Plant Gene 9 (2017) 19–25

Contents lists available at ScienceDirect

Plant Gene

journal homepage: www.elsevier.com/locate/plantgene

Dissecting a co-expression network of basic helix-loop-helix (*bHLH*) genes from phosphate (Pi)-starved soybean (*Glycine max*)

Ertugrul Filiz [a,*], Recep Vatansever [b], Ibrahim Ilker Ozyigit [b]

[a] Duzce University, Department of Crop and Animal Production, Cilimli Vocational School, 81750 Cilimli, Duzce, Turkey
[b] Marmara University, Faculty of Science and Arts, Department of Biology, 34722 Goztepe, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 18 October 2016
Received in revised form 8 December 2016
Accepted 10 December 2016
Available online 16 December 2016

Keywords:
Cis-element
Cluster
Gene module
Seed genes
bHLH
Co-expression
Soybean

ABSTRACT

In this work, a co-expression network-based approach was employed in analysis of basic helix-loop-helix (*bHLH*) transcription factor genes in soybean (*Glycine max*). *bHLHs* as regulatory agents modulate many complex associations related with regular metabolic functions and various stress factors. Co-expression networks are versatile resources to understand complex associations from functional aspects. Using a microarray data from phosphate (Pi)-starved soybean, a co-expression network of *bHLH* genes was constructed. A network was established with 253 nodes (*bHLH* genes) interconnecting 1763 edges (association) and then the network was dissected into 13 distinct clusters to extensively investigate the correlations. Each cluster was individually analyzed with emphasis on seed genes, which could be used as marker/reference genes in development of plant lines with enhanced stress tolerance. The seed genes were involved in very diverse metabolic processes, including stress modulation, metal homeostasis, hormone response and developmental roles. Our network-based clustering approach provides new insight in understanding many uncharacterized plant *bHLHs*.

1. Introduction

Plants are constantly exposed to variety of environmental stresses, which are thought to be the prime culprit for crop losses in the world (Huang et al., 2012). Exposure to stressors leads to differential expression of hundreds-to-thousands of genes, and from these genes the manipulation of regulatory agents like transcription factors (TFs) have been reported to be a powerful approach to improve plant stress tolerance (Nakashima et al., 2009; Golldack et al., 2011; Huang et al., 2013). Several TF families, *DREB*, *WRKY*, *MYB*, *bZIP* and *bHLH* have been reported in the regulation of abiotic stress-related signal transduction network (Yamaguchi-Shinozaki and Shinozaki, 2006; Chen et al., 2002). From these, basic helix-loop-helix (*bHLH*) is one of the largest regulatory protein families in eukaryotic organisms (Kim and Kim, 2006; Chinnusamy et al., 2003). bHLHs consist of ~50–60 amino acid residues with two functionally distinct regions, the basic region and HLH region. The basic region is located at the N-terminus site with about 15 residues, where six residues are usually basic residues. This region was reported to play a crucial role as a DNA-binding motif (Li et al., 2006; Atchley et al., 1999). The helix-loop-helix (HLH) domain is located at the C-terminus site and includes two amphipathic α-helices with a variable length linking loop; this domain has a function as protein-protein interaction motif for homo- or heterodimer formation (Atchley et al., 1999; Ellenberger et al., 1994). Most plant bHLHs have a basic region domain that allow them to recognize E-box (5′-CANNTG-3′) and/or G-Box (5′-CACGTG-3′) sequences in target gene promoters (Castilhos et al., 2014; Robinson et al., 2000). In addition, analysis of 638 plant bHLHs from *Arabidopsis*, poplar, rice, moss and algae revealed 32 subfamilies as clades with high support values (Carretero-Paulet et al., 2010). In these clades, most *bHLH* genes (around 80%) possessed one to three introns at *bHLH* coding region while the rest had no introns (Carretero-Paulet et al., 2010). Plant *bHLHs* have been reported to regulate very diverse biological processes such as flowering (Ito et al., 2012), root hair, chloroplast and fruit development (Nicolas et al., 2013; Tominaga-Wada et al., 2012), nodule vascular patterning (Godiard et al., 2011), light and ROS signaling (Chen et al., 2013), ABA and jasmonate regulation (Li et al., 2007) and circadian rhythm modulation (Hanano et al., 2008). Most *bHLH*-regulated metabolic processes were also related to the regulation of various abiotic stresses (Castilhos et al., 2014). Collectively, bHLHs modulate a network of complex associations related with regular metabolic functions and various stresses.

Phosphorus is an essential macronutrient for plant growth and development thereby its limitation is a constraint for crop yield and quality (Wang et al., 2016a). Pi-starvation in soybean has been reported to cause dwarfed plants, reduction in leaf areas, necrotic spots on lower leaves, increased flowering time and decreased pods during fruiting time (Wang et al., 2016a). However, plants have developed different strategies at molecular and physiological levels to cope with Pi-changes. For example, a root plasma membrane H^+-ATPase was reported to play

* Corresponding author.
E-mail address: ertugrulfiliz@gmail.com (E. Filiz).

http://dx.doi.org/10.1016/j.plgene.2016.12.001
2352-4073/

a role in soybean adaptation to P-limitation (Shen et al., 2006). P-addition to topsoil caused soybean roots to become shallower and also increased P-efficiency, and thereby plant yield (Zhao et al., 2004). Transcriptome analysis of Pi-deficient soybean roots revealed differentially expressed genes, possibly involved in Pi-homeostasis via calcium signaling modulation, nutrient uptake/transport, ROS control, gene transcription and hormonal signaling (Zeng et al., 2016). In a different study, sequencing of RNA libraries from soybean Pi-sufficient and -deficient conditions revealed four P-responsive miRNAs such as miR399, nov_6, nov_9 and nov_10 (Xu et al., 2013). Moreover, a few TFs characterized, including some of the bHLH family, were reported to either positively or negatively regulate subsets of Pi-stress response (Baek et al., 2013). *bHLH32* negatively regulated some Pi-starvation responses (Chen et al., 2007). Rice and maize PTF1 encoding a bHLH TF was involved in Pi-signaling (Yi et al., 2005; Li et al., 2011). Wheat bHLH1 played a crucial role in adaptation to Pi- and N-starvation via transcriptional regulation of a set of genes (Yang et al., 2016). Moreover, Pi-starvation responsive TFs were also reported to exert their functions via a cross-talk between Pi-starvation and phytohormones or photosynthates (Rouached et al., 2010).

In the post-genomic era, sequenced genomes of model plants and high-throughput technologies can produce an extensive collection of gene expression data. Co-expression networks, which are constructed by using expression data, are one of the versatile resources that allow to understand complex network associations from functional aspects (Serin et al., 2016). Thus, a gene co-expression analysis was performed herein to understand correlations of *bHLH* genes in phosphate (Pi)-starved soybean.

2. Materials and methods

2.1. Gene expression data

Microarray gene expression data was retrieved from GEO datasets of NCBI, deposited under GSE78242 accession number (ncbi.nlm.nih.gov/geo/; Wang et al., 2016a). The arrays contained samples of phosphorus treatments from two soybean accessions such as low-P tolerant (CD) and low-P sensitive (YH). Samples were treated with normal (control: half Hoagland with 0.5 mmol/L P) and low (treatment: half Hoagland with 0.005 mmol/L P) phosphorus applications. Plants were hydroponically grown under regimens of 16 h/8 h (day/night) photoperiod and 26–28 °C/22 °C (day/night) temperature. Datasets each with three biological replicates were processed using NCBI's GEO2R tool, which internally employs GEOquery and limma R packages from the Bioconductor project. Calculated log2 gene expression values of Pi-treated samples from low-P tolerant (CD) and low-P sensitive (YH) soybeans were used in further analysis.

2.2. Correlation network

A co-expression network of *bHLH* coding genes was constructed using the ExpressionCorrelation plugin of Cytoscape with Pearson's Correlation Coefficient (PCC) method for −0.95&0.95 cutoff (cytoscape.org/; Shannon et al., 2003). The plugin computes a similarity network from genes in an expression matrix, and in network nodes and edges represent the genes and associations, respectively. Constructed networks were visualized by using Cytoscape.

2.3. Network clustering

The *bHLH* network was dissected into clusters using MCODE plugin of Cytoscape. This plugin identifies the intensely connected regions in large protein-protein interaction networks that may represent molecular complexes. MCODE does not provide any statistical score on the resulting clusters and its algorithm operates in three stages, (i) vertex weighting, (ii) complex prediction and (iii) post-processing (for details in scoring please refer to Bader and Hogue, 2003). The total number of network clusters was identified using the PCC method and 0.9 cutoff with adopted settings: degree cutoff = 2, haircut on, fluff off, node score cutoff = 0.2, K-core = 2, Max depth = 100 and loops not included. Each individual cluster and seed genes in clusters were analyzed by using the Network Analyzer tool in Cytoscape. Statistical significance of "degree", "clustering coefficient" and "closeness centrality" measures is usually attributed to higher values in the given network rather than a certain threshold. 1000 bp upstream regions of seed genes were retrieved from SoyKB (soykb.org) and supplied to PlantCARE database for promoter analysis (bioinformatics.psb.ugent.be/webtools/plantcare/html/).

3. Results and discussion

3.1. Co-expression network analysis

A co-expression network of *bHLH* TFs was constructed using microarray data (GEO: GSE78242) of phosphate (Pi)-starved two soybean accessions, low-P tolerant (CD) and low-P sensitive (YH). Initially, global expression profiles of all samples under low-vs-normal phosphorus treatments were processed using NCBI GEO2R tool and then the calculated expression values were retrieved along with relevant annotation files, including protein domain annotations of genes. Accordingly, a total of 262 entries defining "bHLH" term were collected from the annotation file and then duplicated entries/probes were removed; the remaining 253 *bHLH* genes were supplied for network construction. Co-expression networks are mainly constructed by such major steps: (i) a similarity score (or similarity matrix) is calculated between each pair of genes (e.g. PCC method) from expression values of given samples, (ii) significant expressions (e.g. >0.8) in a similarity matrix are replaced by 1 otherwise with 0, (iii) finally a network is constructed based on whether two genes are connected (1 element) or not (0 element) (Zhang and Horvath, 2005). Thus, possible associations (edges) of a gene in a network also give hints about the values of significant expressions for that gene under the given conditions. Herein, a co-expression network was generated using the Cytoscape plugin "ExpressionCorrelation" with PCC method and for −0.95&0.95 cutoff value. The overall network was established with 253 nodes (*bHLH* genes) interconnecting 1763 edges (association) with a measure of 0.601 clustering coefficient and 0.055 network density (Fig. 1). In a similar study, a network density of 58 *bHLH* rice genes under cold and heat stresses was calculated as 0.29 and 0.37 respectively. The network density indicates the portion of potential connections in a network or in other words, it shows the overall closeness of genes (Srivastava et al., 2016). Herein, based on our *bHLH*-scale network, soybean *bHLH* genes were calculated with a lower network density (0.055), suggesting that the presence of diverse groups of bHLH members in the same network may decrease the network density due to functional diversities associated with weak correlations.

To extensively investigate *GmbHLH* correlations, the main *bHLH* soybean network was dissected into 13 distinct clusters using Cytoscape's MCODE plugin (Table 1; Fig. 2). MCODE finds the potential clusters which are highly interconnected regions in a given network (Bader and Hogue, 2003). Thus, deduced clusters herein emphasize how densely the genes are associated with their neighbors. In addition, each individual cluster and seed genes in those clusters were analyzed using the Network Analyzer tool of Cytoscape. The number of nodes and edges in clusters showed a distribution of 5–41 and 10–162 respectively. In the dissected networks, cluster 4 was identified with the highest number of nodes (41 genes) and edges (161 associations), however cluster 1 possessed the highest cluster score (16.25) with 17 nodes associating with 130 edges due to its higher connectivity per node. Then, each single cluster was individually investigated by putting more emphasis on the seed genes. A "seed" has been defined as the highest scoring node in a gene cluster (Bader and Hogue, 2003). Herein,