumerate all the connected components.

3) Assessing Statistical Significance of the Functional Modules: The results of the previous section only guarantee that the subgraphs output are connected components but there is no clear indication whether the subgraphs could potentially represent functional modules or if their occurrence was purely random. One way to check this would be to compare the connectivity of each component with the connectivity that could be obtained at random for a subgraph with the same number of edges. Here we quantify connectivity using the subgraph density parameter.

The Monte Carlo method [7], [8], a robust statistical significance method, is utilized. For every connected component $S = (V, E)$, we randomly sample subsets of $|E|$ COG pairs each from the set of all possible COG pairs M and calculate the density value $\alpha(S)$ for each subset. We estimate an empirical p-value as R/W, where W is the total number of random subsets generated ($W \sim 1000$) and R is the number of random subsets that produce a test statistics $\alpha()$ greater than or equal to that of $\alpha(S)$. The conserved potential functional modules are those connected subgraphs with a p-value less than or equal to 0.05.

4) Identification of Potential Pathway Cross-talks: The statistically significant connected components are now analyzed for potential pathway cross-talks. This is done by assessing the KEGG metabolic pathway [9]–[11] enrichment, in each component. This enrichment is calculated using the hypergeometric test. If more than one pathway significantly enriches (p-value $\leq$ 0.05) a connected component, we hypothesize that the set of enriching pathways are cross-talking.

III. Experiments and Results

A. Experiments

With the goal of identifying functional modules across multiple organisms and the correlating these modules with phenotypes, we select a group of hydrogen-producing organisms, categorized by method of hydrogen production via *light fermentation* versus *dark fermentation*. The list of organisms used in each experiment can be foundat http://people.engr.ncsu.edu/kpadman/DetectingCrossTalk-Biclusters.html

B. Results

Application of the clustering algorithm resulted in identification of 8 COG clusters associated with organisms capable of light fermentation, and 28 COG clusters associated with dark fermenting organisms. Initial review of each light fermentation cluster shows the presence of a set of 13 identical COGs found across all 8 COG clusters. These "core" COGs are all observed in Cluster 1 and include genes necessary for synthesis of hydrogenase complex(es). However, for the dark fermentation clusters, we did not observe a large set of COGs present across each cluster. For this set of organisms, only two COGs were identified as present across all clusters. This may be due partially to two reasons. First, the selection of species and their diversity has some impact on the types of clusters generated. Second, dark fermentation organisms tend to utilize a greater variety

of fermentation pathways, such as acetate fermentation and butyrate fermentation pathways [12]. Greater variation in fermentation routes will not produce as large of a "core" set of COGs across each cluster.

Due to limitations in space, the only item discussed is the overall characterization of COGs present in Cluster 5 and Cluster 14 for light and dark fermentation, respectively. Clusters described in this study were selected based on whether they show greater variation in the presence of unique COGs and contained the "core" set of COGs described above. Functional associations between COG groups present in each cluster are validated through literature review and prior knowledge

1) Light Fermentation: Nitrogen-fixation is the process, in which nitrogenase catalyzes the conversion of nitrogen gas to ammonia and inadvertently results in the production of hydrogen gas as a byproduct [13], [14]. Two COGs (COG 2710 and COG 1348), which are associated with expression of two key proteins, nitrogenase iron protein (NifH) and molybdenum iron protein [13], were present across all the clusters. Although the presence of these two proteins is essential for nitrogen-fixation to be carried out by light fermenting microorganisms, expression of various genes in other metabolic pathways plays important roles in either directly or indirectly regulating the expression of genes encoding NifH proteins. These proteins include ferric iron regulation proteins (sigK, clpB, and fur-related), ammonia ligase (glnA), and nitrogenase [15]. In this study, glutamate ammonia ligase (glnA), a key gene for nitrogenase (NifH), and genes encoding proteins for iron uptake, are assembled in the same cluster. In *Anabaena*, iron uptake proteins and some nitrogen proteins (e.g., Ntc) have been shown to regulate genes encoding glutamate synthetase (glnA) [16]. Review of the role of glutamine synthetase in *Anabaena* indicates that this enzyme is responsible for regulating nitrogenase activity, thus impacting hydrogen production [16]. The indirect regulation of nitrogenase by iron uptake proteins provides an example of cross-talk between iron and nitrogen-related metabolic pathways. In addition to nitrogenase, proteins associated with the synthesis of uptake or expression of hydrogenase, were identified in 11 of the 19 COGs present in Table I. Hydrogen uptake proteins help with removing excess hydrogen to maintain the reducing environment in cells [17]. We also identified a number of proteins (e.g., Hyd and Hyp) involved in formation of [NiFe]-uptake hydrogenases. The presence of maturation hydrogenase factors (COG0068, COG0298, COG0309) and accessory protein for uptake of nickel and expression (COG0378) are consistent with literature reports describing the structure of hydrogenase complexes. Inclusion of hydrogenase proteins in Table I is likely due to the relationship of hydrogenase proteins with iron uptake genes. To function properly, iron is needed to form the NiFe center present in the large hydrogenase subunit (HupL) [18]. As such, hydrogenase maturation is dependent on cross-talks with iron uptake.

In previous studies by Lopez-Gollomon [16], the nitrogen regulator protein NtcA was found to work together with the iron-uptake protein, Fur, to co-regulate genes involved in various metabolic functions. Metabolic functions co-regulated include transcriptional regulation protein and glutamine synthesis [19]. In this study, genes encoding iron uptake regulator proteins (COG 0735) were clustered together with genes encoding glutamine synthetase (COG0174). The co-appearance of these two COGs suggests cross-talk between iron uptake and ammonia assimilation networks may be occurring. In addition, there is indication that hydrogenase proteins, such as HupUV, are involved in regulating the glutamine synthetase gene, glnAII, in some organisms [13], [20].

2) Dark Fermentation: An example of COG clusters identified in dark fermentative bacteria is present in Table II. In this cluster, 13 different COGs consisting of proteins that are either directly or indirectly responsible for the uptake or production of hydrogen, were present. Of these COGs, 7 are related to the synthesis or expression of [NiFe]-hydrogenase, an enzyme that catalyses the reversible oxidation of molecular hydrogen, and plays a vital role in anaerobic metabolism [20]; the others are involved in nitrogen and iron metabolic pathways that include proteins like nitrogenase, iron uptake proteins, such as Fur (COG0735), ammonia assimilation proteins, such as glutamine synthetase (COG3968), and proteins involved in electron transfer. Previous findings by Butland *et al.* [21] show that the presence of proteins (e.g., HypE, HypD, HupS, HupD) is typically associated with hydrogen uptake [18], [22]. Based on the other genes (e.g., hybG, hupS) present in the cluster, we can predict that [NiFe]-hydrogenase is associated with hydrogen uptake in this group of organisms.

In addition to hydrogenase maturation and expression proteins, Fe-S oxidoreductases were identified. As part of the structure of [NiFe]-hydrogenase, Fe-S metal centers are located on the small subunit of the hydrogenase complex [18], [20]. Because of this, it is expected that iron uptake pathway would cross-talk with hydrogenase-related pathways. Furthermore, iron uptake pathway also cross-talks with nitrogen metabolism in the sense that iron uptake proteins can be involved indirectly in nitrogen metabolism through regulation of nitrogenase and maintaining the reducing environment in the cell through hydrogen uptake (hydrogenase) [19], [23].

It has been shown that cross-talk between iron uptake and nitrogen metabolism enables regulation of ammonia assimilation [14]; it may be possible that the uncharacterized glutamine synthetase protein in Table II is subject to such kind of regulation. Another observation is that, in our result, the gene encoding the uncharacterized glutamine synthetase proteins was only present in a few species, includ-

Table I

CLUSTER 5: THE PRESENCE (1) OR ABSENCE (0) OF COGS FOR LIGHT FERMENTATION RESULTS IN ONE CLUSTER IDENTIFIED BY THE BI-CLUSTERING ALGORITHM. ORGANISMS: *Anabaena variabilis* (AVA), *Anabaena (Nostoc) sp. PCC 7120 (ana)*, *Rhodobacter sphaeroides (rsk)*, *Rhodpseudomonas palustris (rpa)*, AND *Rhodospirillum rubrum (rru)*.

COG ID	COG Description	ava	ana	rsk	rpa	rru
COG0068	Hydrogenase maturation factor	1	1	1	1	1
COG0298	Hydrogenase maturation factor	1	1	1	0	1
COG0309	Hydrogenase maturation factor	1	1	1	1	1
COG0374	Ni,Fe-hydrogenase I large subunit	1	1	1	1	1
COG0375	Zn finger protein HypA/HybF (possibly regulating hydrogenase expression)	1	1	1	1	1
COG0378	Ni2+-binding GTPase involved in regulation of expression and maturation of urease and hydrogenase1	1	1	0	1	1
COG0409	Hydrogenase maturation factor	1	1	1	1	1
COG0680	Ni,Fe-hydrogenase maturation factor	1	1	1	1	1
COG1740	Ni,Fe-hydrogenase I small subunit	1	1	1	1	1
COG0174	Glutamine synthetase	1	1	1	1	1
COG0535	Predicted Fe-S oxidoreductases	1	1	1	1	1
COG0716	Flavodoxins	1	1	0	1	1
COG1348	Nitrogenase subunit NifH (ATPase)	1	1	1	1	1
COG2082	Precorrin isomerase	1	1	1	1	1
COG2710	Nitrogenase molybdenum-iron protein, alpha and beta chains	1	1	1	1	1
COG2370	Hydrogenase/urease accessory protein	1	1	0	0	0
COG1941	Coenzyme F420-reducing hydrogenase, gamma subunit	1	1	0	0	0
COG3259	Coenzyme F420-reducing hydrogenase, alpha subunit	1	1	0	0	0
COG0735	Fe2+Zn2+ uptake regulation proteins	1	1	1	1	1

Table II

CLUSTER 14: THE PRESENCE (1) OR ABSENCE (0) OF COGS FOR DARK FERMENTATION RESULTS IN ONE CLUSTER IDENTIFIED BY THE BI-CLUSTERING ALGORITHM. ORGANISMS: *Bacillus licheniformis* (BLI), *Clostridium acetobutylicum* (CAC), *Clostridium beijerinckii* (CBE), *Clostridium perfringens* (CPF), *Caldicellulosiruptor saccharolytics* (CSC), *Clostridium thermocellum* (CTH), *Escherichia coli* (ECO), AND *Desulfovibrio vulgaris subsp. vulgaris Hildenborough* (DVU).

COG ID	COG Description	bli	cac	cbe	cpf	csc	cth	dvu	eco
COG0298	Hydrogenase maturation factor	0	1	1	0	1	1	0	1
COG0309	Hydrogenase maturation factor	0	1	1	0	1	1	1	0
COG0374	Ni,Fe-hydrogenase I large subunit	0	1	1	0	0	0	1	1
COG0409	Hydrogenase maturation factor	0	1	1	0	1	1	1	0
COG0680	Ni,Fe-hydrogenase maturation factor	0	1	1	0	0	0	1	1
COG1740	Ni,Fe-hydrogenase I small subunit	0	1	1	0	0	0	1	1
COG0535	Predicted Fe-S oxidoreductases	1	1	1	0	1	1	1	0
COG1348	Nitrogenase subunit NifH (ATPase)	0	1	1	0	1	1	1	0
COG2710	Nitrogenase molybdenum-iron protein, alpha and beta chains	0	1	1	0	1	1	1	0
COG0716	Flavodoxins	1	1	1	1	1	1	1	1
COG0735	Fe2+/Zn2+ uptake regulation proteins	1	1	1	1	1	1	1	1
COG2082	Precorrin isomerase	0	1	1	1	0	0	1	0
COG3968	Uncharacterized protein related to glutamine synthetase	0	1	1	0	0	1	1	0

ing *Clostridium acetobutylicum* and *Clostridium beijerinckii*, which both contained nitrogenase and hydrogenase enzymes. It has been demonstrated that, in light fermenting organisms, such as *R. palustris*, glutamine synthetase is regulated by hydrogenase accessory proteins (HupUV) [14]. However, to the best of our knowledge, this relationship has not been described in dark fermentation organisms. This knowledge increases the probability that the uncharacterized glutamine synthetase protein maybe present in the COG cluster owning to its association with nitrogenase proteins, which may further indicate possible cross-talk between ammonia assimilation and nitrogen metabolism.

IV. RELATED WORK

To the best of our knowledge, the method proposed in the paper is the first to provide a solution to the problem of identifying phenotype-related potential pathway cross-talk mechanisms by comparing dozens of phenotype-expressing organismal networks. However, there are other methods in data mining that could be modified to solve this problem.The *frequent subgraph mining* is a method of detecting subgraphs that occur frequently, i.e, the building blocks of a given graph ([24]–[26]). *Network alignment* is a network comparative analysis method that can identify the common subgraphs across all the input organismal networks [27]–[32]. *Clustering* is an approach where vertices of a graph

are grouped based on some similarity measure [33], [34]. The clusters from each input organismal network can be identified and these cluster sets can be compared to identify the conserved ones. The methods described in this section typically compare only a small number of graphs.

V. Conclusion

We have developed a method to identify the conserved functional modules across multiple phenotype-related organisms. This method could allow researchers to detect cross-talk between metabolic pathways by analyzing conserved modules generated. Validation through hypothesis testing proves that the clusters obtained are not random coincidence. To make sure that the result produced by our method is biologically relevant, we further applied the method to a group of hydrogen-producing organisms. Known cross-talks between pathways involved in hydrogen production were identified, predictions of cross-talks among candidate pathways were made.

Acknowledgment

This work was supported in part by the U.S. Department of Energy, Office of Science, the Office of Advanced Scientific Computing Research (ASCR) and the Office of Biological and Environmental Research (BER) and the U.S. National Science Foundation (Expeditions in Computing). The work by A.M.R. was supported by the Delores Auzenne Fellowship and the Alfred P. Sloan Minority PhD Scholarship Program. Oak Ridge National Laboratory is managed by UT-Battelle for the LLC U.S. D.O.E. under contract no. DEAC05-00OR22725.

References

[1] J. A. Hernandez, S. Lopez-Gomollon, A. Muro-Pastor, A. Valladares, M. T. Bes, M. L. Peleato, and M. F. Fillat, "Interaction of FurA from *Anabaena sp.* PCC 7120 with DNA: A reducing environment and the presence of Mn(2+) are positive effectors in the binding to isiB and furA promoters," *Biometals*, vol. 19, pp. 259–268, 2006.

[2] S. Lopez-Gomollon, J. A. Hernandez, S. Pellicer, V. E. Angarica, M. L. Peleato, and M. F. Fillat, "Cross-talk between iron and nitrogen regulatory networks in *Anabaena (Nostoc)* sp. PCC 7120: Identification of overlapping genes in FurA and NtcA regulons," *J. Mol. Biol.*, vol. 374, pp. 267–281, 2007.

[3] Y. Cheng, J. H. Li, L. Shi, L. Wang, A. Latifi, and C. C. Zhang, "A pair of iron-responsive genes encoding protein kinases with a Ser/Thr kinase domain and a His kinase domain are regulated by NtcA in the Cyanobacterium *Anabaena sp.* strain PCC 7120," *J. Bacteriol.*, vol. 188, pp. 4822–4829, 2006.

[4] L. J. Jensen, M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, T. Doerks, P. Julien, A. Roth, M. Simonovic, P. Bork, and C. von Mering, "String 8–a global view on proteins and their functional interactions in 630 organisms." *Nucleic Acids Res*, vol. 37, pp. D412–416, 2009.

[5] R. L. Tatusov, M. Y. Galperin, D. A. Natale, and E. V. Koonin, "The COG database: A tool for genome-scale analysis of protein functions and evolution," *Nucleic Acids Research*, vol. 28, no. 1, pp. 33–36, 2000.

[6] A. Prelić, S. Bleuler, P. Zimmermann, A. Wille, P. Bühlmann, W. Gruissem, L. Hennig, L. Thiele, and E. Zitzler, "A systematic comparison and evaluation of biclustering methods for gene expression data," *Bioinformatics*, vol. 22, pp. 1122–1129, 2006.

[7] B. North, D. Curtis, and P. Sham, "A note on the calculation of empirical *p*-values from Monte Carlo procedures," *Am. J. Hum. Genet*, vol. 71, pp. 439–441, 2002.

[8] B. Zhang, B. Park, T. Karpinets, and N. Samatova, "From pull-down data to protein interaction networks and complexes with biological relevance," *Bioinformatics*, vol. 24(7), pp. 979–986, 2008.

[9] M. Kanehisa and S. Goto, "KEGG: Kyoto encyclopedia of genes and genomes," *Nucleic Acids Res*, vol. 28, pp. 27–30, 2000.

[10] M. Kanehisa, S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Hirakawa, "From genomics to chemical genomics: New developments in KEGG," *Nucleic Acids Res*, vol. 34, no. suppl 1, pp. D354 –D357, 2006.

[11] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa, "KEGG for representation and analysis of molecular networks involving diseases and drugs," *Nucleic Acids Res*, vol. 38, pp. D355 –D360, 2010.

[12] D. White, *The physiology and biochemistry of prokaryotes*. Oxford, England: Oxford University Press, Inc., 2000.

[13] F. E. Rey, E. K. Heiniger, and C. S. Harwood, "Redirection of metabolism for biological hydrogen production," *Appl Environ Microbiol*, vol. 73, pp. 1665–1671, 2007.

[14] F. E. Rey, Y. Oda, and C. S. Harwood, "Regulation of uptake hydrogenase and effects of hydrogen utilization on gene expression in *rhodopseudomonas palustris*," *J. Bacteriol.*, vol. 188, pp. 6143–6152, 2006.

[15] W. C. Lathe, J. M. Williams, M. E. Mangan, and D. Karolchik, "Genomic data resources: Challenges and promises," *Nature Edu.*, vol. 1, no. 3, 2008.

[16] S. Lopez-Gomolln, J. A. Hernndez, S. Pellicer, V. E. Angarica, M. L. Peleato, and M. F. Fillat, "Cross-talk between iron and nitrogen regulatory networks in *anabaena (nostoc)* sp. pcc 7120: Identification of overlapping genes in FurA and NtcA regulons," *J. Mol. Biol.*, vol. 374, pp. 267–281, 2007.

[17] K. L. Kovacs, Z. Bagi, B. Balint, B. d. Fodor, G. Scnadi, R. Csaki, T. Hanczar, A. T. Kovacs, G. Maroti, K. Perei, A. Toth, and G. Rakhely, "Novel approaches to exploit microbial hydrogen metabolism," in *Biohydrogen III*, J. Miyake, Y. Igarashi, and M. Rogner, Eds. USA: Elsevier Inc, 2004.

[18] P. M. Vignais, B. Billoud, and J. Meyer, "Classification and phylogeny of hydrogenases," *FEMS Microbiol Rev*, vol. 25, pp. 455–501, 2001.

[19] J. Miyake, *Biohydrogen*. New York, USA: Plenum Press, 1998.

[20] S. Shima and R. K. Thauer, "A third type of hydrogenase catalyzing H2 activation," *Chem. Rec.*, vol. 7, pp. 37–46, 2007.

[21] G. Butland, J. w. Zhang, W. Yang, A. Sheung, P. Wong, J. F. Greenbalt, A. Emili, and D. B. Zamble, "Interactions of the *escherichia coli* hydrogenase biosynthetic proteins: HybG complex formation," *FEBS Lett*, vol. 580, pp. 677–681, 2006.

[22] O. Guerrini, P. Soucaille, L. Girbal, B. Guigliarelli, C. Leger, B. Burlat, and C. Lger, "Characterization of two 2[4Fe4S] ferredoxins from *clostridium acetobutylicum*," *Curr Microbiol*, vol. 56, pp. 261–267, 2008.

[23] J. Yu and P. Takahashi, "Biophotolysis-based hydrogen production by Cyanobacteria and green microalgae," in *Communicating Current Research and Educational Topics and Trends in Applied Microbiology*, A. Mendez-Vilas, Ed., 2007, vol. 1, pp. 79–89.

[24] M. Koyutürk, A. Grama, and W. Szpankowski, "An efficient algorithm for detecting frequent subgraphs in biological networks," *Bioinformatics*, vol. 20, pp. 200–207, 2004.

[25] A. Inokuchi, T. Washio, and H. Motoda, "An apriori-based algorithm for mining frequent substructures from graph data," in *PKDD*, 2000, pp. 13–23.

[26] J. Huan, W. Wang, and J. Prins, "Efficient mining of frequent subgraphs in the presence of isomorphism," in *ICDM*, 2003, pp. 549–552.

[27] M. Koyuturk, Y. Kim, U. Topkara, S. Subramaniam, W. Szpankowski, and A. Grama, "Pairwise alignment of protein interaction networks," *J Comput Biol*, vol. 13, pp. 182–199, 2006.

[28] M. Narayanan and R. M. Karp, "Comparing protein interaction networks via a graph match-and-split algorithm," *J Comput Biol*, vol. 14, pp. 892–907, 2007.

[29] J. Flannick, A. F. Novak, C. B. Do, B. S. Srinivasan, and S. Batzoglou, "Automatic parameter learning for multiple network alignment," in *RECOMB*, 2008, pp. 214–231.

[30] S. Erten, X. Li, G. Bebek, J. Li, and M. Koyuturk, "Phylogenetic analysis of modularity in protein interaction networks," *BMC Bioinformatics*, vol. 10, pp. 333–346, 2009.

[31] R. Singh, J. Xu, and B. Berger, "Global alignment of multiple protein interaction networks with application to functional orthology detection," *PNAS*, vol. 105, pp. 12 763–12 768, 2008.

[32] R. Singh, X. Jinbo, , and B. Berger, "Pairwise global alignment of protein interaction networks by matching neighborhood topology," in *RECOMB*, 2007, pp. 16–31.

[33] W. Hwang, Y.-R. Cho, A. Zhang, and M. Ramanathan, "A novel functional module detection algorithm for protein-protein interaction networks," *Algorithms Mol Biol*, vol. 1, 2006.

[34] M. Koyuturk, W. Szpankowski, and A. Grama, "Biclustering gene-feature matrices for statistically significant dense patterns," in *CSB*. Washington, DC, USA: IEEE Computer Society, 2004, pp. 480–484.