Discover the Dehydration Response Genes in Boea hygrometrica Transcriptome Using Bayesian Network Approach

Mengmeng Zhang, Lu Wang, Ping Wan*

College of Life Sciences, Capital Normal University, Beijing, China

Email address: wanping@cnu.edu.cn (Ping Wan)
*Corresponding author

To cite this article:
Mengmeng Zhang, Lu Wang, Ping Wan. Discover the Dehydration Response Genes in Boea hygrometrica Transcriptome Using Bayesian Network Approach. Computational Biology and Bioinformatics. Vol. 8, No. 1, 2020, pp. 9-14. doi: 10.11648/j.cbb.20200801.12

Received: March 3, 2020; Accepted: March 18, 2020; Published: March 23, 2020

Abstract: “Drying without dying” is an amazing feature in land plant evolution. Boea hygrometrica is an important resurrection plant model. The current genome and transcriptome analysis have revealed that some biological processes may contribute to its dehydration tolerance, but genes play pivotal roles in the dehydration response remains unclear. Bayesian network approach is a powerful tool for transcriptome data analysis and biological network reconstruction. In this work, by using the Bayesian network approach, we first reconstruct a gene regulation network with the B. hygrometrica transcriptome data. The network contains 1292 genes. Next, we defined the hub node genes in the network and focus on their functions in order to understand the response B. hygrometrica carried out under the dehydration stress. Finally, by an association analysis, we deduce the function of the unknown gene Bhs126_021 which has a degree of 84 in the network. The data-driven strategy we applied in this work not only finds out the knowledge from the knowledge-driven strategy analysis, but also provides novel findings from the B. hygrometrica transcriptome. Our findings give insight of control genes in land plant under the dehydration stress. The data-driven strategy applied in this work can also efficiently analyze other similar transcriptome data sets.

Keywords: Dehydration Response Genes, Boea hygrometrica, Bayesian Network, Transcriptome Analysis

1. Introduction

Boea hygrometrica is a homiochlorophyllous dicot in Gesneriaceae that grows in rocky areas throughout most of China [1]. It is an important plant model for understanding responses to dehydration. In 2015, the draft genome of B. hygrometrica was sequenced. The genome size of B. hygrometrica is about 1.69 Gb. The genome encodes 23,250 genes. The dehydration-induced alteration in gene expression experiments discovered 9,888 differentially expressed genes (DEGs) [2]. Knowledge-based analysis of its transcriptome revealed three major clusters of genes involved in dehydration stress response. Cluster 1 primarily associated with photosynthesis. Cluster 2 was mainly of ABA metabolism and signaling, late embryogenesis abundant proteins (LEAs) and components of ROS protection and detoxification pathways. Cluster 3 primarily encoded proteins for nucleic acid metabolism. However, knowledge-based analysis did not find the genes played key roles in B. hygrometrica under dehydration stress. The key controlling genes for the dehydration tolerance in B. hygrometrica remains unknown.

Bayesian network approach [3] is a promising tool for transcriptome data analysis [4-7] and biological network reconstruction [8-12]. Bayesian network approach is a kind of data-driven analysis method. It is independent of the known knowledge and could mine the novel knowledge merely based on the dataset itself.

To investigate the mechanisms of the dehydration tolerance in B. hygrometrica, in this work we reconstructed the B. hygrometrica gene regulatory network using Bayesian network approach, and discover pivotal control genes in B. hygrometrica against dehydration stress. The pipeline used in this work can also be converted to analyze other transcriptome data.
2. Data & Methods

2.1. Transcriptome Datasets

Genome-wide transcriptome dataset during dehydration in leaf tissues of *B. hygrometrica* was collected from Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/), the accession numbers are GSE48671 and GSE66474.

In this work, we extracted the data of the 1292 two-fold differentially expressed genes to reconstruct the Bayesian network.

2.2. Reconstruction of *B. hygrometrica* Gene regulatory Network Using Bayesian Network Approach

Bayesian network is a probabilistic graphical model that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). It consists of two components: the first component is a directed acyclic graph; and the second component is a set of parameters that quantify the network [9]. A Bayesian network is defined as:

\[ P_b = (X_1, ..., X_n) = \prod_{i=1}^{n} P_b(X_i|pa(X_i)) \]

Where:

- \( X_i \) denotes each variable in DAG, \( pa(X_i) \) denotes all parent nodes of \( X_i \).

In this study, we used the R package bnlearn (http://cran.r-project.org/) to learn the Bayesian network structure.

2.3. Visualization of the Gene Regulatory Network

We use Cytoscape 2.8.3 (http://www.cytoscape.org/) software to visualize the gene regulatory network [13].

3. Result and Discussions

3.1. Gene Regulatory Network of *B. hygrometrica*

First, we reconstructed the *B. hygrometrica* gene regulatory network using Bayesian network approach. The gene regulatory network includes 1292 nodes and 8969 edges (Figure 1). The distribution of the node degree showed that only 114 nodes (less than 10%) have a degree greater than 25 (Figure 2). The distribution follows the power law. The result suggested that *B. hygrometrica* gene regulatory network is a scale-free network. The scale-free networks are remarkably resistant to accidental failures [14]. Since the hub nodes (i.e. nodes having high degree of connectivity) in a scale-free network dominate the overall connectivity of the network, these hub nodes play important roles for maintaining the stability of the network [15].

![Figure 1. Gene regulatory network of *B. hygrometrica*.](image-url)
3.2. Dehydration Response Genes in B. hygrometrica

After reconstructing the gene regulatory network, we focused our analysis on the hub nodes in B. hygrometrica, and expected to find out the key controlling genes involved in the dehydration tolerance in B. hygrometrica.

In this work, we defined the hub node genes as the genes with the degree equal to or greater than 40. We found 58 hub nodes in the B. hygrometrica gene regulatory network. Table 1 shows the 18 hub nodes with degree above 60. Hub nodes play pivotal role in a network. Therefore, these hub node genes are of key genes involved in the dehydration response in B. hygrometrica.

Previous study has shown that three clusters of genes involved in dehydration response [2]. The first cluster is associated with photosynthesis. In this work, by using the Bayesian network approach, we independently discovery that the hub node gene Bhs3_009 with a degree of 79 associates with photosynthesis (Table 1). The second cluster is mainly of ABA metabolism and signaling. In the gene regulatory network, we discovered two hub node genes Bhs6354_003 (52) and Bhs109_092 (43) (Numbers after each gene is the degree of the gene in the network, the same here after) participate in the abscisic acid (ABA) mediated signaling pathway (Table 1). The third cluster encoded proteins for nucleic acid metabolism. In the gene regulatory network we discovered four genes involved in nucleic acid metabolism, including genes Bhs3_009 (79), Bhs211_042 (50), Bhs3_092 (43) and Bhs6835_002 (42) (Table 1). Obviously, by adopting the data-driven strategy, the Bayesian network not only verified the previous three gene clusters, but also discovered the key controlling genes in B. hygrometrica regarding to the dehydration response.

Besides, we also found out seven genes directly response to stress, including genes Bhs85_075 (71), Bhs2169_013 (61), Bhs62_045 (54), Bhs6354_003 (52), Bhs1230_006 (44), Bhs109_092 (43) and Bhs161_011 (41) (Table 1). The results demonstrate that the data-driven strategy based on the Bayesian network is efficient in discovering the important genes in the transcriptome analysis.

| Gene      | Degree | Biological Process                                                                 |
|-----------|--------|----------------------------------------------------------------------------------|
| Bhs126_021| 84     | unknown                                                                          |
| Bhs3_009  | 79     | GO:0016070 RNA metabolic process, GO:0019684 photosynthesis, light reaction       |
| Bhs2862_003| 72   | unknown, GO:0015995 chlorophyll biosynthetic process, GO:0043623 cellular protein complex assembly |
| Bhs85_075 | 71     | unknown, GO:0044267 cellular protein metabolic process                           |
| Bhs1182_005| 71   | unknown                                                                          |
| Bhs3124_012| 71   | unknown                                                                          |
| Bhs29_019 | 70     | GO:0006464 protein modification process, GO:0007167 enzyme linked receptor protein signaling pathway |
| Bhs3495_020| 69   | GO:0051707 response to other organism, GO:0031109 microtubule polymerization or depolymerization |
| Bhs4184_003| 68   | GO:0051234 establishment of localization                                           |
| Bhs211_034| 67     | GO:0006811 ion transport                                                         |
| Bhs23426_001| 67   | unknown                                                                          |
| Bhs1217_008| 66   | unknown                                                                          |
| Bhs13204_001| 66   | unknown                                                                          |
| Bhs22_072 | 63     | GO:0015986 ATP synthesis coupled proton transport                                 |
| Bhs942_007| 63     | GO:0015988 energy coupled proton transport, against electrochemical gradient      |
| Bhs2169_013| 61   | GO:0050996 response to stimulus, GO:0032446 protein modification by small protein conjugation |
| Bhs22_035 | 60     | unknown                                                                          |
| Bhs403_002| 60     | unknown                                                                          |

Figure 2. The distribution of node degrees.
3.3. The Possible Function of the Gene Bhs126_021

The gene Bhs126_021 has the highest degree of 84 in the gene regulatory network suggesting that it may be the most important gene in dehydration response in B. hygrometrica (Table 1). Unfortunately, the function of Bhs126_021 is unknown. We attempted to uncover its function via an association analysis. Since Bhs126_021 regulated seven genes in the regulatory network (Table 2), we investigated the functions of the seven genes, and found that these genes involve in two of three clusters related to dehydration response, including photosynthesis and RNA metabolic process. The result suggested that Bhs126_021 truly involved in dehydration response.

| Gene     | Degree | Biological Process                                                                 |
|----------|--------|-----------------------------------------------------------------------------------|
| Bhs10_105| 54     | GO:0007127 meiosis I                                                              |
| Bhs1834_026| 39    | GO:0048229 gametophyte development                                                 |
| Bhs250010_001| 39 | unknown                                                                           |
| Bhs29_019| 70     | GO:0006464 protein modification process                                            |
| Bhs3_009| 79     | GO:00016070 RNA metabolic process                                                 |
| Bhs3124_012| 71    | GO:0019684 photosynthesis, light reaction                                          |
| Bhs96_022| 50     | GO:0043623 cellular protein complex assembly                                       |

We also investigated the 38 genes with known gene function annotations regulated by the gene Bhs126_021 in the gene regulatory network (Table 3). Among the 38 genes, 11 genes respond to stress, radiation, hormone stimulus, abiotic stimulus or osmotic stress, including genes Bhs6354_003, Bhs63_076, Bhs3141_003, Bhs176_045, Bhs518_033, Bhs6596_001, Bhs194_026, Bhs2167_014, Bhs68_058, Bhs1834_013 and Bhs83_038 (Table 3). We also find genes which involve in photosynthesis (Bhs3271_001), DNA metabolism (Bhs63_076) and ABA mediated signaling pathway (Bhs6354_003). The results also corroborated that the gene Bhs126_021 is a dehydration response gene.

| Genes     | Degree | GO annotation                                                                 |
|-----------|--------|-------------------------------------------------------------------------------|
| Bhs32_040 | 55     | GO:0006810 transport                                                          |
| Bhs942_007| 63     | GO:0044267 cellular protein metabolic process                                 |
| Bhs576_011| 52     | GO:00010467 gene expression                                                   |
| Bhs3271_001| 19   | GO:0009639 response to red or far red light;                                  |
| Bhs6354_003| 52   | GO:0004026 regulation of transcription, RNA-dependent                         |
| Bhs63_076 | 28     | GO:00010467 gene expression                                                   |
| Bhs161_031| 44     | GO:0006464 protein modification process                                        |
| Bhs2041_014| 13   | GO:0008152 metabolic process                                                   |
| Bhs3141_003| 13   | GO:0009631 response to stress                                                 |
| Bhs1799_014| 12   | GO:0006464 protein modification process                                        |
| Bhs225_074| 29     | GO:0023052 signaling                                                          |
| Bhs4430_017| 12   | GO:0007275 multicellular organismal development                               |
| Bhs6659_011| 12   | GO:0003036 developmental process involved in reproduction                     |
| Bhs1655_046| 12   | GO:00096157 plastid organization                                              |
| Bhs1231_005| 11   | GO:0009630 gravitropism                                                       |
| Bhs176_045| 11     | GO:0006351 transcription, DNA-dependent                                       |
| Bhs518_033| 12     | GO:0009725 response to hormone stimulus                                       |
4. Conclusions

The data-driven strategy based on the Bayesian network approach could not only find out the knowledge obtained by the knowledge-driven strategy, but also could discover the novel knowledge in transcriptome data.

Acknowledgements

This study was funded by the scientific research project of Beijing Municipal Commission of education, KM201610028010.

References

[1] Wilson CL, Gazette JB: Floral anatomy in Gesneriaceae. I. Cyrtandroideae. 1974, 135:247-256.
[2] Xiao L, Yang G, Zhang L, Yang X, Zhao S, Ji Z, Zhou Q, Hu M, Wang Y, Chen M, et al: The resurrection genome of Boea hygrometrica: A blueprint for survival of dehydration. *PNAS* 2015, 112:5833-5837.
[3] Pearl J: *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier; 2014.
[4] Gevaert O, Smet FD, Timmerman D, Moreau Y, Moor BDJB: Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. 2006, 22: e184-e190.
[5] Friedman N, Linial M, Nachman I, Pe'er DJ: Ocjb: Using Bayesian networks to analyze expression data. 2000, 7:601-620.
[6] Bernaola N, Michiels M, Larrañaga P, Bielza C: Learning massive interpretable gene regulatory networks of the human brain by merging Bayesian Networks. BioRxiv 2020. doi: 10.1101/2020.02.05.935007.
[7] Hashimoto RF, Kim S, Shmulevich I, Zhang W, Bittner ML, Dougherty ERJB: Growing genetic regulatory networks from seed genes. 2004, 20:1241-1247.
[8] Auliac C, Frouin V, Gidrol X, d'Alché-Buc FJB: Evolutionary approaches for the reverse-engineering of gene regulatory networks: A study on a biologically realistic dataset. 2008, 9:91.
[9] Kotiang S, Eslami A: A Probabilistic Graphical Model for System-Wide Analysis of Gene Regulatory Networks. *Bioinformatics* 2020. Doi:10.1093/bioinformatics/btaa122.
[10] Saint-Antoine MM, Singh A: Network inference in systems biology: recent developments, challenges, and applications. *Current Opinion in Biotechnology* 2020, 63:89-98.
[11] Spirtes P, Glymour C, Scheines R, Kauffman S, Aimale V, Wimberly F: Constructing Bayesian network models of gene expression networks from microarray data. 2000.

[12] Wan P, Yue Z, Xie Z, Gao Q, Yu M, Yang Z, Huang J: Mechanisms of Radiation Resistance in Deinococcus Radiodurans R1 Revealed by the Reconstruction of Gene Regulatory Network Using Bayesian Network approach. Journal of Proteomics & Bioinformatics 2013: S6: 007.

[13] Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker TJB: Cytoscape 2.8: new features for data integration and network visualization. 2011, 27:431-432.

[14] Barabási A-L, Albert RJs: Emergence of scaling in random networks. 1999, 286:509-512.

[15] Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási A-L: The large-scale organization of metabolic networks. Nature 2000, 407:651-654.