Simulation of Differential Expression in Root, Stem and Leaf of Ornamental Cunninghamia Lanceolata Under Internet of Things Environment

Hui Liu(✉)

Heze University, Heze, China
ysh201701@163.com

Abstract. With the promotion of the use of ornamental Cunninghamia lanceolata (Cunninghamia lanceolata L.), the difference of root, stem and leaf specific expression was simulated under the Internet of things environment. In the process of the research, the model was established. The selective expression of genes was carried out and the score output was calculated. Based on the output probability setting of signal states and differential gene identification, the biological function analysis of peculiar genes in roots, stems and leaves of Cunninghamia lanceolata was realized. Through the experimental analysis, the validity of the simulation is proved effectively.

Keywords: Internet of things · Cunninghamia lanceolata · Rhizome and leaf · Difference

1 Introduction

The rapid development of molecular biology in recent years shows the traditional physiology, ecological methods to reveal the resistance of Cunninghamia lanceolata (Cunninghamia lanceolata) information is very limited, and vulnerable to other factors, it is difficult to meet the needs of modern Chinese fir research. Therefore, protein two-dimensional electrophoresis technology emerged as the times require. It can not only reveal the differences in protein and exclude the phenotypic variation caused by environmental differences, but also widen the choice of research materials. Although the cost is large, the method is fast and reliable [1]. The research on proteomics of Cunninghamia lanceolata is relatively few, which is related to the late start of the research on coniferous tree species [2]. Cunninghamia lanceolata is a unique tree species in China. It is also one of the main tree species in the south of China. It is of great importance in the total amount of forest resources in China. For a long time, continuous cultivation of Cunninghamia lanceolata (Cunninghamia lanceolata) resulted in an increasingly obvious decline in the yield of Chinese fir plantation. A large number of studies have shown that phosphorus deficiency is one of the main constraints on plant growth [3]. Low phosphorus tolerance is the nature of a plant widely existing in nature. It has been widely recognized that low phosphorus tolerance of plants and its

G. Gui and L. Yun (Eds.): ADHIP 2019, LNICST 301, pp. 392–400, 2019.
https://doi.org/10.1007/978-3-030-36402-1_42

role in improving plant resistance and uptake of nutrients in ecosystems has been widely recognized. The results showed that phosphorus stress could lead to the decrease of plant dry matter accumulation and root parameters, the change of species and quantity of root specific exudates, and the increase of nutrient accumulation in root. However, most of the current scholars have studied in physiology and ecology, and relatively few in molecular biology. At present, there are few studies on Simulation of rhizome differences. He et al. [4] analyzed the correlations between the varietal parameters of Sirius and to determine the minimum number of parameters that need to be estimated in order to accurately simulate the effect of the genotype and environment on wheat anthesis date. So far, the mechanism of molecular effects of Chinese fir on low phosphorus stress has not been clarified. In order to study the response mechanism of Chinese fir to adversity during its growth and development, especially the mechanism of low phosphorus tolerance, it is particularly important to express differentially from stem to leaf.

2 Simulation of Differential Expression of Ornamental Cunninghamia Lanceolata Root, Stem and Leaf

2.1 Model Establishment

The xylem of Chinese fir roots, stems and leaves is composed of tubular molecules, parenchyma cells and wood fibers. It has been one of the hotspots in plant anatomy, developmental biology and cell biology because of its obvious characteristics in morphology, structure, development and physiological function. These cell types have different functions, all from the cambium meristem primordial cells. The proportion of these cell types is different among different tree species. The structure and composition of stem wood of the same tree also have great variation. This is determined by its internal genetic basis and external environmental conditions. The radial width of the ring is more or less dependent on the species of trees, while some species growing in the tropics may not have obvious rings or may form multiple rings a year, resulting in false rings. In addition to the morphological characteristics of the cell wall, the xylem cells of late wood became longer and lignin components changed (Takabe et al. 1992) [5]. Stress wood can cause secondary growth of trunks and branches in a more favorable direction of (favorable). Along with this brings about the violent development transformation, causes the wood product many aspects also to have the change.

The most important point in designing a generalized hidden Markov model (GHMM), is its topology. Choosing the topology of the model correctly can describe the data more efficiently. Usually, the topology of GHMM is fully connected, that is, every state can be reached from any other state. However, it is obvious that the ergodic topology is not suitable for the poly (A) signal structure derived from the pseudo-Chinese fir mRNA polyadenosine signal model (Fig. 1A).

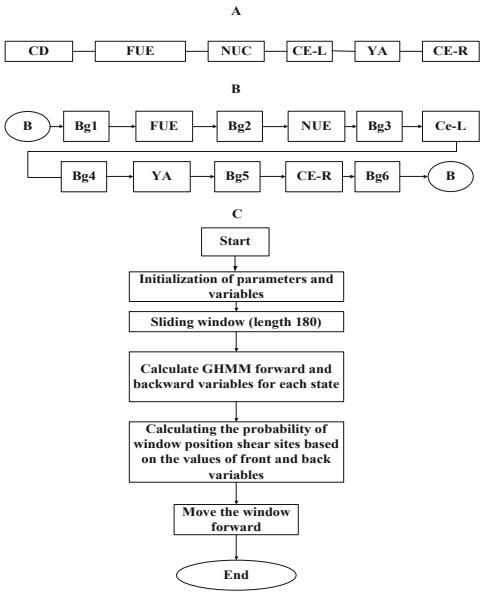

Fig. 1. A signal structure; B topology; C algorithm flow chart

The square represents the fixed-length region, while the brackets represent the variable-length region with a limited range; 3'-UTR represents the 3 'untranslated region; CD represents the coding region; FUE represents the upstream remote control element; NUE represents the upstream near-end control element; CE stands for the shear element, CE.L,CE.R for the left and right side of the poly (A) site, CS for the shear site, that is, the poly (A) site. YA represents the dimer with TA or CA at the left end of CS. It is noted that not all sites are YA, at left end of the site, so other pairs of combinations are also considered in GHMM. In order to simplify the model, it is assumed that the length of each specific signal state is fixed, while the length of the background region outside the signal is variable. At the same time, each state can only be transferred to the next state, but does not exclude the possibility of jumping directly from one signal to another, so the length of each background state in the model is set to possible zero. Finally, it is specified that all paths of the model can only begin in the first state, that is, background 1, and end in the last state IV. The GHMM topology for poly (A) site recognition is shown in figure 3.6B. Among them, B is the starting point of program scanning, Bg is the background state between cis-acting elements, and E is the end point of program scan. According to the design of GHMM model, we construct a computer program to predict the location of poly (A) sites, the core flow of which is shown in figure 3-6C. The main function of the subfunction is to calculate the probability product of the base distribution of the subsequence by the base distribution matrix, and to realize the calculation of the first-order isomeric Markov submodel [6].

2.2 Selective Expression of Genes

The selective expression of genes is very complex, and its influencing factors have both external and internal factors. Now people can make the favorable gene expression of organism by changing the external conditions, inhibit the harmful gene expression, for the benefit of human beings. Treat cancer by placing oncogenes in a state of suppression, such as through specific chemotherapy methods. RNA interference is common in animals, plants and human bodies. It is of great significance for the management of selective expression of genes, the protection of virus infection, the control of active genes and the activation of inhibitory genes. In eukaryotes, most of the genes that contain complex transcription units are genes that encode tissue and development-specific proteins, with the exception of varying numbers of introns. Its original transcription product, pre.rnRNA, can be processed into two or more mRNA, by many different ways, which can encode more than one peptide. This diversity of post-transcriptional processing is regulated by the selective expression of genes. It mainly includes the selection of transcription initiation sites, selective processing and mRNA stability control.

2.3 Score Output Calculation

At present, hidden Markov model (HMM) is a widely used statistical method in bioinformatics, mainly used in linear sequence analysis, model analysis, gene discovery and so on. HMM is developed on the basis of Markov chain [7]. Unlike Markov chain model, HMM is more complex. Its observations and states do not correspond one to one, but are related through a set of probability distributions. As shown in Fig. 1, a hidden Markov model is composed of two random variable sequences, one of which is an unobservable (both hidden) variable state sequence to describe the state transition, and the other is an observable symbol sequence generated by an unobservable state to describe the statistical correspondence between the state and the observed value. For an observer, only the observed value can be seen, but the observer who gets the observed value from which states in the model does not know, which is the so-called "hidden" of the model (Fig. 2).

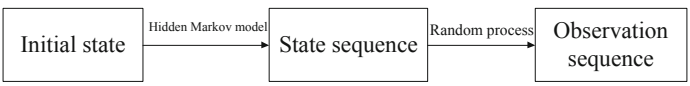

Fig. 2. HMM composition diagram

According to the description of poly (A) signal pattern, the statistical characteristics are mainly concentrated in the nucleotide sequences of upstream 130nt and downstream 50nt of poly (A) site, indicating that this part of the sequence basically reflects the purine around poly (A) site. The pattern of the pyrimidine concentration curve, so we use a 180nt wide sliding window to calculate the output score of the sequence. For an input sequence longer than 180nt, slide along the entire sequence with the step size

of Int until the end of the sequence [8]. In this way, for each base position, according to the previously designed algorithm, a score can be calculated in each window containing that base (Score, is the score of the poly (A) site in a sequence of windows). Take the maximum score as the last score of the poly (A) site at this location, described as follows:

$$\begin{aligned} S\text{core(t)} &= \max\{S(\text{t})\} \\ S(\text{t}) &= \{\log_{10} PS_{\text{w}} + 120\}/2 \end{aligned} \tag{1}$$

$$PS_{\text{w}} = \text{a}_t(8) \cdot \beta_t(9) \tag{2}$$

According to the definition of forward and backward variables, $\text{a}_{\text{t}}(8)$ represents the probability that the observed sequence $O1, O2, \cdots, O\text{t}$ will appear when the eighth state of the model (i.e., YA on the left side of poly (A) site ends at the base position t, $\beta_t(9)$ represents the ninth state in the model (that is, Bg5 on the right side of poly (A). W refers to all window sequences containing base position t. Log values and two constants (120 and 2) are used to adjust the score so that the final output score is in a controllable range.

2.4 Output Probability Setting of Signal Status

After determining the length of the signal, we need to further calculate the output probability of the bases (A, T, C, G) in each signal state (the probability distribution of the observed symbols in the state S_{i}). To this end, the number of signal modes obtained by SignaSleuth from 8 K data sets is used to analyze the base distribution of FUE, CE-L and CE-R signal states (i.e. the base-to-ε output probability $D_\varepsilon = \varepsilon \in \{A, T, C, G\}$), and the calculation formula is as follows:

$$D_\varepsilon = \frac{\sum_{\text{i}=1}^{N} (\varepsilon_{\text{i}} \times W_{\text{i}})}{\sum_{\varepsilon \in \{A,T,C,G\}} (\sum_{\text{i}=1}^{N} (\varepsilon_{\text{i}} \times W_{\text{i}}))} \tag{3}$$

Among them, $W_{\text{i}} = \frac{C_{\text{i}}}{\sum_{\text{i}=1}^{N} C_{\text{i}}}$ is the statistical weight of mode i, and the more the number of occurrences, the greater the weight of mode i. C_{i} is the number of occurrences of mode i in the signal region. ε_{i} is the number of occurrences of base i in mode ε, $1 \leq \text{i} \leq \text{N}$, and IV is the number of signal modes considered.

2.5 Differential Gene Identification

In order to identify the structural differences between WT and pcfs4, the gene (DPG), is based on biological significance. The top of the figure is the structure of the given gene, and the box represents the exon (exon), joining two exon lines representing the short lines below the intron (intron), exon representing the TilingArray probe [9]. The

abundance of Exon l and Exon 2 is the same between WT and pcfs4, but Exon 3 is different. In this way, the ratio of Exon-1 between IV T and pcfs4 is the same as that of Exon-2, which is equal to 1, which is different from the ratio of Exon-3. On the contrary, if the selection utilization of poly (A) locus is not affected by PCFS4, the relative abundance of the two transcripts will be the same in WT and pcfs4, that is, the ratio of measured abundance between wT and pcfs4 will be equal in all three exon.

For a given unknown gene, it is assumed that two major transcripts (also known as transcripts) are derived from its pre.mRNA by selecting two different poly (A) sites for splicing. As described above, the selective polyadenylate of the gene FCA is the same.

Based on the above concepts, hierarchical combination statistical analysis was used to identify differentially structured genes (DPG) and differentially expressed genes (DEG) regulated by PCFS4 factors by using logarithmic converted ratio data to identify differentially structured genes (DPG) and differentially expressed genes (DEG). For each of the selected differentially structured genes, we also use the multiple mean comparison (MMC) method to analyze the differences in all pairings of the gene or transcript. And then determine which or which exorls units cause partial differences in the corresponding genes or transcripts. In order to improve the specificity of the candidate differentially expressed gene (DEG), the double-sided T-test was further applied to the entire transcript of the candidate gene, zero pseudomorphism: all probe ratios of the entire transcript. The average value equals 0. If the candidate gene is considered not to be a differentially expressed gene, the candidate gene is classified as a differentially expressed gene.

2.6 Special Allogeneic Biological Function in Root, Stem and Leaf of Cedar

The number of genes specifically expressed in leaves was the largest, with 20 genes, but mainly concentrated in auxin signal transduction pathway, including 15 genes encoding SAUR protein (9 indole-3-acetic acid inducible protein ARG7 and 6 auxin-induced protein). 2 GH3 (gretchen hagen3) and 1 gene encoding AUX/IAA protein (IAA29) In addition, jasmonic acid pathway transcriptional activator MYC2 (BbHLH28) and ethylene signal transduction activator EIN3 were also specifically expressed in leaves.

3 Methodology

3.1 Test Material

The seeds of Cunninghamia lanceolata were collected from Jiaozhou Experimental Station of Qingdao Agricultural University. After laminating, the seedlings were raised in spring 2016. After 45 days of seedling growth, individual plants with the same growth and development were selected and their roots, stems and leaves were selected for specific gene expression analysis. The leaves of ‘QN-101′4’ at different developmental stages, including sprouting buds (shoot tips), early leaf development (young

leaves), middle developmental (mature leaves) and late developmental (senescent leaves), were used for qRT-PCR analysis.

3.2 Data Detection and Functional Annotation

The genes obtained from mapping were compared with NR (NCBI non-redundant protein library), Swiss-Prot (non-redundant protein database), Gene Ontology database and KEGG pathway database) by Blastx (E < 1e−5), and the gene annotation information was obtained. P-value < 0.05 was used as the threshold to define the metabolic pathways of significant enrichment in genes. Using iTAK software (Yi et al., 2016), using classification in database The defined transcription factor (transcription factor, TF) family was identified by hmmscan [10].

3.3 Fluorescence Quantitative PCR Verification

PrimeScriptTM reagent Kit with gDNA Eraser kit was used to reverse transcript mRNA into cDNA. The expression of differentially expressed genes was verified by Real-time Quantitive PCR. Pbactin is selected as the internal reference gene. The differential gene and the specific primer sequence of the internal reference gene are listed in Table 1:

Table 1. Real-time fluorescence quantitative PCR gene primer sequence.

Name	Gane ID	Forward primer	Reverse primer
BbHLH 28	LOC103964261	AAGACGCATCACCCAACAAC	AAAGCTTTGCCGGGAACATT
BAK 1	LOC103929041	CGGAAGGAGATGCTCTGAGT	TTCCGGAAAGGTCCAAGCT
ARG7	LOC103957880	CGTCTGCCTGCTGTAATTCC	TGTTAGACCGCCCATTGGAT
ARG7	LOC103957860	TTGCAGTCTATGTCGGGGAG	GATCGGTTCGTGTTCACTCA
Auxin-induced protein X10A	LOC103967982	GGGTTCCGTCTTCCATCTGT	TCTTGTCTGCAGGGAATCGT
Auxin-induced protein X10A	LOC103957857	TCAGTGTCCCTTCTATGCCC	GAGATGCACTTGGACAACGG
Squalene monooxygenase	LOC104967056	ACCTCAGATATCGCCGGAAG	TTTGAAAAGCACCCATCGCA
Pbactin	GQ339778.1	ATTGGAGCTGAGAGATTCCGGT	GTCTCATGAATGCCAGCAGCTT

3.4 Analysis of Gene Expression in Different Tissues of Chinese Fir

By comparing with the reference genome, 19 443, 19 567 and 17 876 genes were obtained in roots, stems and leaves of Cunninghamia lanceolata (Cunninghamia lanceolata). There were 845894841 low-expression genes (FPKM < 10) in root, stem and leaf respectively. The number of high-expression tissue-specific genes was lower (26, 12), as shown in Fig. 3.

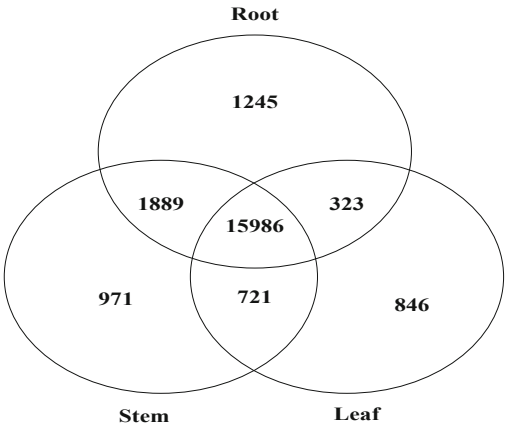

Fig. 3. Analysis of differentially expressed genes in rhizomes and leaves of Cunninghamia lanceolata

These results suggest that there are significant differences in gene expression among different tissues of Cunninghamia lanceolata, and the expression of a large number of tissue-specific genes may be related to the different biological functions of different tissues.

4 Conclusion

It was found that the expression of genes related to hormone signal transduction pathway was mainly enriched in leaves, especially auxin signal transduction pathway. Auxin regulates plant cell division and differentiation, promotes stem and leaf growth, vascular tissue differentiation and apical dominance. The genes associated with auxin signal transduction pathway are mainly expressed in leaves, which may be related to the primary site of auxin synthesis in young leaves.

References

1. Guangxin, Yu., Shuangqing, Z., Xiuhong, Z., et al.: CDNA-SRAP analysis of differential expression in stem and leaf of Aegilops alba. J. Henan Agricultural University **51**(4), 481–486 (2017)
2. Zhang, Y., Jiang, T., Ye, Y., et al.: Cloning and expression specific analysis of strawberry FaCOP1. Horticulture **44**(3), 547–556 (2017)
3. Yan, S.Y., Zhang, Y., Wang, X., et al.: Design of orderly Harvester for stems and leaves vegetables based on Pro/E. Agric. Mechanization Res. **39**(3), 139–143 (2017)
4. He, J., Gouis, J.L., Stratonovitch, P., et al.: Simulation of environmental and genotypic variations of final leaf number and anthesis date for wheat. Eur. J. Agron. **42**(1), 22–33 (2012)

5. Yan, B., Zhu, S., Su, S., et al.: Study on UPLC fingerprinting and chemical pattern recognition of stems and leaves of Scutellaria baicalensis from different areas. J. Nanjing Univ. Tradit. Chin. Med. (6), 24–26 (2017)
6. Chuang, H.S., Wang, Y., Jian, Y., et al.: Study on the finite element modeling method for sugarcane stalk-leaf system. Agric. Mechanization Res. **40**(6), 19–23 (2018)
7. Le, W., Yuan, N., Zhiqiang, C., et al.: Simulation study on the dynamics of cotton leaf age. Jiangsu Agric. Sci. **46**(4), 68–73 (2018)
8. Jing, D., Lu, X., Zhang, K., et al.: Analysis of SNP and allele-specific expression of Gordane hybrid and its parent transcripts. J. Crops **44**(12), 71–79 (2018)
9. Hu, S., Chung, K., Yang, J., et al.: Study on the content difference of flavonoids in stems and leaves of peppermint at different harvest date. Chin. J. Tradit. Chin. Med. **3**, 544–550 (2018)
10. Huan, Z., Yingjie, Y., Dingli, L., et al.: RNA-Seq analysis of the genes specifically expressed in root stem and leaf of Duli L. J. Hortic. **45**(10), 18–31 (2018)