ORIGINAL ARTICLE

Mining of Related Genes with High Efficiency of Phosphorus Utilization Based on Transcriptome Sequencing in Soybean

Bing Wang, Youbin Kong, Hui Du, Wenlong Li, Caiying Zhang, Xihuan Li

Heibei Agricultural University/North China Key Laboratory of Crop Germplasm Resources, Education Ministry of China, Baoding, Hebei, 071001

Abstract: Low phosphorus in soil has become an important limit factor affecting the yield and quality of soybean. The excavation and utilization of high phosphorus efficient related genes is an important prerequisite for the analysis of high pho sphorus mechanism and the improvement of genetic breeding. In this study, the high- and low-efficiency soybean germplasms were used to analyze the root transcriptome data under two different phosphorus conditions through the weight gene co-expression network method.The results showed that there were 15305 high-expressed related genes obtained and were divided into 20 modules, and four of them showed different expressions between these two varieties under two phosphorus treatments. Further analysis results of the Melightcyan module revealed that 268 genes were found in this module, and 13 genes of them were up-regulated with low-phosphorus induction and involved in multiple metabolic pathways. Moreover, the related genes in this module which participate in the phospholipid metabolism pathways showed the most highest expression levels. Finally, combined with the previous reports, six kinds of related genes with high efficient utilization of soybean phosphorus were screened out, which could provide useful candidate genes for the molecular mechanism analysis and breeding improvement in soybean.

Keywords: Soybean; Low-phosphorous Stress; Weighted Gene Co-expression Network Analysis; Transcriptome sequencing; Candidate gene

Introduction

Soybean is an important food and oil crop, not only is the main source of vegetable protein and edible oil, but also has medical and health benefits. It has important application value[1-2]. However, soybean production is often affected by various environmental stresses, of which the lack of available phosphorus is an important factor affecting the yield and quality of 40. Some studies have found that soil phosphorus is mainly absorbed and utilized by plants in the form of phosphate. However, due to the adsorption and fixation of phosphorus by soil, the solubility of phosphorus is low and it is difficult to be directly absorbed and utilized, resulting in soil heredity. Learning phosphorus deficiency[3-5].

Root system is the main organ for plants to obtain water and mineral nutrition. Under low phosphorus stress, plant root system first senses low phosphorus signal and obtains phosphorus source[6-11] to maintain growth and development by changing root system morphology, secreting organic acid, phosphatase, etc. Some studies have found that under low phosphorus stress, root surface area, root volume and root-shoot ratio of wheat are significantly increased[12] and under low phosphorus strip, white lupin excretes organic acid through root row to improve phosphorus absorption efficiency around rhizosphere[13] Therefore, root system plays an important role in phosphorus absorption and utilization of plants. It is of great significance to explore the genes of plant root system responding to low phosphorus stress to improve phosphorus absorption and utilization of plants.

At present, phosphorus efficient gene cloning has been reported in Arabidopsis thaliana, rice, corn, wheat, soybean

doi:10.18063/gse.v4i1.1171

and other plants, such as gmprap14[14], gmpr1[15] ,atpt1[16] and[17] pstol 1. However, due to the complex quantitative character of phosphorus efficient utilization and the large number of genes involved, there is still a lack of genes with definite functions for phosphorus efficient utilization in molecular mechanism analysis and molecular breeding. At present, weighted gene co-expression network analysis (WGCNA) has become an important research method to deal with a large number of different gene expression patterns among multiple samples and is widely used in the medical field. Farber[19] uses WGCNA method to analyze genome-wide association study (GWAS) data and screen out regulatory human beings Key Genes of Bone Mineral Density. Gao, *et al.*[20] used WGCNA Methods to Dig out Candidate Genes Related to Ascorbic Acid Synthesis and Circulation in Tomato Fruit; Wei Dayong *et al.* obtained genes related to glucosinolate content by combining with WGCNA and GWAS methods. Therefore, using WGCNA and WG CNA methods to discover candidate genes for plant root response to low phosphorus stress has become an important method for screening phosphorus efficient genes.

In view of this, this study uses phosphorus efficient specific germplasm Zhonghuang 15, and phosphorus inefficient germplasm Niumao Huang in 2, and species phosphorus. Transcriptome sequencing of root samples was carried out under 60-element treatment, and transcriptome data were analyzed by WGCNA method to screen soybean low phosphorus stress response genes, providing functional genes for analysis of phosphorus efficient molecular mechanism and molecular breeding.

1. Materials and methods

1.1 Soybean materials to be tested

This article uses the phosphorus efficient germplasm Zhonghuang 15, and the phosphorus inefficient germplasm Niumao Huang as materials.

1.2 Planting of soybean materials under test

Selecte plump and uniform seeds of medium yellow 15 and ox hair yellow, and sow in vermiculite; After emergence of seedlings 7 d, root samples were taken as controls (marked as 0 d), and 2 phosphorus treatments were set up, in which KH 2 po 4 was used as phosphorus source (1.0 mmol/L) and phytic phosphorus was used as phosphorus source (1.0 mmol/l) in low phosphorus treatment. During plant growth, nutrient solution was poured every 7 d and samples were taken 7 d, 28 d, 49 d and 70 d after treatment. After quick freezing with liquid nitrogen, the samples are stored in an ultra-low temperature refrigerator until all samples are taken. After the sample is completed, the transcriptome will be sequenced.

1.3 Method for constructing coexpression network of soybean differential gene weight

Using R ,language pack ,WGCNA and software to construct soybean differential gene weight co-expression network, specifically referring to Langfelder ,et al[22] method, the basic process is briefly described as follows: Firstly, assume that each differential gene network obeys scale-free distribution, select appropriate weight values, and calculate the dissimilarity coefficients of different nodes; Then, according to the correlation of expression levels between different genes, gene clustering scores are constructed. And analysis of graphs, and division of organizational modules according to the clustering relationship between different genes, and then merging the modules with similar representation modes according to the similarity of module characteristic values; Finally, according to the relationship between the tissue module and the sample traits, the candidate genes related to the sample traits are mined.

1.4 Enrichment analysis of differentiated genes in soybean issue modules

In order to clarify the function of soybean candidate genes in each organization module, use AmiGO 2 and software (http://amigo. Geneontology.org/ AMIGO/LANDING) carried out GO enrichment and tissue localization analysis on each module, and used KEGG at the same time (https://www.kegg.jp/kegg/kegg2.html) analysis of metabolic pathways involved in different genes.

2. Results and analysis

2.1 Construction of weight co-expression network of soybean low phosphorus pesponse genes

In order to construct the soybean low phosphorus response gene weight co-expression network, this study first treated the obtained 2 and 3 varieties with different phosphorus. The differentially expressed genes under 85 conditions were screened and filtered, and 15,305 highly expressed genes were obtained for network construction after removing low expression genes. On this basis, according to the principle of scale-free network, the software package WGCNA and WG CNA are used to calculate the threshold value discovery (Figure 1). When the threshold value is 7 ,the correlation coefficient between genes reaches a plateau, so the threshold value 7 is taken as the parameter value for constructing the soybean low phosphorus response gene weight co-expression network.

Further, the topological overlap matrix algorithm was used to evaluate the correlation of expression patterns among selected genes, and dynamic pruning method 90 was used to divide clustering modules of differential genes (Figure 2a). The results showed that 41 tissue modules (Figure 2b) were obtained, and the modules continued to be combined and found according to the similarity of expression patterns of differential genes of each module (Figure 2c), Finally, 20 and 20 tissue modules were obtained, of which the blue module contained the largest number of differential genes (6202 and) followed by the brown module (2295 and purple module. The number of genes included was the lowest (68), and 1809 genes were not included in any module.

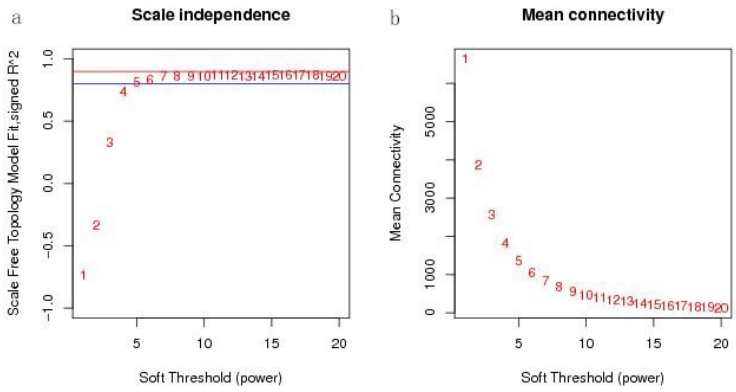

Figure 1. The selection of soft threshold of related genes to low phosphorus. in soybean a: The index figure of scale free network model; b: The figure of average link degree.

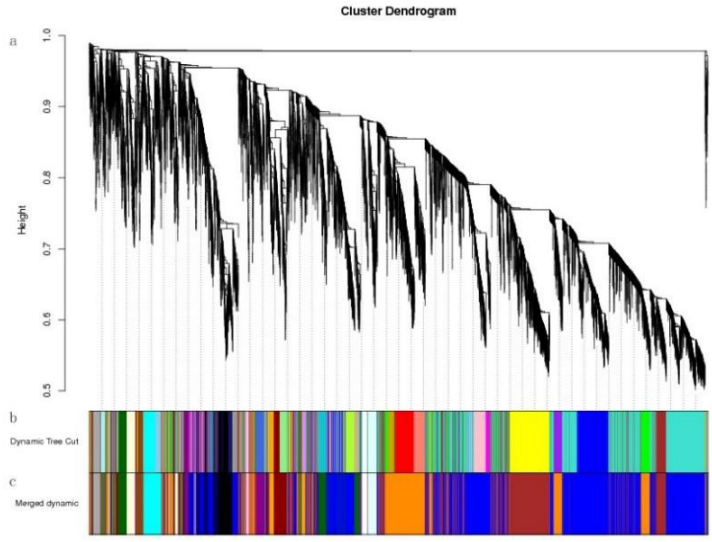

Figure 2. Clustering analysis and module division of related genes to low phosphorus in soybean. a: Clustering analysis; b: Gene modules obtained from the dynamic tree cut method; c: Gene modules after the merged analysis.

2.2 Differential expression analysis of soybean low phosphorus response tissue module in different phosphorus treatment samples

By analyzing the expression patterns of 20 tissue modules under the treatment of 2 phosphorus in the tested variety, the results showed that (Figure 3), there were 8 individual tissue modules have great differences among different

phosphorus treatments, namely brown, dark red and ochre, among the phosphorus efficient varieties Zhonghuang, 15 and 15. Color, Orange, Deep Red, Dark Grey, Light Blue, Deep Orange Module; There are 5 and 2" tissue modules in phosphorus inefficient varieties of cattle hair yellow. There is a big difference between different phosphorus treatments, namely ochre, blue, orange red, light blue and dark orange modules. In step 110, the expression differences of the above-mentioned modules between Zhonghuang 15 and Niumao Huang were comprehensively analyzed. The results showed (Figure 3) that the four tissue modules were significantly different between the two varieties. At the same time, it was found that there were purple acid phosphatases and phospholipids in the differentially expressed blue modules. Transp<rt0>g atpase gene, dark orange module contains transcription factors PHR1, WRKYs, BHLHs involved in phosphorus response process, brown module contains plant expansion protein gene, etc. As the above modules and genes have great differences between different phosphorus treatments and different varieties, they can be used for subsequent screening of genes related to high phosphorus utilization in soybean.

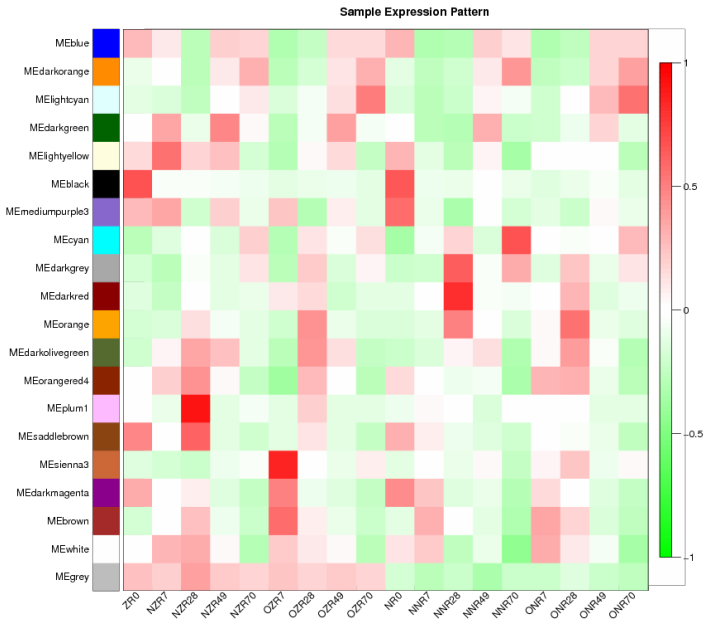

Figure 3. The expressions of related gene modules response to low phosphorus in different soybean samples. ZR: Root of Zhong-huang 15; NZR: Root of Zhong-huang 15 under normal phosphorus condition; OZR: Root of Zhong-huang 15 under low phosphorus condition; NR: Root of Niu-mao-huang; NNR: Root of Niu-mao-huang under normal phosphorus condition; ONR: Root of Niu-mao-huang under low phosphorus condition.

2.3 Functional enrichment analysis of soybean low phosphorus response gene tissue module

In order to further study the biological functions of candidate genes in the above-mentioned tissue modules, the AmiGO 2 software was used to analyze each module. it was found that 20 tissue modules could be enriched to the corresponding GO pathway, and the differentially expressed light blue (Melightcyan) module was further analyzed, which contained 268 genes, including 2 pathways involved in phosphorus stress response.

(GO:0016036) genes, 3 genes involved in the tricarboxylic acid cycle pathway (GO:0006099) and 3 genes involved in the citric acid metabolism pathway (GO:0006101), In addition, it also contains 2 2 genes involved in stress response pathways (GO:0042594) and 3 genes involved in reactive oxygen species response (GO:0009060). the above genes are obviously induced to up-regulate expression under low phosphorus treatment, which indicates that they are closely related to the efficient utilization of phosphorus during soybean growth and development.

At the same time, KEGG and software were used to analyze 268 and candidate genes contained in the tissue module. The results showed (Figure 4) that the above genes participated in 20 and 20 metabolic pathways. Among them, the genes involved in glycerophospholipid metabolism, vitamin B6, metabolism and ether lipid metabolic pathways were up-regulated by low phosphorus induction with high expression. However, genes involved in glycosylphosphatidylinositol biosynthesis, photosynthesis-related protein metabolism, thiamine metabolic pathway and

other genes are down-regulated by low phosphorus induction. “

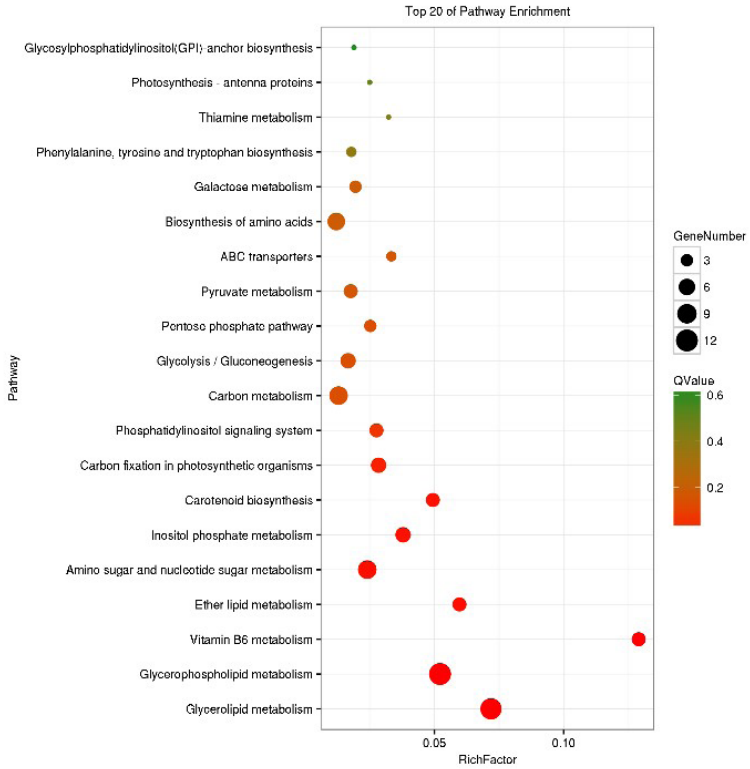

Figure 4. KO pathway enrichments of soybean related genes in Melightcyan module.

2.4 Construction of phospholipid metabolism interaction network of soybean low phosphorus inducing gene

Since the genes involved in glycerophospholipid metabolic pathway in the light blue module are most up-regulated by low phosphorus induction, this study further uses Cytoscape software to carry out network visualization processing on candidate genes of the metabolic pathway and construct gene interaction network diagram. the results show that (Figure 5), LYPLA2, PLDP1, PSS1, PECT1, PAH1 and other genes participate in phospholipid metabolism. In the process, MGD2, DGD2 and other genes participate in the transformation process of phospholipids to glycolipids, SQD1, SQD2 and other genes participate in the transformation process of phospholipids to thiolipids. In addition, candidate genes of SPX and domain are also found in this metabolic pathway, which are reported to be involved in inorganic phosphorus transport and signal transmission in plants.

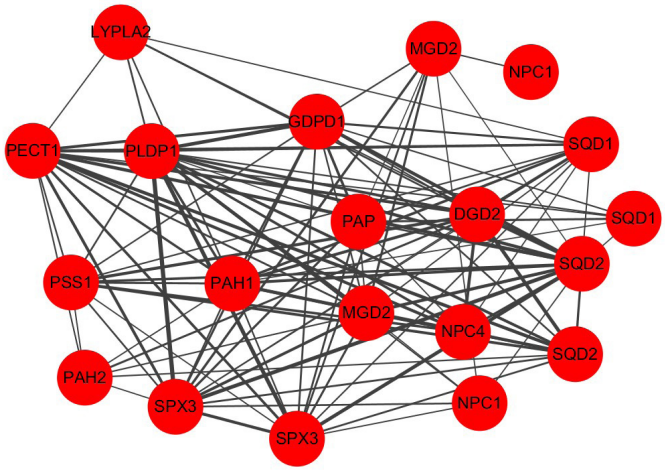

Figure 5. Gene co-expression network of the soybean related genes in Melightcyan module.

2.5 Selection of candidate Genes for soybean response to low phosphorus stress

A comprehensive analysis was made of the differentially expressed genes of the 2 soybean varieties tested

under the 2 phosphorus treatment conditions, combined with the currently reported classification of genes related to efficient utilization of phosphorus in plants. In this study, 6 types of candidate genes responding to low phosphorus stress in soybean were selected, including purple acid phosphatase, phospholipids transport ATP enzyme, transcription factor, extender protein, phospholipid metabolism and transporter gene, etc. At present, the research group is carrying out biological function research on the above-mentioned low phosphorus response genes, with a view to providing functional genes for plant phosphorus efficient molecular breeding and mechanism analysis.

3. Discussion

Transcriptome sequence is an important research method for screening and discover candidate genes for plant traits at present, and has been applied in various plant[23-25]. Currently, there are many reports that transcriptome sequencing combined with genome-wide association study or isolated population linkage analysis has been used to obtain functional genes for important plant traits[26-27]. Co-expression network analysis of weighted genes is currently a common method for studying transcriptome sequencing data. The basic idea of this method is to first divide the differentially expressed genes into significantly related tissue modules by analyzing their correlation, and then find the tissue modules and candidate genes most related to the target traits by analyzing the correlation between each tissue module and a specific phenotype or trait[28].

In this study, the root transcriptome of different soybean varieties under the treatment of two phosphorus was analyzed by the method of weighted gene co-expression network. The data were analyzed to obtain 20 tissue modules, and the expression patterns of differential genes in the 20 modules under different varieties and treatments were further analyzed. The results showed that the expression patterns of candidate genes in each tissue module were different, of which 8 module genes in phosphorus efficient variety and 15 responded to low phosphorus stress, and 5 tissue modules in phosphorus inefficient variety Niumao Huang responded to low phosphorus stress, and the expression patterns of you, 4 and 5 modules are quite different between the two types of varieties, which may be closely related to the efficient utilization of phosphorus in phosphorus-efficient varieties and 15, laying a foundation for subsequent screening and discovery of phosphorus-efficient genes.

In addition, reports have confirmed that phosphorus is not only an important component of compounds in organisms, but also participates in the metabolism of macromolecular substances such as sugars, proteins and lipids (e.g. phospholipids) in plants. In recent years, there have been reports on the function of phospholipids in the efficient utilization of phosphorus in plants[30-31]. In this study, based on the analysis of the differential gene expression of 20 tissue modules obtained, it was found that there are some candidate genes involved in glycerophospholipid metabolic pathway and up-regulated in response to low phosphorus stress in the “Melightcyan” tissue module. furthermore, the interaction network of these genes was visualized and analyzed. the results showed that there are many genes (e.g., LYPLA2, PLDP1, PSS1, PECT1, PAH1, MGD2, DGD2, SQD1, SQD2, SPX3, etc.) It has direct or indirect interaction with glycerophospholipid metabolism and forms a complex interaction network, in which homologous genes of some genes have been reported to participate in phospholipid metabolism processes in model plants such as Arabidopsis thaliana, such as MGD, DGD, SQD and the like[32]. Therefore, the research group is currently conducting in-depth research on the above candidate genes, with a view to providing functional genes for plant phosphorus efficient molecular breeding.

References

1. Liu Ling, Liao Hong, Wang Xiurong, etc. Adaptability of Soybean with Different Root Configurations to Low Phosphorus and Its Relationship with Phosphorus Efficiency [J] Chinese Agricultural Sciences. 2008,41(4):1089-1099.
2. Yao minlei, Zhang jiyao, Zhou Xi, *et al.* analysis of digital gene expression profile of soybean response to low phosphorus stress [J] soybean science, 2016 ,35 (2) :213-221.
3. Xia Xiao, Liu leping, Chen chunfang. effects of phosphorus stress on root morphology of Cinnamomum camphora seedlings [J] Hubei forestry science and technology, 2011 ,6 :31-33.
4. Gu Mian, Chen Aiqun, Xu Guohua. Plant Phosphorus Deficiency and Mycorrhizal Signal Transduction Network [J]

Journal of Nanjing Agricultural University, 2012 ,35 (5) :133-146.
5. Deng Meiju, Wang Fei, Mao Chuanzao. Plant Phosphate Transporter and Its Molecular Regulation Mechanism
6. Xu Jianding .Effects of Soil Water and Nitrogen and Phosphorus Nutrition on Plant Root Growth [J] Shanxi Soil and Water Conservation Science and Technology, 2010, 3: 12-15.
7. Li Wei, Hu Zhizhi, Ni Yu, *et al.* Relationship between Root Water Lift and Drought Tolerance in Alfalfa and Maize-Study on Mechanism of Plant Root Water Lift II [J] Journal of Grassland, 2007 ,15 (6) :515-518.
8. Chiou T J, Lin S I. Signaling network in sensing phosphate availability in plants. [J]. Annual Review of Plant Biology, 2010, 62: 185-206.
9. Li Xihuan, Chang Wensuo, Zhang Caiying. Research Progress on Genes to Improve Phosphorus Nutrition Efficiency (Candidate) in Plants [J] Journal of Plant Genetic Resources, 2012.13(1):83-97.
10. Narang R A, Bruene A, Altmann T. Analysis of phosphate acquisition efficiency in different Arabidopsis accessions[J]. Plant Physiology, 2000, 124: 1786-1799.
11. Dai Kaijie, Where, Guan Huilin, etc. Research Progress on Plants and Low Phosphorus Environment-Induction, Adaptation and Countermeasures [J] .Journal of Ecology, 2006, 25(12):1580-1585.
12. Ma Q F, Renge Z. Phosphorus acquisition and wheat growth are influenced by shoot phosphorus status and soil phosphorus distribution in asplit-root system[J]. Journal of Plant Nutrition and Soil Science, 2008, 171(2): 266-271.
13. Tanaka Min, Wang Chen. Comparison of Organic Acids Secreted by Root Exudation and Non-root Exudation Areas of Phosphorus-deficient White Lupin [J] Journal of Plant Physiology and Molecular Biology. 2000,26(4):317-322.
14. Kong Y B, Li X H, Wang B, *et al.* The soybean purple acid phosphatase GmPAP14 predominantly enhances external phytate utilization in plants.[J]. Frontiers in Plant Science, 2018, 9: 292.
15. Wu B, Li X H, Liu C, *et al.* Genetic transformation and function analysis of transcription factor GmPHR1 in soybean[J]. Soybean Science, 2013, 3 : 302-305.
16. Mulchhal U S, Pardo J M, Raghothama K G. Phosphate transporters from the higher plant Arabidopsis thaliana [J]. Proc Natl Acad Sci,1996, 93: 10519-10523.
17. Gamuyao R, Chin J H, Pariascatanaka J, *et al.* The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency[J].Nature, 2012, 488(7412): 535-539.
18. Wang Tao, Jiang Qinghua, Peng Jiajie, *et al.* Review of Research on Construction and Analysis Methods of Gene Co-expression Networks[J] Intelligent Computers and Applications, 2014, 4(6):47-50+53.
19. Farber C R. Systems-level analysis of genome-wide association data[J]. G3-Genes Genom Genet, 2013, 3(1): 119-129
20. Chao G, Zheng J, Shan L, *et al.* Deciphering ascorbic acid regulatory pathways in ripening tomato fruit using a weighted gene correlation network analysis approach[J]. Journal of Integrative Plant Biology, 2013, 55(11):1080-1091.
21. Wei Dayong, Cui Yixin, Xiong Qing, *et al.* Identification of Candidate Bases for Glucosinolate Content in Rape Seeds by Genome-wide Association Mapping and Co-expression Network Analysis. Cause [J]. Journal of Crops, 2018, 44 (5): 629-641.
22. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis[J]. BMC bioinformatics, 2008, 9(1): 559.
23. Zhu Lin, Yuan Meng, Gao Hongxiu, *et al.* Transcriptome Analysis of Low Temperature Response in Rice Seedling Stage [J] Journal of North China Agriculture, 2018 ,33 (5) :40-51.
24. Qin Wenting, Yang Caiqiong, Yang Feng, *et al.* Transcriptome Analysis of Soybean Seed Aging [J]. Journal of Northeast Agricultural University, 2018 ,49 (3) :1-9.
25. Zhu Jianxiong, Su Hang, Zeng Xiu, *et al.* Analysis of Transcriptome Quality of Maize Flooding Stress [J]. Molecular Plant Breeding, 2017, 15 (2): 532-537.
26. Xian Xiaohua, Wang Jia, Xu Xinfu, *et al.* Integration GWAS and WGCNA Analysis and Excavation of Yellow Seed Microeffective Loci of Brassica napus [J]. Acta Crop Sinica, 2018 ,44 (8) : 1105-1113.
27. Darlington T M, Ehringer M A, Larson C, *et al.* Transcriptome analysis of inbred long sleep and inbred short sleep mice[J]. Genes, Brain and Behavior, 2013, 12(2): 263-274.

28. Ghazalpour A, Doss S, Zhang B, *et al.* Integrating genetic and network analysis to characterize genes related to mouse weight[J]. PLoS Genetics, 2006, 2(8): e130.
29. Hammond J P, Broadley M R, White P J. Genetic responses to phosphorus deficiency[J]. Annals of Botany, 2004, 94(3): 323-332.
30. Li M Y, Welti R, Wang X M. Quantitative profiling of Arabidopsis polar glycerolipids in response to phosphorus starvation. Roles of Phospholipases D zeta 1 and D zeta 2 in phosphatidylcholine hydrolysis and digalactosyldiacylglycerol accumulation in phosphorus-starved plants[J] .Plant Physiology, 2006, 142(2): 750-761.
31. Cruz- Ramirez A, Oropeza- Aburto A, Razo- Hernandez F, *et al.* Phospholipase DZ2 plays an important role in extraplastidic galactolipid biosynthesis and phosphate recycling in Arabidopsis roots[J]. Proceedings of the National Academy of Sciences of the United States of America, 2006, 103(17): 6765-6770
32. Nakamura Y. Plant phospholipid diversity: emerging functions in metabolism and protein-lipid interactions[J]. Trends in Plant Science, 2017, 22(12): 1027-1040.