Cloning and Sequence Analysis of a H2B Gene from Aloe
Jiangchuan Du1,2, Lingting Guan1,2, Xiaohong Liu1,2*
1Key Laboratory of Southwest China Wildlife Resource Conservation (Ministry of Education), China West Normal University, Nanchong City, Sichuan Province, China
2College of Life Science, China West Normal University, Nanchong City, Sichuan Province, China

Abstract: Aloe is a perennial succulent plant, widely applied in landscaping and medicine fields. In this study, a clone was randomly selected from an aloe full-length cDNA library, and used for sequence determination and bioinformatics analysis. The results showed that it contained a full-length H2B cDNA gene sequence, the histone H2B gene is 711 bp in length and harboring a 471-bp open reading frame, encoding 156 amino acids. Furthermore, the cloned H2B gene was analyzed by bioinformatics methods, including characteristic, spatial structure and functional sites. The results are beneficial for understanding the sequence and function of H2B gene in other plants.

Keywords: aloe, H2B, gene, bioinformatics analysis

INTRODUCTION
Histone is the important structural component informing nucleosome together with genomic DNA in a eukaryotic cell, related to gene regulation [1, 2]. There are five major families for histone, including H1, H2A, H2B, H3, and H4. H2A, H2B, H3 and H4 are as the core histones, generally, each of them show dimer. Then, the four distinct dimers are gathered into one octameric nucleosome core. H1 protein binds to the "linker DNA" (approximately 20-80 bp in length) region between nucleosomes, and help to stabilize the chromatin fiber [3]. Finally, 146 bp or so of DNA wrap around this core particle, forming into nucleosome.

According to previous literature, histone H2B has many functions. For example, Histone H2B monoubiquitination can facilitate the rapid modulation of gene expression during Arabidopsis thaliana photo morphogenesis [4], repress floral transition [5], and regulate the dynamics of microtubules during the defense response to Verticillium dahliae toxins in Arabidopsis thaliana [6]. Besides, it play an important role in DNA replication and repairing, transcriptional regulation as well as meiosis [4, 7, 8].

Aloe is a succulent perennial herbaceous plant, belongs to the lily family, growing in dry-heat environment. It has many functions, such as resistance to bacteria and virus, eliminating inflammation, enhancing immunity, reducing the blood sugar, reducing blood lipid as well as repressing hepatitis and tumor. Additionally, it is often used to treat burns, headache, gonorrhea, constipation, etc. [9, 10]. The major medicinal components of aloe are aloemodin, aloeresin, aloesin and 2,5-dimethyl-8-C-β-D-glucopyranosyl-7-hydroxy-chromone [11]. Furthermore, aloe an ornamental plant, can eliminate indoor formaldehyde, speed up skin's metabolism, benefit our health by eating. However at present, aloe products can’t meet market demands, due to its wide use and slow growth.

From previous studies on aloe, most of researchers focused on analyzing its chemical composition, pharmacological action of functional components, plant tissue culture and plant physiology. Whereas in genetics and molecular biology, only few studies were reported. To date, H2B gene of aloe has not been researched according to published literature. Therefore, a cDNA clone containing H2B gene was here studied by sequencing and bioinformatics analysis.

MATERIALS AND METHODS
Plant material
In this experiment, the aloe material is Aloe chinensis Baker, provided by our laboratory. Its leaves were used for RNA isolation and full-length cDNA library construction.

Sequencing of H2B gene
The clone containing H2B gene (known after sequence analysis), derived from an established full-length cDNA library by SMART (switching mechanism at 5’ end of RNA transcript) technology, was randomly selected and used for sequencing.
Bioinformatics analysis of H2B gene

For the sequence of H2B gene, GenScan software and DNAMAN 6.0 analyzed it, and its nucleotide sequence and deduced amino acid sequence were compared with other sequences through database search using online bioinformatics tools (available at www.ncbi.nlm.nih.gov/orffinder/). Afterwards, physical characteristic of H2B protein were analyzed using ExPASy tool (web.expasy.org/compute_pi/), and the spatial structure of H2B protein was predicted by SWISS-MODEL software (swissmodel.expasy.org/).

RESULTS AND ANALYSIS
Sequencing result and ORF analysis

The sequencing result of selected clone was shown in Figure 1. Its length is 711 bp, with 208 A (29.3%), 196 C (27.6%), 175 G (24.6%) and 132 T (18.6%); its molecular weight (kDa) is 219.45 and 438.35 for ssDNA and dsDNA, respectively.

The obtained gene sequence was further deduced its ORF (open reading frame) sequence, and the results revealed that it has two ORFs, of which the longer is 471 bp in length, and encodes 156 amino acids (Figure 2). The deduced amino acids sequence was further compared with other sequences in GeneBank databases, and it is highly homologous with the H2B sequences from other species (Figure 3), with identity up to 87.97%, 87.18%, 87.17%, 87.17% and 86.79% with the reported H2B protein in GeneBank from *Phoenix dactylifera* (XP_008795023.1), *Dendrobium catenatum* (XP_020679983.1), *Arabidopsis thaliana* (XP_020571758.1), *Dendrobium catenatum* (XP_020679983.1), *Phalaenopsis equestris* (XP_020571758.1), *Arabidopsis thaliana* (XP_020274442.1). Therefore, the cloned gene here from aloe should be a H2B gene. Further analysis displayed that the molecular formula and molecular weight of the protein encoded by the cloned H2B gene are C_{760}H_{1277}N_{211}O_{223}S_{3} and molecular weight 17034.96, respectively.

![Fig-1: The sequence of the selected clone from full-length library in aloe](image)

![Fig-2: The deduced amino acids of the cloned gene from aloe](image)
Fig-3: Comparison of the amino acid sequence of H2B protein among different plants. Identical amino acids are indicated by blue background

Bioinformatics analysis
Physical characteristic prediction

Primary structure analysis for the protein encoded by the cloned H2B gene revealed that its theoretical pI is 9.97; aliphatic index and grand average of hydropathicity are 66.47 and -0.821, respectively. Another notable feature of the protein is that the number of negatively charged residues is 19, total number of positively charged residues is 38, and the instability index (II) is 22.78. Therefore, we can predict that the H2B protein is a stable and alkaline protein.

The H2B protein encoded by the gene cloned here was predicted for its hydrophobicity, and the results was as shown in Figure 4, the x-axis represents 156 amino acids of the H2B protein and the y-axis represents hydrophobic fraction (positive value as hydrophobicity and negative value as hydrophilicity). According the figure, it was found that both the 22nd and the 58th have the minimum value -3.467, suggesting that the two sites have the highest hydrophilicity, while the 94th has a highest value, up to 1.411, demonstrating that the site possesses the highest hydrophobicity.

Fig-4: Hydrophilicity of H2B protein encoded by the H2B gene cloned from aloe
Spatial structure prediction

The secondary structure of H2B protein encoded by the H2B gene cloned here was predicted, and the results showed that the percentages of alpha helix, extended strand and random coil in the secondary structure were 55.77%, 7.69%, and 36.54%, respectively (Figure 5).

Furthermore, the H2B protein encoded by the H2B gene from aloe was predicted their tertiary structure together with other five H2B protein, including Phoenix dactylifera (XP_008795023.1), Phalaenopsis equestris (XP_020571758.1), Dendrobium catenatum (XP_020679983.1), Arabidopsis thaliana (NP_200799.1), Asparagus officinalis (XP_020274442.1) and Aloe chinensis Baker, and the results was displayed in Figure 6. Comparing the 3D model map, it was found that Aloe chinensis Baker is highly similar with Asparagus officinalis in the tertiary structure, but different from Phoenix dactylifera, Phalaenopsis equestris, Dendrobium catenatum and Arabidopsis thaliana.

Functional sites

After analyzing the functional sites of the amino acid sequences encoded by H2B genes, it was found that there are one cAMP- and cGMP-dependent protein kinase phosphorylation site, four Protein kinase C phosphorylation sites, one Casein kinase II phosphorylation site, one N-myristoylation site, two Amidation site, five Microbodies C-terminal targeting signal and one Histone H2B signature in this encoded protein (Figure 7.)
CONCLUSIONS
In this study, a clone from an aloe cDNA library was selected and sequenced, and the result of comparing with other sequences in GeneBank showed that the obtained gene in this experiment is a H2B histone gene. Moreover, its physical characteristics and spatial structure were analyzed in detail by informatics means. The cloned H2B gene from aloe can be further studied on its biological function, and this work is in progress. Simultaneously, the analyzing methods for the cloned gene here are beneficial for other genes isolated from other species.

ACKNOWLEDGMENT
The authors would like to thank the Scientific Research Fund of Sichuan Provincial Education Department of China (13ZA0012) and the Talent Fund of China West Normal University (17YC338).

REFERENCES