Bioinformatics analysis of BAN gene in Arabidopsis thaliana

Yongguang Han1,2, Ligang Ma1,2, Le Zhao1,2, Weisheng Feng1,2, Xiaoke Zheng1,2,*

1 School of Pharmacy, Henan University of Chinese Medicine, Zhengzhou, China
2 Collaborative Innovation Center for Respiratory Disease Diagnosis and Treatment & Chinese Medicine Development of Henan Province, Zhengzhou, China

*Corresponding author e-mail: zhengxk.2006@163.com

Abstract. The LAR encoded BAN gen play a key role in flavonoid synthesis. The protein might be hydrophilic and possibly localized in cytoplasm, non-transmembrane region. The protein had a high degree of homology with Capsella rubella, reaching 91%.

1. Introduce
Flavonoid metabolism is an important metabolic process in plants. Proanthocyanidin (PA) pathway is one of the most important ways to metabolize flavonoids [1,2]. The chemical compounds in PA pathway play an important role in plant life activities of the metabolites, such as avoiding ultraviolet damage, protecting plant to prevent the invasion. In addition, this kind of material has antibacterial, antiviral, pathogens, antioxidant activity, antitumor, anti-osteoporosis and other biological activities [3-,5]. The flavonoids also play an important role in agriculture, such as the content of PA in grass is too high to affect the digestion of ruminant animal resulting in ruminant animal abdominal discomfort. Therefore, it is of great significance to study the metabolic pathways of flavonoids and their regulation of biological reactions. BAN(NAD(P)-binding Rossmann-fold superfamily protein) gene encodes leucoanthocyanidin reductase (LAR) [6-9]. LAR can convert leucoanthocyanidins to 2,3-trans-flavan-3-ols and negatively regulate flavonoid biosynthesis. At present, some studies have confirmed that the expression of BAN is positively correlated with the accumulation of condensed tannins, but it has been reported that LAR activity is not consistent with the accumulation of condensed tannins. LAR gene expression and condensed tannin synthesis may also be affected by intraspecific and interspecific differences [10-14]. Therefore, it is of great scientific significance and practical application value to find out the structure, transcriptional function and translation of BAN gene.

2. Materials and Methods

2.1. Data
The BAN gene sequence information of Arabidopsis and the amino acid sequence information of LAR protein are downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/). The gene sequence information of Brassica napus, Capsella rubella, Theobroma cacao, Mangifera indica, Brassica oleracea, Parrya nudicaulis, Eucalyptus grandis, Setaria italica, Triticum urartu, Dichanthelium oligosanthes, Cattleya hybrid cultivar, Tulipa gesneriana, Muscari armeniacum, Curcuma alismatifolia, Lilium regale, Picea sitchensis, Amborella trichopoda and Selaginella moellendorffii were downloaded from the Blast software network. DNAMAN software was used to analyze the number of amino acids encoded by the gene sequence.
2.2. Physicochemical analysis of protein
The obtained amino acid sequence was preserved in text format and subcellular location analysis was carried out by online analysis software Target P 1.1 Server, analyzed the conserved domain related information of the protein by InterPro Scan online software and determined the protein transmembrane region by TMHMM online software. The ExPASy Proteomics Server online software was used to predict the basic physicochemical properties of the protein, including amino acid length, relative molecular mass, isoelectric point, lipophilicity and hydrophilic of protein.

2.3. Analysis of multiple sequence alignment
Software DNAMAN was used to compare multiple sequences. The amino acid sequence of target protein was taken as sample and BLAST online software was used to select amino acid sequences from NCBI gene library. The amino acid sequences of these species, including Arabidopsis, were compared and analyzed by the multi sequence alignment program in DNAMAN software.

2.4. Modeling of three-dimensional structure of protein and prediction of two stage structure
An online software SWISS-MODEL was used to model three-dimensional homology, and then the structure of the protein was boldly conjectured. SPOAM online software was used to analyze the proportion of the two levels of the protein.

2.5. The construction of phylogenetic tree
The MEGA5 software is used to generate the phylogenetic tree. The construction of the evolutionary tree uses Neighbor-Joining (NJ) and the check parameter Bootstrap is repeated 500 times.

3. Results and discussion

3.1. Physicochemical properties of BAN gene encoding protein in Arabidopsis thaliana
The protein encoded by the BAN gene contains 340 amino acid residues. The physical and chemical properties of the protein encoded AtBAN gene were predicted and analyzed by ExPASy Proteomics Server online software Protparam. The molecular formula of LAR is C_{1703}H_{2694}N_{434}O_{515}S_{13} and the relative molecular mass is 37.9kDa and the isoelectric point pI is 5.89. The negative residual base (Asp + Glu) is 43 and the positive residual base (Arg + Lys) is 38. The instability coefficient of the protein is 26.20, which indicates that the protein is more stable. The fat coefficient of protein is 91.21 and the total average hydrophobicity index is -0.176, so it is speculated that the protein encoded by AtBAN is hydrophilic protein.

![AtLAR amino acid sequence](image)

The cDNA coding region sequence of the AtBAN gene predicted by DNAMAN is shown in Figure 2.
3.2. Prediction and analysis tertiary structure of the protein

The tertiary structure of the LAR protein is predicted in SWISS-MODEL online site, as shown in Figure 3. The secondary protein structure was also analyzed by SOPMA software. The content of the alpha helix, beta-pleated sheet beta-turn and random coil is about 39.41%, 10.59%, 22.65% and 27.35%, respectively.

The Interpro Scan online software was also used to analyze the structure of the protein. The results showed that the protein had a NAD (P) binding domain. This indicates that the protein has a core Rossmann type folding NAD- and NADP binding domain, which consists of 3 layers of α/β/α.

Figure 2. Prediction of the sequence of AtBAN coding region of plants

Figure 3. The three-dimensional model of LAR
3.3. **Subcellular localization of AtLAR in Arabidopsis thaliana**
The subcellular location of AtLAR protein was predicted by online software Target P1.1 Server and the prediction result was shown as Table 1. It shows that the AtLAR may be located in the cytoplasm.

**Table 1.** subcellular mapping of AtLAR

| Name | len | cTP | mTP | SP | other | Loc | RC |
|------|-----|-----|-----|----|-------|-----|----|
| Sequence | 340 | 0.062 | 0.173 | 0.154 | 0.750 | - | 3 |
| Cutoff | | 0.000 | 0.000 | 0.000 | 0.000 | |

Notes: (Len represents the length of the sequence. The mTP representative sequence contains the mitochondrial targeting peptide, and the cTP representative sequence contains the chloroplast transport peptide. SP represents a sequence of signal peptides that contain secretory pathways, and other is represented elsewhere in the cell, Loc represents the prediction of the subcellular location based on the score, RC indicated that the reliability level of the prediction was 1 to 5, the smaller the value, the higher the reliability of the prediction.

3.4. **Transmembrane region prediction and conservative region prediction of AtLAR**

![Figure 4. Prediction results of AtLAR protein transmembrane region](image)

Using TMHMM Server v.2.0 to predict the transmembrane region of AtBAN, the prediction results showed that the AtLAR protein had no transmembrane domain. The number of predicted TMH was zero. That indicated the protein had no transmembrane helix.

3.5. **Analysis of multiple alignment between AtBAN amino acid sequences**
The amino acid sequence of AtLAR was analyzed by DNAMAN and the sequence alignment was found in Figure 5. The results show that the amino acid sequence of Capsella rubella was closest to the amino acid sequence of Arabidopsis thaliana, reaching 91%, followed by Brassica napus, reaching 89%.
3.6. *AtLAR* Amino acid sequence alignment and evolutionary analysis

Submitting the amino acid sequence of *AtLAR* to the protein sequence database of NCBI in the National Biotechnology Information Center of the United States for BLASTP search similarity sequence. The results show that the LAR of *Capsella rubella* was the most similar to the LAR sequence of *Arabidopsis*, reaching more than 90%.

**Figure 5.** Multiple sequence alignment

**Figure 6.** Phylogenetic tree analysis of plant LAR

(1. Magnoliophyta. 2. Monocots. 3. dicotyledonous plants. 4. Lycopsida. 5. spermatophyte)
4. Conclusion
The bioinformatics and phylogenetic analysis of The LAR enzyme encoded by AtBAN were carried out. The protein might be hydrophilic and possibly localized in cytoplasm, non-transmembrane region. The protein had a high degree of homology with Capsella rubella, reaching 91%.

Acknowledgments
This work was financially supported by the National Key Basic Research Development Program of China (973 Program) (2013CB531802) and Doctoral Research Fund of Henan University of Traditional Chinese Medicine (Grant No. BSJJ2014-14).

References
[1] GAN Bei Yang hongyu. Metabolic Approach of Flavonoids and Its Regulation in Arabidopsis thaliana[J] Journal of Anhui Agri. Sci. 2008, 36(13):5290-5292.
[2] ZHOU Zhe, ZHANG Cai-Xia, ZHANG Li-Yi, WANG Qiang, LI Wu-Xing, TIAN Yi, CONG Pei-Hua. Bioinformatics and Expression Analysis of the LysM Gene Family in Apple. China Agriculture Science, 2014, 47(13): 2602-2612
[3] Xie D Y, Dixon R A. Role of anthocyanin reductase, encoded by BANYULS in plant flavonoid biosynthesis.[J]. Science, 2003, 299(5605):396.
[4] Devic M, Guilleminot J, Debeaujon I, et al. The BANYULS gene encodes a DFR-like protein and is a marker of early seed coat development.[J]. Plant Journal for Cell & Molecular Biology, 1999, 19(4):387.
[5] Quattrocchio F, Verweij W, Koes R. Flavonoids: a colorful model for the regulation and evolution of biochemical pathways[J]. Trends in Plant Science, 2005, 10(5):236.
[6] Wang XM, Dong J, Jiang JS.et.al. Cloning and Bioinformatic Analysis of MsBAN Gene from "Medicago Sativa" chinese journal of grassland 2012, 34(2):8-15.
[7] CHEN Chun-Yan,MA Hui-Ling,DONG Wen-Ke. Cloning and Expression Analysis of a Leucoanthocyanidin Reductase (LAR) Gene from Onobrychis viciifolia cv.Gansu[J]. Acta Prataculturae Sinica, 2015, 24(6): 177-187.
[8] ZHANG Hao, YOU Shi-Dong, GAO Jing, ZHANG Hai-Li, LI Sheng-Hui, XING Ji-Hong, WANG Feng-Ru, DONG Jin-Gao. PMP Expression Characteristics and Analysis of the Function in Arabidopsis. China Agriculture Science, 2014, 47(15): 3094-3102.
[9] Talbert P B, Adler H T, Parks D W, et al. The REVOLUTA gene is necessary for apical meristem development and for limiting cell divisions in the leaves and stems of Arabidopsis thaliana[J]. Development (Cambridge, England), 1995, 121(9):2723.
[10] Kemp L E, Bond C S, Hunter W N. Structure of 2C-Methyl-D-Erythritol 2,4-Cyclodiphosphate Synthase: An Essential Enzyme for Isoprenoid Biosynthesis and Target for Antimicrobial Drug Development[J]. Proceedings of the National Academy of Sciences, 2002, 99(10):6591.
[11] Chen J, Xiao Y, Di P, et al. Molecular cloning and characterization of a 2C-methyl-d-erythritol 2,4-cyclodiphosphate synthase gene from Cephalotaxus harringtonia[J]. Molecular Biology Reports, 2009, 36(7):1749-1756.
[12] Baima S, Possenti M, Matteucci A, et al. The arabidopsis ATHB-8 HD-zip protein acts as a differentiation-promoting transcription factor of the vascular meristems.[J]. Plant physiology, 2001, 126(2):643-655.
[13] Eshed Y, Feinberg K, Poliak S, et al. Gliomedin Mediates Schwann Cell-Axon Interaction and the Molecular Assembly of the Nodes of Ranvier[J]. Neuron, 2005, 47(2):215-229.
[14] Baudry A, Caboche M, Lepiniec L. TT8 controls its own expression in a feedback regulation involving TTG1 and homologous MYB and bHLH factors, allowing a strong and cell-specific accumulation of flavonoids in Arabidopsis thaliana[J]. Plant Journal for Cell & Molecular Biology, 2006, 46(5):768-79.