

CLONING AND BIOINFORMATIC ANALYSIS OF GENE ENCODING
WD40-REPEAT PROTEIN IN ARABIDOPSIS THALIANA

HY Duan, CE Zhou, JL Zhao and YQ Zhou
College of Life Science, Henan Normal University, Xinxiang, China

ABSTRACT: In this study, one gene encoding WD40-repeat protein in *Arabidopsis thaliana* was isolated by RT-PCR, and there are four glycosylation sites in this putative WD40-repeat protein, but the transmembrane domain is not found. Compared with other WD40-repeat proteins in *Arabidopsis thaliana*, this putative WD40-repeat protein shows higher similarity with WD40 domain-containing protein (accession number np-192182) and is 58.2%, however the similarity with other WD40-repeat proteins is approximately 27%-30%. Besides, the comparison with WD40-repeat proteins in 40 species indicated that this putative WD40-repeat protein in *Arabidopsis thaliana* has obvious similarity with WD40-repeat proteins in Dicotyledon and Monocotyledon, yet shows lower similarity with WD40-repeat protein in Mermithidae, and the function and mechanism of this putative WD40-repeat protein in *Arabidopsis thaliana* need to be further studied.

KEYWORDS: *Arabidopsis thaliana*, WD40-repeat protein, bioinformatics analysis

INTRODUCTION

WD40 motif is also called Trp-Asp or WD40, composed of approximately 40-60 amino acid residues which typically contains conserved GH (glycine-histidine) and WD (tryptophan-aspartic acid) dipeptide sequences at the N-terminal or C-terminal region, respectively, and a conserved core in which only a few of amino acid residues are highly conserved (Neer *et al.*, 1994; Smith *et al.*, 1999). Since WD40 motif was firstly found in β subunit of G protein (Fong *et al.*, 1986), WD40 motif has been found in a lot of regulatory proteins from mammal, drosophila, yeast and plant, furthermore genome analysis revealed that WD40 motif generally exists in 1%-2% proteins of eukaryote, but very few in prokaryotes (Madrona and Wilson, 2004), further analysis showed that WD40 motif of WD-repeat protein acts as bracket of protein-protein interaction and can participate in several interactions between proteins at the same time, and then could regulate and control assembling of many protein complexes (Smith *et al.*, 1999). In addition to conserved WD40 motif, other domains which could facilitate interaction between proteins are also discovered in some WD-repeat proteins, such as ring finger zinc binding domain (RING), coil-coil region (COIL), LisH motif, etc. (Neer *et al.*, 1994; Madrona and Wilson, 2004; Duan *et al.*, 2007). Therefore, members of WD40-repeat protein family have various functions, could participate in many kinds of biological mechanism, such as signal transduction, cell cycle, RNA processing,

chromatin assembly, transcription, and so on (Madrona and Wilson, 2004; Duan *et al.*, 2007; Duan and Ding, 2007).

At present, WD40 motif has been poorly studied in plants and prokaryotes, but the available research revealed that WD40-repeat protein could play a crucially important role in the development of floral organs, regulation of photoperiod, formation of meristem and others (Deng *et al.*, 1991; Henderson *et al.*, 2005). Moreover, although most WD40-repeat proteins were extremely conserved in eukaryote, there are also significant differences among their functions (Steven and Philip, 2003), so function and regulatory mechanism of most WD40-repeat proteins are not still clear. In this research, one gene encoding WD40-repeat protein was cloned from *Arabidopsis thaliana* by RT-PCR method, the characteristics, structure and function of this gene and its encoding product were analyzed and predicted by a series of bioinformatics software, which would not only reveal the regulatory mechanism of WD40-repeat protein in plant development, but also conduce to understand deep interaction mechanism of WD40-repeat protein and other proteins.

MATERIALS AND METHODS

2.1. Plant materials and growth conditions

Seeds of wild-type *Arabidopsis thaliana* (Colombia type) were soaked in sterile water for 30min, surface-sterilized for 30s with 75% ethanol, then sterilized with 5% sodium hypochlorite for 10min, and washed several

times. Subsequently, after vernalization for 3d in the dark at 4°C, seeds were sown on MS solid medium and cultured at 22°C/18°C with a 16h light and 8h dark photoperiod. Seedlings of *Arabidopsis thaliana* with two leaves were transplanted in vermiculite and nutritional soil, and cultured at the above culture condition. Two weeks later, seedlings were harvested, quickly immersed in liquid nitrogen, and then stored at -80°C for RNA extraction.

2.2. Total RNA extraction

Total RNA was extracted from *Arabidopsis thaliana* seedlings with total RNA isolation system (Promega Corporation, America). The yield and purity of total RNA were determined by spectrophotometer at 260nm, and the integrity of total RNA was determined by denaturing agarose gel electrophoresis. Total RNA was solvated into water without RNAase, and then stored at -80°C.

2.3. Cloning and sequencing of target gene

RT-PCR primers were designed according to the conserved sequence of gene encoding WD40-repeat protein, are AF5'-CTCTCCTTTAGTTAGACAAAGCC-3', AR5'-CAATTTTCAATCCTGAAACTC-3', respectively, and synthesized commercially at GeneCore BioTechnologies Co. Ltd. (Sangon, Shang-hai, China). One microgram total RNA of *Arabidopsis thaliana* seedlings was used for RT-PCR with One Step RNA PCR Kit (AMV) (TaKaRa biotechnology Co., Ltd. Japan), and RT-PCR products were separated by 1.2% agarose gel electrophoresis. After RT-PCR products were retrieved by freeze-thaw method, cDNA fragment was ligated into pMD18-T Vector, and then was sequenced. Sequencing of cDNA fragment was performed by GeneCore BioTechnologies Co. Ltd. (Sangon, Shanghai, China) using ABI373A automatic sequencer.

2.4. Bioinformatics analysis of data

DNA sequence data were analyzed by DNAMAN software and the nucleotide and amino acid sequences were compared in NCBI (National Center for Biotechnology Information) database using BLAST analysis sever. The transmembrane domain of amino acid sequences was analyzed in <http://www.expasy.org/tools>, furthermore prediction of glycosylation site, phosphorylation site and others were carried out by scanProsite (<http://cn.expasy.org/prosite/>). The similarity among some WD40-repeat proteins from *Arabidopsis thaliana* was analyzed with DNAMAN software, furthermore the homologous gene encoding WD40-repeat

protein was respectively searched and screened from some species by Blast, and the phylogenetic tree was constructed in order to explore genetic evolution of WD40-repeat genes by DNAMAN software.

RESULTS

3.1. Cloning of target gene encoding WD40-repeat protein

Approximately one cDNA fragments with 1300bp was isolated from *Arabidopsis thaliana* by RT-PCR and was sequenced. Sequence analysis revealed that this cDNA fragment contains a single open reading frame, encodes WD40-repeat protein and is same to gene with accession number np_190535 from *Arabidopsis thaliana* by BLAST. The total cDNA sequence with 1388bp and its predicted amino acid sequence with 317 amino acid residues were shown in Figure (1).

```

1 CTCTCCTTTAGTTAGACAAAGCCCTCTCTCCAAAACCCATAAATCAGCCAAATTCAAAT
61 TCCGCTAACCAATTTCCATGGCAGAGGAAATCCAGCAACAGCATCATTCCCTCCATAC
      M A E E I P A T A S F T P Y
121 GTCCATTCACAAAATCTAACATCTCACAATGGCGCGTCTCTCCGGTAAATTCCTCC
      V H S Q T L T S H N R A V S S V K F S S
181 GATGGACGCTCTCTCCGATCCGCTCCGCGGACAAACAAATCCGCACTTACACAATCAAC
      D G R L L A S A S A D K T I R T Y T I N
241 ACCATTACAGATCCAAATGGCGGACGCGCTACAGAAATTCACGGCCATGAAAACGGTATC
      T I N D P I A E P V Q E E F T G H E N G I
301 TCCGATGTTCATCTCTCAGACGGGAGGTTTATAGTCTCAGCTCCGATGACAAAACC
      S D V A F S S D A R F I V S A S D D K T
361 TTAAGCATGGGACGTTGAAACTGGTTCATTGATCAAGACGCTTATTGGACACACTAAT
      L K L W D V E T G S L I K T L I G H T N
421 TACGCTCTGTGTCAATTCATCTCAATCAATCAATATGATGATCTGGTTCGTTGAT
      Y A F C V N F N P Q S N M I V S G S F D
481 GAAACTGTGGATCTGGGATGTCACACTCTGGAAAGTGTGTTGAAAGTTCCTCCCGGCAT
      E T V R I W D V T T G K C L K V L P A H
541 TCTGATCCTGTTACTGCTGTGATTTAATAGAGATGGGTCTCTCATTGTTCCAGTAGC
      S D P V T A V D F N R D G S L I V S S S
601 TATGATGGGTTGTGTGATATATGGGATCTGGGACTGGTCAATGTTGTGAAAACCTGATT
      Y D G L C R I W D S G T G H C V E T L I
661 GATGATGAGAATCCTCTGTTGTTGTTAGATTCTCTCCTAATGGCAAGTTTATCTCTC
      D E N P P V S F V R F S P N G K F I L
721 GTGGTACACTTGATAACACCGCTGAGGTTGTGGAACATTTGCTGTCTAAGTTCCTCAAA
      V G T L D N T L R L W N I S S A K F L K
781 ACATACACTGGCCAGTGAAGCAGACAGTATTCGATTCCTCTCCGTTCTCCGTCACAAAT
      T Y T G H V N A Q Y C I S S A F S V T N
841 GGAAGCGAATAGTCACTGATCGGATCCGAGGACAACTGTGTACACATGTGGGAGTAACTCC
      G K R I V S G S E D N C V H M W E L N S
901 AAGAACTGTACAGAAACTTGAGGGTCACTACTGAGCGCTCATGAACGTAGCATGCCAC
      K K L L Q K L E G H T E T V M N V A C H
961 CCGACAGAGAACTTGTGATCGCATCAGGCTCGCTCGACAGACAGTAAAGGATTTGGACACAG
      P T E N L I A S G S L D K T V R I W T Q
1021 AAGAAGAATAAATCATCACTCTCTCTCTCCAGAAATCAACATGATGAGTACATGCCACC
      K K E *
1081 TTCTTCATGCAGTAATCATCTGTTGCTTACTAAGCTTTGATGATTGATTCAGCCTCAIT
1141 ATGTAAAAGTCTAATTTTGTGTTGGTGAAGGGAAGTGAACAGACTGTTCTAACTGTT
1201 AATAGTTCCTGTGTACTCAGAACTATTAGGTAGGCTTAAAGGCAAGACTCTCCGCA
1261 GTGCAAAAGCAATTTGAGAGAATGAACACAAATCTTCATAGGTTTACCCCTTACTCAITG
1321 ATCTATTAAGCTGATGCAATGCTTACCAATCACTATTGGTTATAAAGTATGTTCCAGGAT
1381 GAAAAATTG

```

Figure 1: Full-length cDNA and putative amino acid sequences of target gene.

In addition, the putative amino acid sequence of target gene (np-190535) and other WD40-repeat proteins in *Arabidopsis thaliana* were compared by DNAMAN software, as shown in Figure (2), this putative WD40-repeat protein of

target gene shows higher similarity with transducin/WD40 domain-containing protein (accession number np-192182) and is 58.2%, however the similarity with other WD40-repeat proteins is approximately 27%-30%, such as transducin/WD40 domain-containing protein with accession number np-196473, np-001078546, np-568993, np-181681, or np-199205, Prp8 binding protein (accession number np-181905) and Notchless protein-like protein (accession number np-200094).

Homology matrix of 9 sequences

| | | | | | | | | | |
|--------------|-------|--------|-------|-------|-------|-------|-------|-------|------|
| np_200094 | 100% | | | | | | | | |
| np_196473 | 19.4% | 100% | | | | | | | |
| np_001078546 | 19.4% | 100.0% | 100% | | | | | | |
| np_192182 | 26.6% | 26.5% | 26.5% | 100% | | | | | |
| np_568993 | 19.7% | 23.6% | 23.6% | 29.4% | 100% | | | | |
| np_190535 | 27.1% | 27.2% | 27.2% | 58.2% | 28.8% | 100% | | | |
| np_181905 | 22.5% | 20.9% | 20.9% | 26.5% | 25.7% | 30.0% | 100% | | |
| np_181681 | 17.4% | 15.1% | 15.1% | 22.1% | 19.7% | 27.8% | 20.1% | 100% | |
| np_199205 | 17.6% | 45.5% | 45.5% | 25.3% | 24.6% | 29.5% | 20.5% | 16.8% | 100% |

Figure 2: Similarity between putative amino acid sequence of target gene and other WD40-repeat proteins in *Arabidopsis thaliana*; The putative protein of target gene (accession number np-190535), Notchless protein-like protein (accession number np-200094), Prp8 binding protein (accession number np-181905), and transducin/WD40 domain-containing protein with accession number np-196473, np-001078546, np-192182, np-568993, np-181681 or np-199205, respectively, were compared by DNAMAN software.

3.2. Bioinformatics analysis of target gene

By analysis of this putative amino acid sequence of target gene with PROSITE software, we found there is conserved WD40 motif (Figure 3), which exhibits higher similarity to WD40 motifs in beta subunit of G-protein. Typically, WD40 motif is composed of 40-60 amino acid residues contains one variable domain and one conserved core with two dipeptide sequences, GH and WD at its N- or C-terminal, respectively. Secondary structure of this putative protein of target gene was also predicted with DNAMAN software (Figure 4), yet the transmembrane domain was not found by some bioinformatic resources (results not shown).

In addition, phosphorylation and glycosylation are very important posttranslational modification of protein and have significant effects on structure and function of protein

(Zhao and Liu, 2003), this putative protein of target gene also has several interesting sites, and four glycosylation sites were shown in Figure (5).

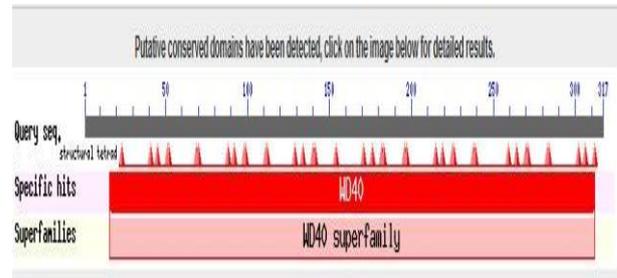


Figure 3: The conserved domain predicted in this putative protein of target gene.

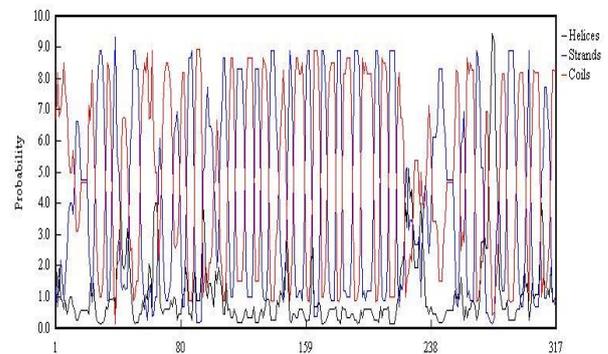


Figure 4: The putative secondary structure of this amino acid sequence encoded by target gene.

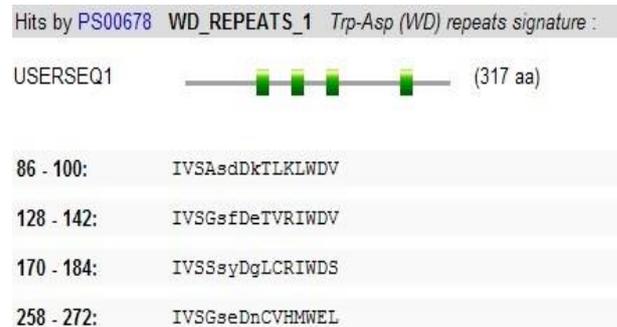


Figure 5: Glycosylation sites in this amino acid sequence encoded by target gene.

3.3. Phylogenetic analysis of WD40-repeat genes

Database searches with the putative amino acid sequences of target gene from *Arabidopsis thaliana* for homologous WD40-repeat proteins from other organisms were performed by BLAST, and some homologous sequences were found in 40 species. The putative amino acid sequences of target gene from *Arabidopsis thaliana* shows obvious similarity to WD40-repeat protein from Monocotyledons and Dicotyledons, especially the similarity with WD40-repeat protein from Dicotyledons is very

high and more than 60%. Moreover, this putative protein of target gene also shows similarity to WD40-repeat protein of mammalian, aves animals and other eukaryotes, and the similarity between putative protein of target gene and WD40-repeat protein from Monocotyledons and other species is alike (approximately 60%).

To further analyze the phylogenic relationships of target gene and other WD40-repeat genes, one gene tree was constructed by DNAMAN software (Figure 6). As shown in Figure (6), these WD40-repeat proteins can be divided into two main clusters, this putative protein of target gene from *Arabidopsis thaliana* was a group lonely, and the other cluster was composed of WD40-repeat proteins from animals and plants. In animal branch, WD repeat-containing protein 5 (accession number np_001006198) from *Gallus* belonging to aves animals, WD repeat domain 5 (accession number xp_002199389) from *Taeniopygia guttata* belonging to aves animals and some WD40-repeat proteins from mammals were clustered, such as unnamed protein product (accession number bae25694) from *Mus musculus* mouse, hypothetical protein (accession number cab66159) from *Homo sapiens*, WD repeat-containing protein 5 (accession number xp_850117) from *Canis lupus familiaris*, WD repeat-containing protein 5 (accession number np_001098945) from *Bos Taurus*, WD repeat-containing protein 5-like (accession number xp_001380141) from *Monodelphis domestica*, and WD repeat-containing protein 5-like (accession number xp_001506016) from *Ornithorhynchus anatinus*, however some WD40-repeat protein from other mammals were clustered into the other branch, such as WD repeat-containing protein 5B (accession number np_001127213) from *Pongo abelii*, WD repeat-containing protein 5B-like (accession number xp_001502155) from *Equus caballus* horse and others.

Given that homologues of this target gene are found in plant, yeast and mammalian, inferring that this target gene from *Arabidopsis thaliana* is strongly conserved through evolution, and may have important function in the development processes of eukaryotes.

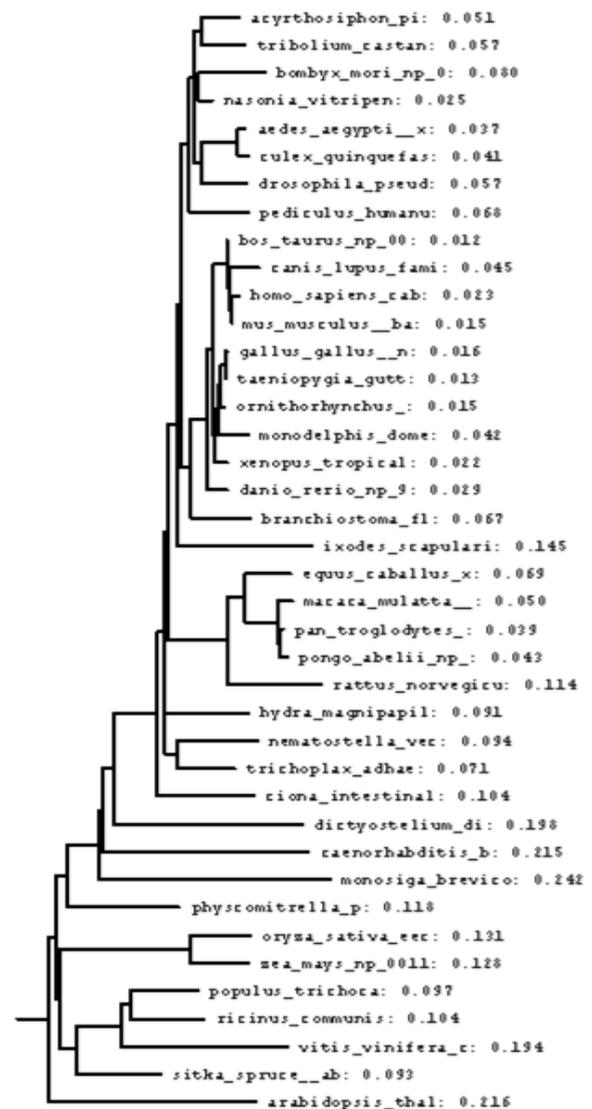


Figure 6: Phylogenetic tree of target gene from *Arabidopsis thaliana* and WD40-repeat genes from other organism; Information of WD40-repeat protein and species was described respectively as follows: WD-repeat protein (accession number eef33490) from *Ricinus communis*, unknown (accession number abk22482) from *Sitka spruce*, hypothetical protein VITISV_026465 (accession number can67116) from *Vitis vinifera*, hypothetical protein POPTRDRAFT_760587 (accession number xp_002307312) from *Populus trichocarpa*, uncharacterized protein LOC100276374 (accession number np_001143652) from *Zea mays*, hypothetical protein OsI_13357 (accession number eec76102) from *Oryza sativa Japonica Group*, WD repeat-containing protein 5 (accession number np_001011411) from *Xenopus laevis*, WD 40-repeat (accession number xp_971564) from *Tribolium castaneum* Herbst, WD40 domain-containing protein (accession number

np_190535) from *Arabidopsis thaliana*, WD 40-repeat (accession number xp_001948719) from *Acyrtosiphon pisum* Harris acyrtosiphon, wd-repeat protein (accession number xp_001662893) from *Aedes aegypti*, WD40-repeat (accession number xp_001868580) from *Culex quinquefasciatus*, GA14510 (accession number xp_001355636) from *Drosophila ananassae*, WD40-repeat (accession number xp_001599787) from *Nasonia vitripennis*, WD40-repeat (accession number np_001037087) from *Bombyx mori*, WD-repeat protein (accession number eeb15014) from *Pediculus humanus corporis*, THO complex subunit (accession number eec12084) from *Xodes scapularis*, unnamed protein product (accession number bae25694) from *Mus musculus* mouse, hypothetical protein (accession number cab66159) from *Homo sapiens*, WD repeat-containing protein 5 (accession number xp_850117) from *Canis lupus familiaris*, WD repeat-containing protein 5 (accession number np_001098945) from *Bos taurus*, WD repeat-containing protein 5 (accession number np_001006198) from *Gallus gallus*, WD repeat domain 5 (accession number xp_002199389) from *Taeniopygia guttata*, WD repeat-containing protein 5-like (accession number xp_001380141) from *Monodelphis domestica*, and WD repeat-containing protein 5-like (accession number xp_001506016) from *Ornithorhynchus anatinus*, WD repeat-containing protein 5 (accession number np_998264) from *Danio rerio*, hypothetical protein BRAFLDRAFT_281422 (accession number xp_002219410) from *Branchiostoma floridae*, similar to WD repeat domain 5 (accession number xp_002127700) from *Ciona intestinalis*, similar to predicted protein (accession number xp_002162258) from *Hydra magnipapillata*, predicted protein (accession number xp_001625397) from *Nematostella vectensis*, histone H3 recognition and presentation by the Wdr5 module of the Mll1 complex (accession number xp_002109498) from *Trichoplax adhaerens* Schulze, WD repeat-containing protein 5B-like (accession number xp_001502155) from *Equus caballus* horse, WD repeat-containing protein 5B-like (accession number xp_001112263) from *Macaca mulatta*, WD repeat-containing protein 5B (accession number xp_516691) from *Pan troglodytes*, WD repeat-containing protein 5B (accession number np_001127213) from *Pongo abelii*, WD repeat-containing protein 5B (accession number np_001019937) from *Rattus norvegicus*, WD40 repeat protein (accession number xp_001772433) from *Physcomitrella patens*

subsp *Cpatens*, WD40 repeat-containing protein (accession number xp_637302) from *Dictyostelium discoideum*, hypothetical protein CBG09206 (accession number xp_001665591) from *Caenorhabditis briggsae*, hypothetical protein (accession number xp_001750091) from *Monosiga brevicollis*. The scale bar indicates the branch length.

DISCUSSION

WD40 motif has been identified in many proteins with diverse function and is thought to be one domain of protein interaction (Neer *et al.*, 1994), and WD40-repeat proteins are very conserved in eukaryotes and mediate interaction among proteins mainly by these conserved and repeated WD40 motifs (Duan *et al.*, 2007; Duan and Ding, 2007). In this research, one 1388bp cDNA fragment was isolated from *Arabidopsis thaliana* by RT-PCR, there are conserved WD40 motifs in its encoding product, yet the transmembrane domain of protein is not found, especially this target gene is same to the gene with accession number np_190535 from *Arabidopsis thaliana*. Therefore, this target gene belongs to one member of the WD40-repeat protein family which interacts with other proteins through WD40 motif and play important roles in some processes of cell and organism (Neer *et al.*, 1994; Baum *et al.*, 2004; Zhu *et al.*, 2004). As compared with other WD40-repeat proteins from *Arabidopsis thaliana*, the putative WD40-repeat protein of target gene shows lower similarity with 7 proteins and is approximately 27%-30% except of WD40 domain-containing protein (accession number np_192182). In addition, some homologous sequences of this target gene from 40 species were found in database by BLAST, the putative protein of target gene has obvious similarity with WD40-repeat protein from Dicotyledons, secondly is the similarity with WD40-repeat protein from Monocotyledons and other species, yet is more distantly related with predicted protein from *Nematostella vectensis* (accession number xp_001625397), and the specific reasons were not clear. It's worth noting that the similarity of the putative protein of target gene with WD40-repeat protein (accession number xp_001665591) from *Caenorhabditis briggsae* is the lowest (49.5%), while is 52.8% with WD40-repeat protein (accession number xp_001750091) from *Monosiga brevicollis* which is the most primitive protozoa in Animal. Furthermore, WD40-repeat proteins from aves animals and some mammalians were clustered, but some WD40-repeat proteins from other mammalians were clustered into the other

branch. The above results showed the discrepancy exists between evolution of gene and evolution of species, inferring some of WD40-repeat genes analyzed in this research might not belong to orthologous gene, which need to be further studied.

REFERENCES

- Baum S, Bittins M, Frey S, Seedorf M. Asc1p, a WD40-domain containing adaptor protein, is required for the interaction of the RNA-binding protein Scp160p with polysomes. *Biochem J* 2004;3:823-830.
- Deng XW, Caspar T, Quail PH. cop1: a regulatory locus involved in light-controlled development and gene expression in Arabidopsis. *Genes Dev* 1991;5(7):1172-1178.
- Duan HY, Ding XS, Sun FC. WD-repeat protein. *Chinese Journal of Biochemistry and Molecular Biology* 2007;23(2):101-105.
- Duan HY, Ding XS. Research Advances in Plant WD-repeat Proteins. *Journal of Henan Agricultural Sciences* 2007;5:11-13.
- Fong HK, Hurley JB, Hopking RS, Miake-Lye R, Johnson MS. Repetitive segmental structure of the transduction β subunit; homology with the CDC4 gene and identification of mRNAs. *Proc Natl Acad Sci USA* 1986;83:2162-2166.
- Henderson IR, Liu F, Drea S. An allelic series reveals essential roles for FY in plant development in addition to flowering-time control. *Development* 2005;132(16):3597-3607.
- Madrona AY, Wilson DK. The structure of Ski8p, a protein regulating mRNA degradation: Implications for WD protein structure. *Protein Sci* 2004;13(6):1557-1565.
- Neer EJ, Schmidt CJ, Nambudripad R, Smith TF. The ancient regulatory-protein family of WD-repeat. *Nature* 1994;371:297-300.
- Smith TF, Gaitatzes C, Saxena K, Neer EJ. The WD repeat: a common architecture for diverse function. *Trends Biochem Sci* 1999;24:181-185.
- Steven VN, Philip L. The WD-repeat proteins superfamily in Arabidopsis: conservation and divergence in structure and function. *BMC Genomics* 2003;4(50):1471-2164.
- Zhao HL, Liu ZM. Protein Glycoengineering. *Progress In Biotechnology* 2003;9:18-20.
- Zhu Y, Wang Y, Xia C, Li D, Li Y, Zeng W, Yuan W, Liu H, Zhu C, Wu X, Liu M. WDR26: a novel Gbeta-like protein, suppresses MAPK signaling pathway. *J Cell Biochem* 2004;3:579-587.