Article

Unraveling the WRKY transcription factors network in *Arabidopsis Thaliana* by integrative approach

Mouna Choura[1], **Ahmed Rebaï**[2], **Khaled Masmoudi**[3]

[1]Laboratory of Plant Protection and Improvement, Center of Biotechnology of Sfax, University of Sfax, Route Sidi Mansour Km 6, P.O.Box 1177, 3018 Sfax, Tunisia

[2]Molecular and Cellular Diagnosis Processes, Center of Biotechnology of Sfax, University of Sfax, Route Sidi Mansour Km 6, P.O.Box 1177, 3018 Sfax, Tunisia

[3]International Center for Biosaline Agriculture (ICBA), P.O.Box 14660, Dubai, United Arab Emirates

E-mail: mouna.choura@cbs.rnrt.tn

Received October 28 2014; Accepted December 5 2014; Published online June 1 2015

Abstract

The WRKY transcription factors superfamily are involved in diverse biological processes in plants including response to biotic and abiotic stresses and plant immunity. Protein-protein interaction network is a useful approach for understanding these complex processes. The availability of *Arabidopsis Thaliana* interactome offers a good opportunity to do get a global view of protein network. In this work, we have constructed the WRKY transcription factor network by combining different sources of evidence and we characterized its topological features using computational tools. We found that WRKY network is a hub-based network involving multifunctional proteins denoted as hubs such as WRKY 70, WRKY40, WRKY 53, WRKY 60, WRKY 33 and WRKY 51. Functional annotation showed seven functional modules particularly involved in biotic stress and defense responses. Furthermore, the gene ontology and pathway enrichment analysis revealed that WRKY proteins are mainly involved in plant-pathogen interaction pathways and their functions are directly related to the stress response and immune system process.

Keywords hub; *in silico*; network; plant immunity; stress; WRKY.

Network Biology
ISSN 2220-8879
URL: http://www.iaees.org/publications/journals/nb/online-version.asp
RSS: http://www.iaees.org/publications/journals/nb/rss.xml
E-mail: networkbiology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: InternationalAcademyofEcologyand Environmental Sciences

1 Introduction

Plants are frequently exposed to various stresses such as drought, salinity, cold and pathogen attacks. WRKY are large protein family of zinc-finger transcriptional regulators in higher plants involved in biological processes and in biotic and abiotic stress response (Ramamoorthy et al., 2008). They are involved in the

regulation of various physiological programs that are unique to plants, including plant pathogen defense, senescence, trichome development (Eulgem, 2000), hormone signaling (Chen et al., 2010; Shang et al., 2010) and secondary metabolism (Wang et al., 2010; Suttipanta et al., 2011). Thus, WRKY proteins play important roles in plant. The WRKY proteins are represented with 74 members in *Arabidopsis* and more than 100 members in rice (Wu et al., 2005; Zhang and Wang, 2005). It is evident that one dimensional annotation is no longer enough to study regulatory proteins that often act with one or more protein partners. Few WRKY – interacting proteins studies were reviewed by Chi et al. (2013) showing the complex regulatory and functional network of WRKY transcription factors and giving insights into biological processes that they do regulate. The availability of the Arabidopsis interactome (Arabidopsis Interactome Mapping Consortium 2011) is encouraging to improve our understanding of WRKY transcription factors family.

Although WRKY proteins were relatively recently discovered class of sequence-specific DNA-binding transcription factors (Rushton et al., 2010), they were extensively studied. Consequently, the literature corpus for WRKY is very large; over 400 articles were retrieved when querying with the generic keyword “WRKY” without counting the large number of protein family members and synonymous.

Our goal is to get a global understanding of comprehensive WRKY protein interactions network in *Arabidopsis thaliana* taking into account available interaction data sources.

For such systems biology analysis, computational approach is needed, where high throughput experiments like yeast two-hybrid (Y2H) analyses, microarrays and text mining of literature are integrated. We then revealed some statistical properties of the network in order to explore the topology and functions of the WRKY network in *Arabidopsis thaliana*.

2 Methods

2.1 Dataset

We queried the Uniprot (release 2014_08) (http://www.uniprot.org/) for WRKY transcription factor in *Arabidopsis thalina* based on (Eulgem, 2000) and we retrieved 74 protein identifiers by excluding hypothetical proteins.

2.2 WRKY-WRKY interaction network construction

The WRKY-WRKY interaction network was performed by version 9.05 STRING (http://string-db.org/) (Franceschini et al. 2013), which is a database of known and predicted protein interactions. The interactions include physical and functional associations derived from genomic context, high-throughput experiments, co-expression and previous knowledge. The interaction data were then assigned by confidence score for each source of evidence.

2.3 Statistical analysis and topological features of WRKY-WRKY network

In a protein-protein interaction (PPI) network, a node denotes a protein and an edge denotes an interaction between two proteins. For each protein set, we applied four topological measures to assess its role in the network: degree, clustering coefficient, betweenness, and shortest-path distance. First, for a node in PPI network, the degree measures the number of links for a node to other nodes. Second, the clustering coefficient of a node is the ratio of the observed number of direct connections between the node’s immediate network neighbours over the maximum possible number of such connections. Third, the betweenness of a node is defined as the number of shortest paths between all possible pairs of nodes in the network that traverse the node. Fourth, for a pair of selected nodes in the network, there are many alternative paths between them. The

path with the smallest number of links is defined as the shortest path. The number of links passing through in the shortest path is defined as shortest-path distance. We tested whether the interaction network is scale-free or not by plotting the node degree distribution. A network may resemble a scale-free topology if the distribution follows a power-law.

The topological and statistical significance of network have been calculated using Cytoscape plugins Network Analyzer (Assenov et al., 2008) and CentiScaPe (Scardoni et al., 2009).

More explanations of calculated parameters are available at http://med.bioinf.mpi-inf.mpg.de/netanalyzer/help/2.6.1/index.html.

3 Results and Discussion

3.1 WRKY Network Construction and statistical properties

We integrated results derived from different sources including genomic context, high-throughput experiments, coexpression (conserved), homology among species and text mining related to WRKY transcription factor family proteins of *Arabidopsis thaliana.* We obtained a protein network made up of 72 interactions as described in supplemental material file 1. Among these interactions, only eight are validated experimentally mainly by two-hybrid assay, sixty five protein interactors are coexpressed and all protein pairs are co-cited in bibliography. This finding provides a starting point to validate the other interactions.

By removing disconnected nodes (proteins), advanced view shows highly connected network constituted of 26 WRKY proteins (Fig. 1). Topological parameters of WRKY network are listed in Table 1 (See supplemental material for more details). By plotting the degree distribution in a log scale, we obtained a linear regression R^2 value of 0.35. This is inferring the presence of significant number of negative feedback loops.

Table 1 Network parameters calculated by Network Analyzer.

Number of nodes	26
Number of edges	71
Clustering Coefficient	0.430
Network diamater	6
Network radius	3
Network density	0.218
Network heterogeneity	0.645
Average number of neighbors	5.462
Characteristic path length	2.625

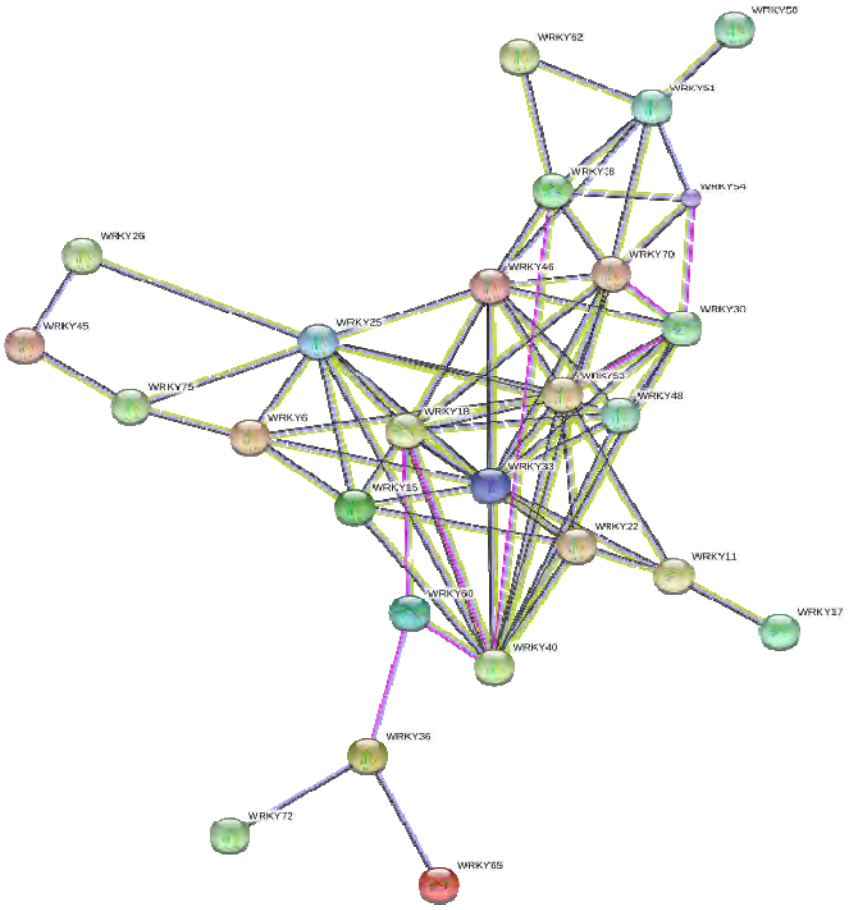

Fig. 1 Evidence view of WRKY proteins network in *Arabidopsis thaliana*.

3.2 Hubs identification

Highly connected nodes are usually defined as 'Hubs'. Based on topological parameters, we have identified protein hubs having often higher degree, clustering coefficient, betweenness and shortest-path distance. By combining these parameters, we have listed the highly ranked nodes in Table 2 (See supplemental material file 2 for complete list). The hubs are thought to maintain network robustness by having higher connectivity in the whole network and may mediate key biological pathways such as signal transduction in the WRKY protein network compared to the other proteins. Our results are in good agreement with previous studies providing strong evidence of WRKY70 role in plant senescence and defense signaling pathways (Ulker et al., 2007) and its cooperation with WRKY54 as negative regulator (Besseau et al., 2012). Moreover, Birkenbihl et al. (2012), revealed the transcriptional regulator role of WRKY33 in *Arabidopsis thaliana* upon *Botrytis cinerea* infection by targeting redox homeostasis, salicylic acid signaling, ethylene jasmonic acid mediated cross communication, and camalexin biosynthesis. Nevertheless, these reports consider that they are still crucial components to identify the full resistance mechanism.

Table 2 Potential hubs in the WRKY network.

Protein	Degree	Clustering coefficient	Betweenness
WRKY40	12	0.5	137
WRKY33	12	0.5	132
WRKY46	10	0.6	125.5
WRKY25	10	0.5	94
WRKY18	10	0.6	75
WRKY53	10	0.6	72.7
WRKY70	9	0.6	67.3

3.3 Functional modules in WRKY network

Among the seventy two interactions, we have found seven functional modules (Fig. 2) particularly related to stress response including response to chitin and organic compounds (13 nodes), response to salicylic acid (8 nodes) and defense response to bacterium (13 nodes) and fungus (7 nodes). Additionally, we have found small modules related to calmodulin binding (7 nodes), immune effector process (3 nodes), jasmonic acid mediated pathway (3 nodes) and leaf and organ senescence (3 nodes). Some proteins are present in multiple functional modules such as WRKY 70, WRKY40, WRKY 53, WRKY 60, WRKY 33 and WRKY 51. It could be argued that such proteins are multifunctional and participate in different biological processes. It is worthy to note that these proteins are aforementioned as hubs of the WRKY network. Thus, it would be benefit to further investigate their roles particularly with respect to stress response.

3.4 Pathways and Gene Ontology (GO) enrichment in WRKY network

The enrichment analysis of the network revealed the functional features of WRKY proteins. From the KEGG pathway enrichment, we found that WRKY proteins are included mainly in plant-pathogen interaction pathway (Fisher's exact test, FDR p-value =1.5 10^{-2}). (http://www.genome.jp/kegg/pathway/ath/ ath04626.html). Then the enriched GO functions underlying the WRKY proteins are particularly related to response to stress, immune system process, transport and signal transduction (supplementary material file 3). Moreover, this may explain the highly connected architecture of WRKY network. Such system, upon stress, should dynamically respond with the cooperation of all interactors. This is in line with a previous report underlying the role of WRKY transcription factors in plant system immunity (Jones and Dangl, 2006).

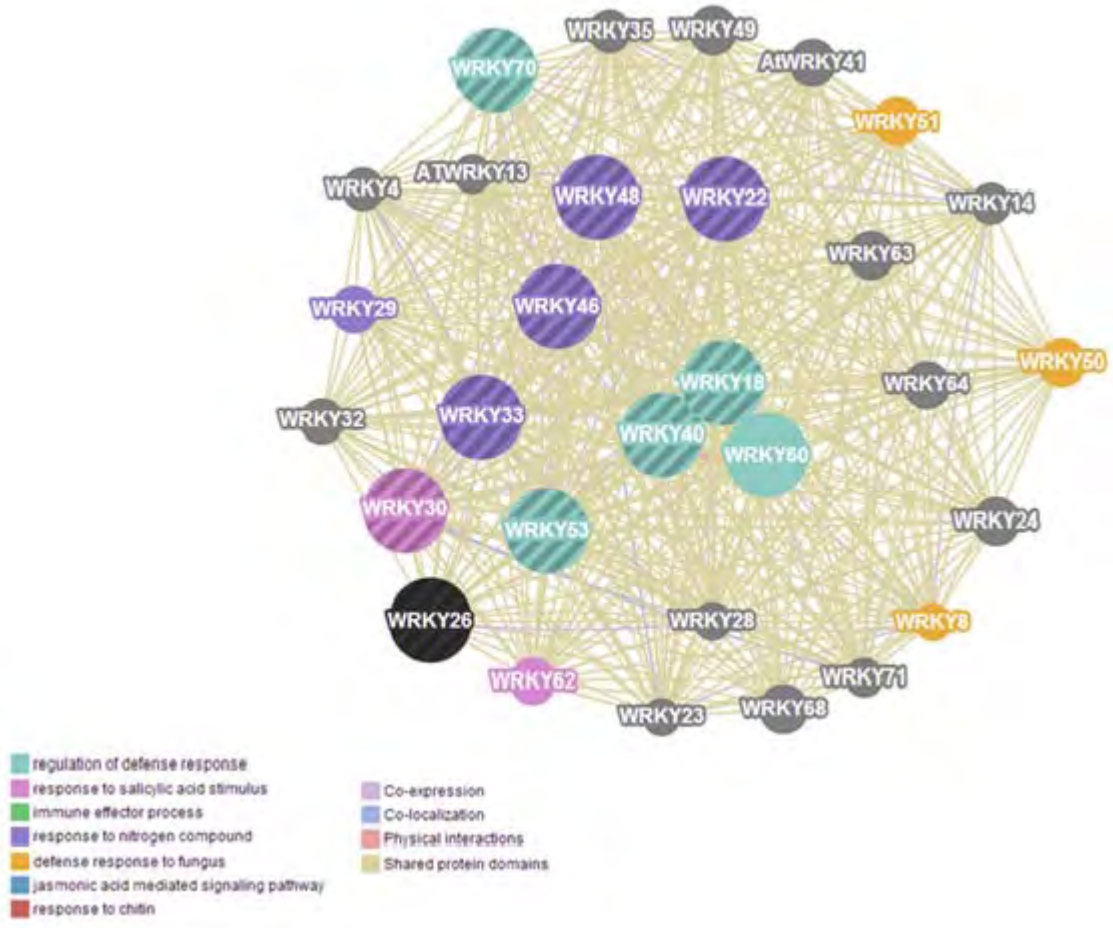

Fig. 2 Functional modules in the WRKY proteins network in *Arabidopsis thaliana.*

4 Conclusions

In this study, we have constructed a WRKY transcription factors network in *Arabidopsis thaliana* using computational tools. By applying integrative approach combining genomic context, high throughput experiments, database content and text mining sources, we get a highly connected network of 72 interactions connected by more than ten hub proteins. This is a fundamental step to uncover plant-pathogen interaction pathway and stress response mediated by WRKY proteins which are not fully established so far. We believe that our approach may be considered as a starting point to address these questions.

Acknowledgements

This work was supported by the Ministry of Higher Education and Scientific Research, Tunisia.

References

Arabidopsis Interactome Mapping Consortium. 2011. Evidence for network evolution in an Arabidopsis interactome map. Science, 333(6042): 601-607

Assenov Y, Ramírez F, Schelhorn SE, et al. 2008. Computing topological parameters of biological networks. Bioinformatics, 24(2): 282-284

Besseau S, Li J, Palva ET. 2012. WRKY54 and WRKY70 co-operate as negative regulators of leaf senescence in Arabidopsis thaliana. Journal of Experimental Botany, 63(7): 2667-2679

Birkenbihl RP, Diezel C, Somssich IE. 2012. Arabidopsis WRKY33 is a key transcriptional regulator of hormonal and metabolic responses toward Botrytis cinerea infection. Plant Physiology, 159(1): 266-285

Chen H, Lai Z, Shi J, Xiao Y, Chen Z, Xu X. 2010. Roles of Arabidopsis WRKY18, WRKY40 and WRKY60 transcription factors in plant responses to abscisic acid and abiotic stress. BMC Plant Biology, 10: 281

Chi Y, Yang Y, Zhou Y, et al. 2013. Protein-Protein interactions in the regulation of WRKY transcription factors. Molecular Plant, 6(2): 287-300

Eulgem T, Rushton PJ, Robatzek S, Somssich IE. 2000. The WRKY superfamily of plant transcription factors. Trends in Plant Science, 5(5): 199-206

Franceschini A, Szklarczyk D, Frankild S, et al. 2013. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Research, 41(D1): D808-D815

Jones J D G, Dangl J L. 2006. The plant immune system. Nature, 444(7117): 323-329

Ramamoorthy R, Jiang SY, Kumar N, Venkatesh PN, Ramachandran S. 2008. A comprehensive transcriptional profiling of the WRKY gene family in rice under various abiotic and phytohormone treatments. Plant and Cell Physiology, 49(6): 865-879

Rushton PJ, Somssich IE, Ringler P, Shen QJ. 2010. WRKY transcription factors. Trends in Plant Science, 15 (5): 247-258

Scardoni G, Petterlini M, Laudanna C. 2009. Analyzing biological network parameters with CentiScaPe. Bioinformatics, 25(21): 2857-2859

Shang Y, Yan L, Liu ZQ, et al. 2010. The Mg-chelatase H subunit of Arabidopsis antagonizes a group of WRKY transcription repressors to relieve ABA-responsive genes of inhibition. Plant Cell, 22(6): 1909-1935

Suttipanta N, Pattanaik S, Kulshrestha M, et al. 2011. The transcription factor CrWRKY1 positively regulates the terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. Plant Physiology, 157(4): 2081-2093

Ulker B, Shahid Mukhtar M, Somssich IE. 2007. The WRKY70 transcription factor of Arabidopsis influences both the plant senescence and defense signaling pathways. Planta, 226(1): 125-137

UniProt Consortium. 2013. Update on activities at the Universal Protein Resource (UniProt) in 2013. 2014. Nucleic Acids Research, 41(D1): D43-D47

Wang H, Avci U, Nakashima J, et al. 2010. Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants. Proceedings of the National Academy of Sciences of USA, 107(51): 22338–22343

Wu KL, Guo ZJ, Wang HH, Li J. 2005. The WRKY family of transcription factors in rice and Arabidopsis and their origins. DNA Research, 12(1): 9-26

Zhang Y, Wang L. 2005. The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. BMC Evolutionary Biology, 5: 1