

Algorithm for the Construction of a Global Enzymatic Network to be Used for Gene Network Reconstruction

Andrés Quintero¹, Jorge Ramírez², Luis Guillermo Leal³ and Liliana López-Kleine^{4,*}

¹Universidad Nacional de Colombia, Department of Biology, Master Student, Universidad Nacional de Colombia-Sede Bogotá; ²Universidad Nacional de Colombia, School of Mathematics, Associate Professor, Universidad Nacional de Colombia-Sede Medellín; ³Universidad Nacional de Colombia, Department of Statistics, Master Student, Universidad Nacional de Colombia-Sede Bogotá; ⁴Universidad Nacional de Colombia, Department of Statistics, Associate Professor, Universidad Nacional de Colombia-Sede Bogotá, Colombia

Abstract: Relationships between genes are best represented using networks constructed from information of different types, with metabolic information being the most valuable and widely used for genetic network reconstruction. Other types of information are usually also available, and it would be desirable to systematically include them in algorithms for network reconstruction. Here, we present an algorithm to construct a global metabolic network that uses all available enzymatic and metabolic information about the organism. We construct a global enzymatic network (GEN) with a total of 4226 nodes (EC numbers) and 42723 edges representing all known metabolic reactions. As an example we use microarray data for *Arabidopsis thaliana* and combine it with the metabolic network constructing a final gene interaction network for this organism with 8212 nodes (genes) and 4606,901 edges. All scripts are available to be used for any organism for which genomic data is available.

Received on: May 19, 2014- Revised on: July 28, 2014- Accepted on: August 05, 2014

Keywords: EC number, Gene network reconstruction, Global enzymatic network, KEGG, Perl.

INTRODUCTION

One of the major goals in systems biology is to understand how functional relationships between genes under specific conditions determine changes in the organism's behavior and cell physiology. Co-expression networks and genome-scale metabolic models, have been successfully applied to advance this kind of biological knowledge [1-4].

There are many strategies which allow the construction of such co-expression networks [1-3], however, only a few of those are designed to integrate more than one type of genomic data [4-8], namely: gene expression, sequence homology, cell location, vicinity of chromosomal genes, sites binding to transcription factors, fusion events and phylogenetic profiles. Although all these types of data provide information for the construction of networks, the most valuable type are metabolic networks, since they directly give information on the relationship between cellular entities (enzymes) and metabolites. Moreover, metabolic reactions are global for all organisms and therefore a global metabolic network containing all known metabolic reactions is of special interest for biological network reconstruction of any kind.

One technique that has been implemented for the reconstruction of metabolic networks, and one that integrates various types of genomic data is Kernel Canonical Correlation Analysis (KCCA) [6, 8, 9]. One major difficulty with KCCA

occurs, however, in the necessary tuning of arbitrary parameters (for example thresholds) for the network construction. Methodologies have also been developed for integrating genomic data types into partially known networks [10], generating networks of lesser genomic coverage. Similarly, other proposed methodologies are able to add similarity results on networks established previously as gold standards [9]. Other techniques rely on probabilistic approaches such as Boolean or Bayesian networks [11] or fully probabilistic descriptions [4], which however, have the drawback of being computationally non-practical for databases with a large number of genes. New and effective methodologies that integrate different types of genomic data are therefore needed.

Of particular interest would be novel approaches that do not depend on arbitrary parameters and that are applicable to a large number of organisms. Such flexible and relatively simple methods will help advance biological knowledge of both model and less studied organisms through the generation of functional gene predictions leading to the formulation of new biological hypotheses.

In this work we propose a general methodology for constructing gene networks using information on metabolic reactions and gene expression data. Our strategy follows the following basic steps: *i*) construction of the global enzymatic network (GEN); *ii*) construction of the organismic enzymatic network (OEN) using a gene similarity matrix based on expression data of an organism of interest; *iii*) integration of both networks in order to obtain a final weighted organismic enzymatic network (WOEN). The implementation of the methodology is available under request as Perl scripts.

*Address correspondence to this author at the Universidad Nacional de Colombia, Department of Statistics, Associate Professor, Universidad Nacional de Colombia-Sede Bogotá, Colombia; Tel: 57 3165000, 13157; E-mail: llopezk@unal.edu.co

To illustrate the proposed methodology, we consider the model organism *Arabidopsis thaliana* and construct WOEN using microarray data and the GEN constructed separately. The used GEN has a total of 4226 nodes (EC numbers) and 42723 edges, and the final WOEN of *Arabidopsis thaliana* has 8212 nodes (genes) and 4606,901 edges, the increase in the number of nodes between the GEN and the WOEN, is explained by the fact that in an organism more than one enzyme can have the same enzymatic activity. The GEN we construct here can be used in combination with gene expression data from the same or any other organism in order to construct the associated gene interaction networks. In that sense, we are using the GEN as a starting point in order to obtain WOENs representing gene coregulations that will depend on the microarray data used. Moreover, any other genomic data type that is also representable as a similarity matrix can be included in the construction by combination with the GEN and could therefore enrich the obtained WOEN.

MATERIALS AND METHODS

Genomic Databases

The Genome Expression Omnibus at NCBI (<http://www.ncbi.nlm.nih.gov/geo/>) was queried for microarray datasets of the model plant *Arabidopsis thaliana*. A total of sixteen experiments from *Arabidopsis* in response to several pathogens were used (accession numbers: GSE28800, GSE26973, GSE5513, GSE28800, GSE5752, GSE5513, GSE8319, GSE5752, GSE12856, GSE8319, GSE13739, GSE12856, GSE14961, GSE13739, GSE15236, GSE14961, GSE16472, GSE15236, GSE16497, GSE16472, GSE17382, GSE16497, GSE17875, GSE17382, GSE19273, GSE17875, GSE20188, GSE19273, GSE21920 and GSE20188).

In addition to the gene expression data, information on all known metabolic reactions was obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG). The KEGG API was used to download the list of up to date EC numbers, enzymes related to each EC number and enzyme sequence. Perl scripts were developed to connect to the REST based KEGG API service, and download all metabolic information required for the global metabolic network; this ensures that any future users of the scripts will use up to date information.

Preprocessing of Microarray Data and Gene-gene Similarity Matrix

The downloaded datasets were independently pre-processed for noise reduction, quantile normalization and log₂ transformation. The RMA (Robust Multiarray Average) method was applied to Affymetrix data using R affy library [12, 13]. Probe IDs were converted to gene IDs. Single probes that matched more than one gene were removed [14]. For those multiple probes that matched a single gene, the maximum expression among the multiple probes was assigned to the gene as suggested in Dozmorov [15]. Genes common to all databases were used for construction of the gene-gene similarity matrix. Microarrays expression data was used to assess the gene expression similarity matrix (GESM). The similarity in GESM for gene pairs was obtained using the mutual information coefficient [16].

Global Enzymatic Network Construction

Using the names, codes and reactions of biochemical activities performed by enzymes, which are defined by the Enzyme Commission (EC) [17], a global enzymatic network (GEN) was constructed, in which the nodes are enzymatic activities (EC numbers), and two activities are connected by an edge if they share at least one metabolite, either as substrate or product [18].

For the construction of the GEN the following strategy was applied, for which several Perl scripts (Table 1) were developed (see Fig. 1, GEN construction):

1. Enzymatic Reactions File Construction

A list of all currently known EC numbers was retrieved from the KEGG API, and for each one of the EC numbers, all metabolites involved in the reactions catalyzed by the enzymes annotated with that EC number were downloaded, so that a file with EC numbers and associated metabolites was created. The Perl script used to accomplish this was called "1_Fetch_EC_met.pl" and does not need any input; it connects directly to the KEGG database and retrieves all information needed.

Scripts "1_metabolite_name-code_hash.pl" and "2_reactions_metID.pl" were created for construction of GEN from ftp KEGG data: *i*) "1_metabolite_name-code_hash.pl" uses as input one file with the identification code, name and alternative names of all biological compounds to build a list of equivalences ("dictionary") relating names and alternate names of metabolites with an identification code (used in KEGG). The goal of this step is to create the metabolic network with a less complex and easy to understand nomenclature for future users; *ii*) using the equivalence list, the metabolic reactions were converted into reactions with codified metabolites. Furthermore, as each EC number can have many reaction variants in different organisms, we summarized this information relating each metabolite to one EC number using the main reaction as handle. This step is achieved with the script "2_reactions_metID.pl", that uses information of each known enzymatic activity (names of the activity, reaction and involved metabolites) and the list containing information on names and codes of metabolites.

2. Filtering of Metabolites

As some metabolites, for example H₂O or NADH, are very common and used in many enzymatic reactions, they have to be withdrawn to avoid an overconnected network. Following Kharchenko [18], the forty most common metabolites were filtered (Supplementary Table 1). This filtering was done with the perl script "3_filter_reactions_slim.pl", which allows filtering the first *n* most common metabolites.

3. Comparison of Enzymatic Activities and Construction of the GEN

Once the *n*(=40) most common metabolites were filtered, the GEN was constructed, connecting two nodes (EC numbers) if they share at least one metabolite. This step is achieved using the script "4_network_construction.pl".

Table 1. Description of all the scripts used for the GEN, OEN and WOEN reconstruction. [^]scripts written in italics are for users that want to use KEGG ftp database downloaded data; *input and output files names written in quotations marks for easy differentiation with running text, names used can change depending of the user preferences.

Script Abbreviation	Script [^]	Description	Input Files*	Output Files*
GEN Construction				
1.2SGEN	1-2_Fetch_EC_met.pl	Downloads EC numbers and reactions data and prints a simplified reaction	NA	"reactions_slim"
-	<i>1_metabolite_name-code_hash.pl</i>	Creates a equivalence list of complete names and codes for metabolites	Compound information file from KEGG ftp	"metabolite_name_code"
-	<i>2_reactions_metID.pl</i>	Prints a simplified reaction	"metabolite_name_code"	"reactions_slim"
3SGEN	3_filter_reactions_slim.pl	Filter n most common metabolites	"reactions_slim"	"reactions_slim.filter_n"
4SGEN	4_network_constuction.pl	Prints the GEN as a list of node pairs	"reactions_slim.filter_n"	"network.reactions_slim.filter_n"; "eclist"
OEN Construction				
1SOEN	1_Fetch_genes.pl	Downloads all GSE and BLAST them against genome of interest	"eclist"	"Blasted_genes.list"
2SOEN	2_DataBase_network.pl	Assigns edges within genes comparing associated enzymatic activities in GEN	"Blasted_genes.list"; "network.reactions_slim.filter_n"	"Gen-Gen.list"
-	<i>3_Gen-gen_adjacency_matrix.pl</i>	Prints the OEN as an adjacency matrix	"Gen-Gen.list"	"Gen-Gen_adjacency.matrix"
WOEN Construction				
1SWOEN	4_Weight_adjacency_matrix.pl	Weights the OEN using the GESM	"Gen-Gen_adjacency.matrix"; a GESM valid file	"Gen-Gen_similarity.matrix"

Organismic Enzymatic Network Construction

In the organismic enzymatic network (OEN) nodes are genes from the genome of the organism of interest, and two genes are connected by an edge if the EC numbers associated to each one of the genes share a metabolite. The following strategy was designed to construct the OEN (Fig. 1, OEN construction):

1. Match of Enzymatic Activities to Genes in the Genome of Interest

The first step, is to assess sequence homology between all known enzymes assigned to each one of the EC numbers (Gold Standard Enzymes (GSE)) [18] present in the GEN and all protein-coding genes in the genome of the organism of interest (Fig. 1). This step is achieved with the Perl script "1_Fetch_genes.pl", which downloads from KEGG amino-acid (aa) sequences of all known GSE. Then, this script performs a BLASTP homology search comparing these aa sequences and all known protein-coding genes of the organism of interest. The E-value threshold can be chosen, and for this implementation we used 1×10^{-5} . Because this search is computationally expensive, the script was developed to run the

BLASTP for subsets of EC numbers, providing therefore the option of parallel running over different subsets, and finally combine the results into a single file. The result is a file with a list of protein coding genes and associated EC numbers.

2. Construction of the OEN

Once the list of enzymes of the organism of interest has been obtained and related to the corresponding EC numbers of the GEN, the OEN can be constructed (Fig. 1). The Perl script "2_DataBase_network.pl" does this by searching in the GEN, and linking up genes to define the edges in the OEN, if the associated EC numbers are linked by an edge in the GEN.

3. Representation of the OEN as an Adjacency Matrix

Finally, the script "3_Gen-gen_adjacency_matrix.pl" represents the OEN in terms of an adjacency matrix. This matrix is a matrix containing 1 if an edge exists between enzymes coded by the genes of the organism and 0 elsewhere. All protein-coding genes, for which a metabolic activity could not be associated by the BLAST search, are not included in the matrix.

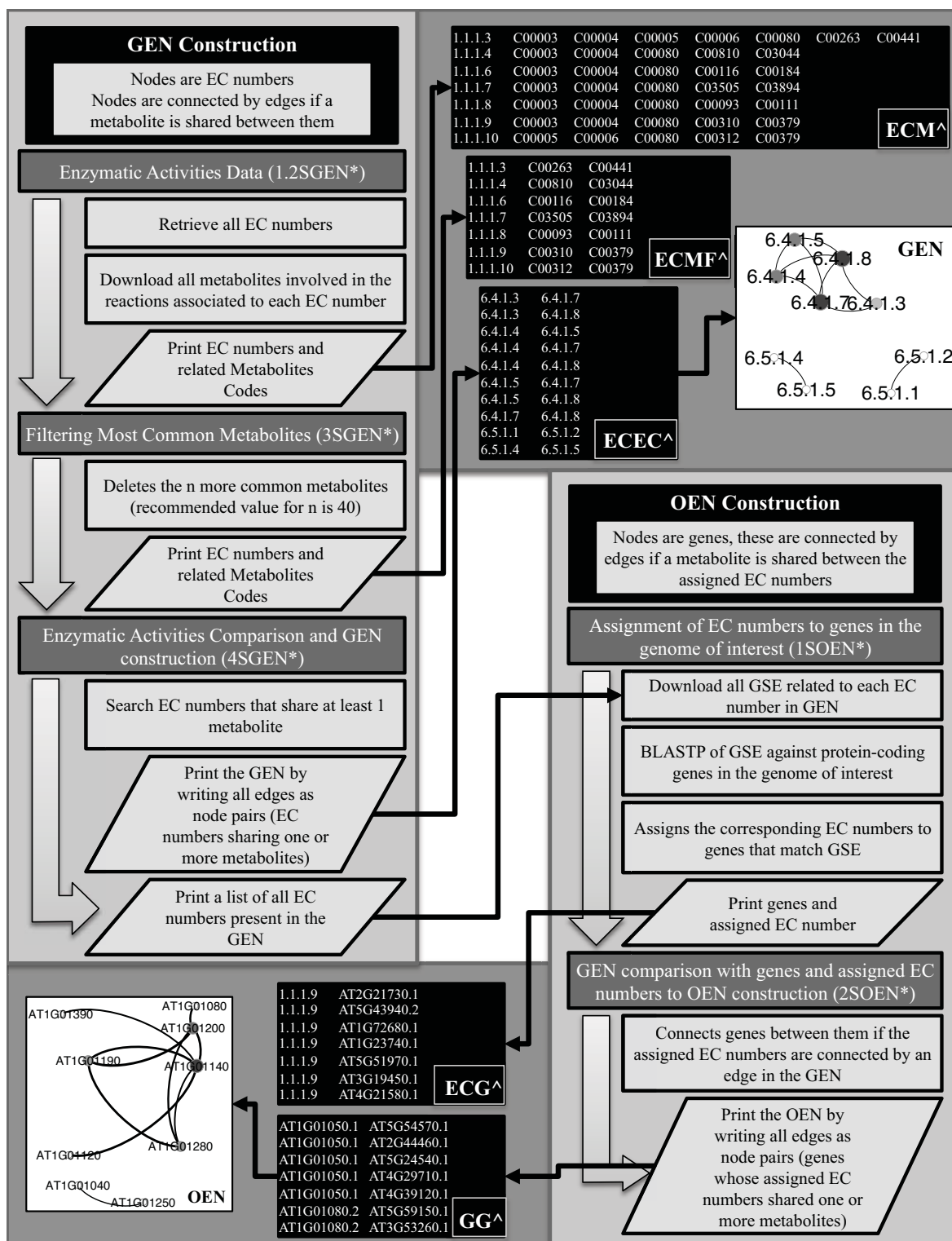


Fig. (1). Workflow of the procedure for the GEN and OEN construction. * 1.2SGEN, 3SGEN, 4SGEN, 1SOEN and 2SOEN represent the corresponding abbreviation for the scripts in (Table 1). Black boxes marked with ^ represent examples of the files generated by the related scripts; ECM: first column are EC numbers and the other columns are codified metabolites associated to each EC number; ECMF: ECM after applying the filter for the most common metabolites; ECEC: each line is a pair of EC numbers representing an edge in the GEN; ECG: first column are EC numbers and second column are the genes assigned to each EC number; GG: each line is a pair of genes representing an edge in the OEN.

Weighted Organismic Enzymatic Network Construction

At this point a final network (weighted organismic enzymatic network (WOEN)) is obtained combining the OEN

and the GESM. This WOEN is represented as an adjacency matrix (Fig. 2, WOEN construction). Edges could be weighted using the script “4_Weight_adjacency_matrix.pl”.

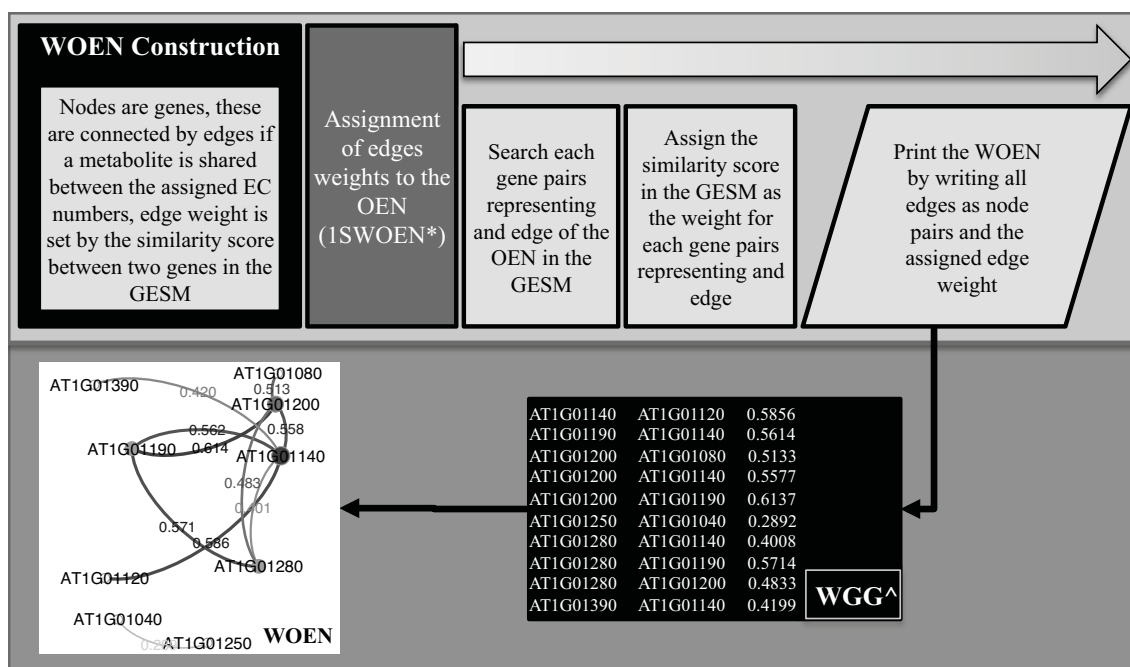


Fig. (2). Workflow of the procedure for the WOEN construction. * 1SWOEN represents the corresponding abbreviation for the script in (Table 1). Black box marked with ^ represent an example of the files generated by the related script; WGG: each line is a pair of genes representing an edge with the assigned weight in the WOEN.

Topological Analysis of Networks

Different topological properties of the obtained network were computed in order to reveal changes in the network configuration at each step of our methodology. We consider the networks: GEN, OEN and WOEN; and the following topological properties: number of nodes, number of edges, clustering coefficient, average path length and centralization [19]. This analysis was performed using the R library igraph [20].

Validation of Edges Linking Immunity Related Genes (IRGs)

The WOEN was mined to validate some functional relationships between immunity related genes (IRGs). Given their importance on immune processes, four IRGs were selected (FLS2, CLV1, RPS2 and RPS4), their edges were filtered and compared with interactions previously reported in literature.

RESULTS AND DISCUSSION

GESM for *Arabidopsis thaliana* Microarray Data

After calculating the similarity measurements between all gene pairs, we obtained a GESM of 8,212x8,212 genes for the microarray data of *Arabidopsis thaliana*. This similarity represents the amount of coordinated activity between all pairs of genes and contains all known genes of the organisms, regardless whether their products participate or not in enzymatic activities. Similarities between genes on this matrix strongly depend on the gene expression data used.

GEN Construction

Using all enzymatic activities data known to date, a GEN (Table 2) was constructed (representing 4,226 enzymatic

activities, with a final number of 2'043,335 associated GSE in a GEN constructed after filtering for the 40 most common metabolites). The GEN was drawn using Gephi [21] (Fig. 3). This network is available at, as a list of connected pairs of enzymatic activities.

OEN and WOEN Construction for *Arabidopsis thaliana*

Using the proposed methodology a OEN and subsequently a WOEN (Table 2) was constructed for *Arabidopsis thaliana* in order to illustrate the methodology. These networks do not aim to represent overall genetic networks, although if very general gene expression data is used (representing many different conditions), a more general network could be achieved. Nevertheless, a detailed analysis of the network allowed us to retrieve several interesting relationships between genes previously reported in the literature (see section on validation).

Topological Analysis of Networks

One way of identifying the effect of the different steps of the proposed algorithm is to track the changes in topological properties of the obtained networks (Table 2). It was found, for example, that GEN is a small graph compared to the subsequent constructed OEN and WOEN. After the BLASTP homology search was performed on GEN, a huge amount of hits per enzyme was revealed. Consequently, the number of edges increased drastically when passing to OEN.

Due to the high number of edges, we expected the OEN to have a higher clustering coefficient; however, during the BLASTP step, the number of nodes augmented simultaneously. As a consequence, both networks exhibit approximately the same degree of clustering. This result supposes that modules of highly connected nodes can be observed

indistinctly at the level of enzymes (GEN) or genes (OEN and WOEN).

Table 2. Topological variables measured in the global networks.

Variable	GEN	OEN	WOEN
Nodes	4,226	9,829	8,212
Edges	42,753	6,444,453	4,606,901
Clustering coefficient	0.52	0.48	0.49
Average path length	4.22	2.03	2.02
Centralization	0.05	0.01	0.01

Same clustering coefficients do not mean equal connectivity in the process represented. To evaluate how edges improve the graph global connectivity, we calculated the centralization and average path length for each network (Table 2). The higher number of nodes in OEN generated a better connectivity as each pair of nodes is separated by 2 edges contrasting the 4 edges in GEN. The better connectivity found in OEN and WOEN also means that graphs are neither centralized nor hubs-dependent. Besides, in metabolic networks a low average path length is an indicative of more efficient transfer processes [22].

The weighting step reduced the number of nodes and edges in the network. Despite this size reduction, the value

of the topological properties of WOEN was about the same. We conclude that genes without expression data are not relevant for the system representation. However, these genes could not be identified using just pathways data. It must be pointed out that expression data allowed us to weight the functional relationships, but also to drop irrelevant nodes from the final WOEN. Finally, our topological analysis results are comparable to those from other methodologies [23].

Validation of Edges Linking Immunity Related Genes (IRGs) Based on Previous Works

Four of the most important IRGs were searched on the WOEN and their edges were compared with literature (Table 3). One of them, *FLS2*, was found linked to *BAK1* and *BRI1*. The protein *FLS2* is a LRR receptor-like serine/threonine-protein kinase that recognizes peptide from flagellin and triggers plant immunity [24]. On the other hand, *BAK1* can regulate the tradeoff between immunity and responses to hormones. While *BRI1* is a receptor of the growth hormone-brassinosteroid [25]. *BAK1* is not only a co-receptor of *FLS2* but also interacts with *BAK1* as reported previously [24, 26, 27].

Some *FLS2* edges are also verified by the work of Qi and Tsuda [28]. They propose that *FLS2* forms a PTI (MAMP-triggered immunity) signaling complex with *RPM1*, *RPS2* and *RPS5* [28]. Lu, Lin, Gao, Wu, Cheng and Avila [29] indicate that *PUB12* and *PUB13* promote flagellin-induced *FLS2* degradation. Besides, the protein *PBL1* interacts with *FLS2* and they are rapidly phosphorylated upon *FLS2* activation by its ligand *flg22* [27]. Finally, Mersmann, Bourdais,

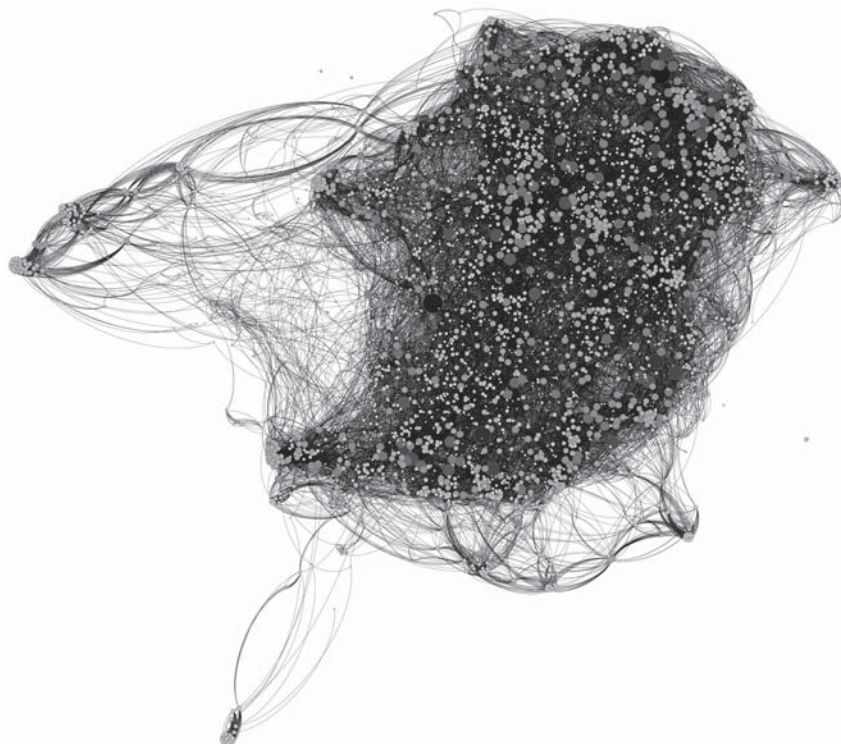


Fig. (3). Global metabolic network constructed with the proposed algorithm. Node size and color intensity represent the connectivity of the node. This network has total of 4226 nodes (EC numbers) and 42723 edges and shows a typical scale-free network topology, with a few number of largely connected nodes or hubs (these can be seen as dark grey and black colored nodes), and a large number of nodes with few connections (these can be seen as light grey and white colored nodes). This figure was constructed using Gephi.

Rietz and Robatzek [30] validated the ETR1-FLS2 interaction and suggest a requirement of ethylene signaling for FLS2 expression.

Table 3. WOEN edges validated for a selection of IRGs.

IRG	Prediction	References
FLS2 (AT5G46330)	<i>BAK1</i> (AT4G33430)	[24, 26]
	<i>BIK1</i> (AT2G39660)	[26, 27]
	<i>PBL1</i> (AT3G55450)	[27]
	<i>ETR1</i> (AT1G66340)	[30]
	<i>PUB12</i> (AT2G28830)	[29]
	<i>PUB13</i> (AT3G46510)	[29]
	<i>RPM1</i> (AT3G07040)	[28]
	<i>RPS2</i> (AT4G26090)	[28]
	<i>RPS5</i> (AT1G12220)	[28]
<i>CLV1</i> (AT1G75820)	<i>PSY1</i> (AT1G72300)	[32]
RPS2 (AT4G26090)	<i>ATSK41</i> (AT1G09840)	[31]
	<i>FLS2</i> (AT5G46330)	[28]
<i>RPS4</i> (AT5G45250)	<i>SGT1A</i> (AT4G23570)	[34]

In addition to the RPS2-FLS2 interaction, the complex between RPS2 and ATSK41 or AtHIR1 was also reported [31]. RPS2 activates effector-triggered immunity (ETI) after recognizing the bacterial effector protein AvrRpt2. ATSK41 is a hypersensitive protein that is enriched in the plasma membrane. It was identified to be a component of RPS2 complexes [31]. Qi, Tsuda, Nguyen, Wang, Lin, Murphy, Glazebrook, Thordal-Christensen and Katagiri [31] showed that ATSK41 and RPS2 are physically associated and contribute to ETI in presence of *Pseudomonas syringae* pv. tomato DC3000.

Other edges predicted for *CLV1* and *RPS4* are referred in (Table 3). *CLV1* controls shoot and floral meristem size. Equally, *PSY1* is a tyrosine-sulfated peptide hormone. This hormone stimulates cellular proliferation and maintenance of root stem cells [32]. *PSY1* and other secreted peptide hormones such as *CLE2*, suffer post-translational modifications and could function as ligands of *CLV1* [32]. Finally, the R protein *RPS4* specifies resistance to *Pseudomonas syringae* pv. tomato expressing *avrRps4*. *SGT1* is an ubiquitin ligase-associated protein proved to have a role in host and non-host resistance [33]. The work of Li, Li, Bi, Cheng, Li and Zhang [34] suggested that *SGT1* conforms a complex that negatively regulates *RPS4* accumulation. All in all, the functional predictions for these Arabidopsis IRGs are well documented and therefore, the WOENs can be mined for potential IRGs.

CONCLUSION

The proposed procedure allows obtaining a global enzymatic network and a gene network for any organism for which genomic data is available. Topological analyses showed the

graph transformation at each step. The tendency of nodes to cluster remains constant along the process, while an improvement in connectivity and noise reduction was observed after the blast search and expression data integration. The WOEN edges are reinforced with the biological data found in literature. Furthermore, our results from the merging of immunity microarray data and the obtained metabolic network, predict a strong relationship between some genes immune processes in Arabidopsis.

AVAILABILITY AND REQUIREMENTS

Pre-processing of microarray data and similarity measurements between genes based on the gene expression profiles, can be obtained with the aid of several software packages. We recommend using R (R Development Core Team, 2011). Perl is free software and the scripts described in this work can be run in any LINUX/UNIX operating system with a running installation of BLAST+. Scripts are available under request at llopezk@unal.edu.co.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This research has been partially financed by a Grant of the Colombian Banco de la República (Project # 3202).

AUTHORS' CONTRIBUTIONS

- Liliana López-Kleine proposed and directed the Banco de la República Project proposing. Jorge Ramírez was the co-director of this Project. They design the strategy of the software together. Andrés Quintero wrote all scripts and implemented the software for the *A. thaliana* data in Perl. Luis Leal analyzed the topology of the obtained network and reviewed biological information on immunity processes of *A. thaliana*
- Andrés Quintero is a biologist and actually is a student of the Master of Science – Biology program at Universidad Nacional de Colombia – Sede Bogotá. His main research area is reconstruction of biological networks.
- Jorge Ramírez is a civil engineer with a MSc and a PhD in mathematics. He currently holds an associate professorship at the Mathematics department of Universidad Nacional de Colombia. His research is mainly focused on stochastic processes and their applications to problems in the natural sciences.
- Luis Leal is a chemical engineer with MSc in statistics at Universidad Nacional de Colombia. His research focuses on statistical approaches to genomics.
- Liliana López-Kleine is a biologist with PhD in statistics and actually associate profesor at the Universidad Nacional – Sede Bogotá Statistics Department. Her main research area is Statistical Genomics.

SUPPLEMENTARY MATERIALS

Supplementary material is available on the publisher's web site along with the published article.

ABBREVIATIONS

GESM	=	Gene Expression Similarity Matrix
GEN	=	Global Enzymatic Network
OEN	=	Organismic Enzymatic Network
GSE	=	Gold Standard Enzymes
aa	=	Aminoacid
WOEN	=	Weighted Organismic Enzymatic Network
IRGs	=	Immunity Related Genes

REFERENCES

- Almaas, E.; Oltvai, Z. N.; Barabási, A.-L. The Activity Reaction Core and Plasticity of Metabolic Networks. *PLoS Comput. Biol.*, **2005**, *1*, e68.
- Christensen, C.; Thakar, J.; Albert, R. Systems-Level Insights into Cellular Regulation: Inferring, Analysing, and Modelling Intracellular Networks. *IET Syst. Biol.*, **2007**, *1*, 61.
- Oberhardt, M. a.; Palsson, B. Ø.; Papin, J. a. Applications of Genome-Scale Metabolic Reconstructions. *Mol. Syst. Biol.*, **2009**, *5*, 320.
- Plata, G.; Fuhrer, T.; Hsiao, T.-L.; Sauer, U.; Vitkup, D. Global Probabilistic Annotation of Metabolic Networks Enables Enzyme Discovery. *Nat. Chem. Biol.*, **2012**, 1-7.
- Gilman, S. R.; Chang, J.; Xu, B.; Bawa, T. S.; Gogos, J. a.; Karayiorgou, M.; Vitkup, D. Diverse Types of Genetic Variation Converge on Functional Gene Networks Involved in Schizophrenia. *Nat. Neurosci.*, **2012**, 1-7.
- Lopez Kleine, L.; Monnet, V.; Pechoux, C.; Trubuil, A. Role of Bacterial Peptidase F Inferred by Statistical Analysis and Further Experimental Validation. *HFSP J.*, **2008**, *2*, 29-41.
- Meyer, P.; Alexopoulos, L. G.; Bonk, T.; Califano, A.; Cho, C. R.; de la Fuente, A.; de Graaf, D.; Hartemink, A. J.; Hoeng, J.; Ivanov, N. V.; Koeppl, H.; Linding, R.; Marbach, D.; Norel, R.; Peitsch, M. C.; Rice, J. J.; Royyuru, A.; Schacherer, F.; Sprengel, J.; Stolle, K.; Vitkup, D.; Stolovitzky, G. Verification of Systems Biology Research in the Age of Collaborative Competition. *Nat. Biotechnol.*, **2011**, *29*, 811-815.
- Yamanishi, Y.; Vert, J.-P.; Nakaya, A.; Kanehisa, M. Extraction of Correlated Gene Clusters from Multiple Genomic Data by Generalized Kernel Canonical Correlation Analysis. *Bioinformatics*, **2003**, *19*, i323-i330.
- Yamanishi, Y.; Vert, J.-P.; Kanehisa, M. Protein Network Inference from Multiple Genomic Data: A Supervised Approach. *Bioinformatics*, **2004**, *20* Suppl 1, i363-70.
- Chen, L.; Vitkup, D. Predicting Genes for Orphan Metabolic Activities Using Phylogenetic Profiles. *Genome Biol.*, **2006**, *7*, R17.
- Price, N.; Shmulevich, I. Biochemical and Statistical Network Models for Systems Biology. *Curr. Opin. Biotechnol.*, **2007**, *18*, 365-370.
- Gautier, L.; Cope, L.; Bolstad, B. M.; Irizarry, R. A. Affy--Analysis of Affymetrix GeneChip Data at the Probe Level. *Bioinformatics*, **2004**, *20*, 307-315.
- Irizarry, R. A.; Hobbs, B.; Collin, F.; Beazer-Barclay, Y. D.; Antonellis, K. J.; Scherf, U.; Speed, T. P. Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. *Biostatistics*, **2003**, *4*, 249-264.
- Atias, O.; Chor, B.; Chamovitz, D. A. Large-Scale Analysis of Arabidopsis Transcription Reveals a Basal Co-Regulation Network. *BMC Syst. Biol.*, **2009**, *3*, 86.
- Dozmorov, M. G.; Giles, C. B.; Wren, J. D. Predicting Gene Ontology from a Global Meta-Analysis of 1-Color Microarray Experiments. *BMC Bioinformatics*, **2011**, *12* Suppl 1, S14.
- Leal, L. G.; López, C.; López-Kleine, L. Construction and Comparison of Gene Co-Expression Networks Based on Immunity Microarray Data from Arabidopsis, Rice, Soybean, Tomato and Cassava. In *Advances in Computational Biology SE - 3*; Castillo, L. F.; Cristancho, M.; Isaza, G.; Pinzón, A.; Rodríguez, J. M. C., Eds. Springer International Publishing, **2014**, 232, 13-19.
- Barrett, A.J.; Cantor, C.R.; Liébecq, C.; Moss, G.P.; S.W. *Enzyme Nomenclature*; Academic Press: New York, **1992**; pp. 5-22.
- Kharchenko, P.; Church, G. M.; Vitkup, D. Expression Dynamics of a Cellular Metabolic Network. *Mol. Syst. Biol.*, **2005**, *1*, 2005.0016.
- Costa, L. D. F.; Rodrigues, F. A.; Travieso, G.; Boas, P. R. V. Characterization of Complex Networks: A Survey of Measurements. *Adv. Phys.*, **2005**, *56*, 78.
- Csardi, G.; Nepusz, T. The Igraph Software Package for Complex Network Research. *Int. J.*, **2006**, *1*, 1695.
- Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks, **2009**.
- Mazuric, A.; Bonchev, D.; Schwikowski, B.; Buck, G. a. Evolution of Metabolic Network Organization. *BMC Syst. Biol.*, **2010**, *4*, 59.
- Radrich, K.; Tsuruoka, Y.; Dobson, P.; Gevorgyan, A.; Swainston, N.; Baart, G.; Schwartz, J.-M. Integration of Metabolic Databases for the Reconstruction of Genome-Scale Metabolic Networks. *BMC Syst. Biol.*, **2010**, *4*, 114.
- Chinchilla, D.; Zipfel, C.; Robatzek, S.; Kemmerling, B.; Nürnberger, T.; Jones, J. D. G.; Felix, G.; Boller, T. A Flagellin-Induced Complex of the Receptor FLS2 and BAK1 Initiates Plant Defence. *Nature*, **2007**, *448*, 497-500.
- Chinchilla, D.; Shan, L.; He, P.; De Vries, S.; Kemmerling, B. One for All: The Receptor-Associated Kinase BAK1. *Trends Plant Sci.*, **2009**, *14*, 535-541.
- Lu, D.; Wu, S.; Gao, X.; Zhang, Y.; Shan, L.; He, P. A Receptor-like Cytoplasmic Kinase, BIK1, Associates with a Flagellin Receptor Complex to Initiate Plant Innate Immunity. *Proc. Natl. Acad. Sci. USA.*, **2010**, *107*, 496-501.
- Zhang, J.; Li, W.; Xiang, T.; Liu, Z.; Laluk, K.; Ding, X.; Zou, Y.; Gao, M.; Zhang, X.; Chen, S.; Mengiste, T.; Zhang, Y.; Zhou, J.-M. Receptor-like Cytoplasmic Kinases Integrate Signaling from Multiple Plant Immune Receptors and Are Targeted by a Pseudomonas Syringae Effector. *Cell Host. Microbe.*, **2010**, *7*, 290-301.
- Qi, Y.; Tsuda, K. Physical Association of Pattern - triggered Immunity (PTI) and Effector - triggered Immunity (ETI) Immune Receptors in Arabidopsis. *Mol. plant*, **2011**, *12*, 702-708.
- Lu, D.; Lin, W.; Gao, X.; Wu, S.; Cheng, C.; Avila, J. Direct Ubiquitination of Pattern Recognition Receptor FLS2 Attenuates Plant Innate Immunity. *Science (80)*, **2011**, *332*, 1439-1442.
- Mersmann, S.; Bourdais, G.; Rietz, S.; Robatzek, S. Ethylene Signaling Regulates Accumulation of the FLS2 Receptor and Is Required for the Oxidative Burst Contributing to Plant Immunity. *Plant Physiol.*, **2010**, *154*, 391-400.
- Qi, Y.; Tsuda, K.; Nguyen, L. V.; Wang, X.; Lin, J.; Murphy, A. S.; Glazebrook, J.; Thordal-Christensen, H.; Katagiri, F. Physical Association of Arabidopsis Hypersensitive Induced Reaction Proteins (HIRs) with the Immune Receptor RPS2. *J. Biol. Chem.*, **2011**, *286*, 31297-31307.
- Matsubayashi, Y. Post-Translational Modifications in Secreted Peptide Hormones in Plants. *Plant Cell Physiol.*, **2011**, *52*, 5-13.
- Peart, J. R.; Lu, R.; Sadanandom, A.; Malcuit, I.; Moffett, P.; Brice, D. C.; Schausser, L.; Jaggard, D. a W.; Xiao, S.; Coleman, M. J.; Dow, M.; Jones, J. D. G.; Shirasu, K.; Baulcombe, D. C. Ubiquitin Ligase-Associated Protein SGT1 Is Required for Host and Nonhost Disease Resistance in Plants. *Proc. Natl. Acad. Sci. USA.*, **2002**, *99*, 10865-10869.
- Li, Y.; Li, S.; Bi, D.; Cheng, Y. T.; Li, X.; Zhang, Y. SRFR1 Negatively Regulates Plant NB-LRR Resistance Protein Accumulation to Prevent Autoimmunity. *PLoS Pathog.*, **2010**, *6*, e1001111.