Molecular characteristics and cloning of two pepper genes \textit{AN2} and \textit{UPA20}

H.X. Chen\textsuperscript{1}, H. Jiang\textsuperscript{1}, J.L. Huo\textsuperscript{2}, J.F. Wen\textsuperscript{3}, H.S. Zhu\textsuperscript{2}, C.H. Ma\textsuperscript{2}, H. Zhou\textsuperscript{2}, J.H. Lv\textsuperscript{2} and M.H. Deng\textsuperscript{2}

\textsuperscript{1}College of Horticulture, Hunan Agricultural University, Changsha, China  
\textsuperscript{2}Yunnan Agricultural University, Kunming, China  
\textsuperscript{3}Kunming University of Science and Technology, Kunming, China

*These authors contributed equally to this study.  
Corresponding author: J.F. Wen / M.H. Deng  
E-mail: wenjf888@163.com / dengminghua2013@163.com

Received May 16, 2013  
Accepted October 1, 2013  
Published January 17, 2014  
DOI http://dx.doi.org/10.4238/2014.January.17.6

\textbf{ABSTRACT.} The complete coding sequences (CDSs) of “Yunnan Purple Pepper No.1” (\textit{Capsicum annuum} L.) \textit{AN2} and \textit{UPA20} genes were amplified using the reverse transcriptase polymerase chain reaction on the basis of the conserved sequence information of some Solanaceae plants and known highly homologous pepper expressed sequence tags. The nucleotide sequence analysis of these 2 genes revealed that pepper \textit{AN2} gene encoded a protein of 263 amino acids that has high homology with the \textit{AN2}-like protein of 4 species: tobacco, tomato, potato, and petunia. The \textit{UPA20} gene encoded a protein of 341 amino acids that has high homology with the \textit{AN2}-like protein of 3 species: tobacco, petunia, and tomato. The tissue expression analysis indicated that the pepper \textit{AN2} gene was overexpressed in the pericarp and placenta; moderately in stems, flowers, and seeds; and weakly in the roots, leaves, and pericarp. The pepper \textit{UPA20} gene was overexpressed in the flowers and seeds; moderately expressed in the roots and stems; and weakly expressed in the leaves and placenta. Our findings might form the
basis for further research on these 2 pepper genes.

Key words: Gene expression profile; Capsicum annuum L.; AN2; UPA20; Bioinformatic analysis

INTRODUCTION

Pepper (Capsicum annuum L.) is a dicotyledonous flowering member of Solanaceae, which includes potato, tomato, eggplant, and African eggplants (Hunziker, 2001; Knapp, 2002). This plant has been cultivated in the Yunnan Province of China for many years. The Yunnan purple pepper No.1 (YNPP1) is special in that the whole plant is violet black, especially the fruit. Shades ranging from violet to black in C. annuum L. are attributed to anthocyanin accumulation.

Anthocyanins are responsible for the purple, blue, and pink coloration of plant parts and are widely found in plant species. Color of flowers and fruits provided by anthocyanins are helpful for attracting pollinators and seed-dispersing animals (Schaefer et al., 2004; Grotewold, 2006). Anthocyanins are beneficial for plants since they protect them from various stresses, such as strong sunlight and active oxygen species (Nagata et al., 2003; Gould, 2004).

The anthocyanin biosynthesis pathway is well understood (Schwinn and Davies, 2004). The activity of enzymes involved in anthocyanin biosynthesis is mainly regulated by transcription factors at the transcriptional level (Schwinn and Davies, 2004; Koes et al., 2005).

To determine whether transcription factors regulate anthocyanin biosynthesis in pepper, we isolated and characterized the regulatory genes, AN2, and UPA20. Their expression patterns might provide a foundation for further understanding the biochemical functions of AN2 and UPA20 in pepper.

MATERIAL AND METHODS

Sample collection

All plants were obtained from the College of Horticulture and Landscape, Yunnan Agricultural University. Pepper tissues (root, stem, leaf, flower, pericarp, placenta, and seed) were instantly frozen in liquid nitrogen and stored at -80°C before use.

Total RNA extraction and first-strand cDNA synthesis

Total RNA from pepper was extracted using Trizol (TaKaRa, China), and cDNA was synthesized using the High Fidelity PrimeScript RT-PCR kit (TaKaRa) according to manufacturer protocol.

PCR amplification

PCR was performed to isolate the pepper genes from pooled cDNAs obtained from
different tissues. Reactions were performed as described previously (Deng et al., 2012; Huo et al., 2012). The PCR program of the \textit{AN2} and \textit{UPA20} genes consisted of denaturation at 94°C for 3 min, followed by 34 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 2 min, and a final extension at 72°C for 10 min with 4°C hold.

The mRNA and amino acid sequences for \textit{AN2} and \textit{UPA20} from various plant species archived at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) were used to locate conserved regions by using multiple sequence alignment with CLUSTALW 1.8. The following primers were designed: AN2-F: 5'-CAAAGAGCAAGGGACAC-3', AN2-R: 5'-TTGGGATAGTACGAACAGA-3'; UPA20-F: 5'-TGTGACATTTCTGAGACCCCAT-3', UPA20-R: 5'-AAATACTCCAGCCTTGATA-3'.

**Reverse transcriptase (RT)-PCR expression profile**

The RT-PCR primers for expression profile analysis of pepper \textit{AN2} and \textit{UPA20} were the same as those used for PCR. PCR was optimized to ensure sufficient product intensity within the linear phase of amplification.

**Bioinformatics**

PCR amplification was repeated 5 times. Amplification products were cloned into pMD18-T (TaKaRa) and bidirectionally sequenced. At least 10 independent clones were sequenced for each PCR product. Sequencing data were edited and aligned in DNASTAR (DNAStrar Inc., USA). The cDNA sequences were predicted using GenScan (http://genes.mit.edu/GENSCAN.html).

The theoretical molecular weight (Mw), isoelectric point (pl), amino acid composition, atomic composition, extinction coefficients, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) of the putative proteins were predicted using ProtParam (http://web.expasy.org/protparam/). Signal peptide, subcellular localization, membrane-spanning regions, and PFAM domains were identified using the SignalP 4.0 server (http://www.cbs.dtu.dk/services/SignalP/), PSort (http://psort.hgc.jp/), TMpred (http://www.ch.embnet.org/software/TMPRED_form.html), and SMART (http://smart.embl-heidelberg.de), respectively. The Blastp program and conserved domain architecture retrieval tool were used to identify similar proteins and conserved domains (http://www.ncbi.nlm.nih.gov/Blast). Alignment of the nucleotide sequences and deduced amino acid sequences was performed using ClustalX, and phylogenetic trees were constructed in ClustalX and Mega 4.0 with standard parameters. Secondary structures of the deduced amino acid sequences were predicted in SOPMA (http://npsa-pbil.ibcp.fr/).

**RESULTS**

**RT-PCR of pepper \textit{AN2} and \textit{UPA20}**

RT-PCR with pooled tissue cDNAs for pepper \textit{AN2} and \textit{UPA20} genes yielded products of 1096 and 1672 bp (Figure 1).
Sequence prediction showed that the cDNAs represent 2 genes that encode 263- and 341-amino acid proteins, respectively.

**Bioinformatic analysis**

The complete coding sequences (CDSs) and the encoded amino acids are presented in Figures 2 and 3.

The pI of the putative proteins of AN2 and UPA20 was 5.17 and 5.10, respectively, and Mw was 63,681 and 84,941 Da, respectively. Total number of negatively charged residues (Asp + Glu) and total number of positively charged residues (Arg + Lys) of the 2 pepper genes were 0. Their formulas were C_{2380}H_{3973}N_{789}O_{988}S_{132} and C_{3171}H_{5321}N_{1023}O_{1315}S_{191}; total number of atoms, 8262 and 11,021; extinction coefficient (assuming all pairs of Cys residues form cystines), 8250 and 11,875; instability index, 45.24 and 45.12 (indicating unstable); aliphatic index, 34.73 and 34.41; and GRAVY, 0.774 and 0.813, respectively. The extinction coefficient (assuming all Cys residues are reduced) and estimated half-life (mammalian reticulocytes, in vitro) of the 2 pepper genes were 0 and 4.4 h, respectively.

**Figure 1.** RT-PCR of pepper AN2 and UPA20. Lane M = DL2000 DNA marker; A. PCR product for AN2; B. PCR product for UPA20.

**Figure 2.** Complete CDS and amino acid sequence of AN2 in pepper. *Stop codon.
The putative AN2 and UPA20 proteins were also analyzed. The AN2 conserved domains were identified as those belonging to the SANT superfamily. UPA20 conserved domains were identified as HLH superfamily.

The results of secondary structure prediction showed that the deduced pepper AN2 protein contained 33.59% alpha helices, 6.87% extended strands, 4.96% beta turns, and 54.58% random coils. The deduced pepper UPA20 protein contained 23.82% alpha helices, 11.76% extended strands, 4.12% beta turns, and 60.29% random coils.

The pepper genes did not contain putative signal peptides (Bendtsen et al., 2004) and were not potential membrane proteins (Moller et al., 2001). Pepper AN2 was probably located with 45% likelihood in the cytoplasm and UPA20 was probably located with 88% likelihood in the nucleus (Nakai and Horton, 1999).

Sequence analysis and evolutionary relationships

The homology of the pepper AN2 gene was determined and analyzed. The results showed that the pepper AN2 protein had high homology with the AN2 protein of 4 other plant species (Figure 4). UPA20 shared homology with UPA20 protein of the species shown in Figure 5.
Phylogenetic trees were constructed on the basis of the alignments as shown in Figures 6 and 7. Pepper AN2 and UPA20 were most closely related to AN2 and UPA20 of potato and tomato.

![Figure 6. Phylogenetic tree of AN2 genes.](image)

![Figure 7. Phylogenetic tree of UPA20 genes.](image)
Molecular cloning and characteristics of two pepper genes

Tissue expression

The pepper AN2 gene was highly expressed in the pericarp and placenta; moderately expressed in the stems, flowers, and seeds; and weakly expressed in the roots, leaves, and pericarp (Figure 8). The UPA20 gene was highly expressed in the flowers and seeds; moderately expressed in the roots and stems; and weakly expressed in the leaves and placenta.

Figure 8. Tissue expression of pepper AN2 and UPA20; actin served as the internal control. Lane M = DL2000 DNA marker; lane 1 = root; lane 2 = stem; lane 3 = leaf; lane 4 = flower; lane 5 = pericarp; lane 6 = placenta; and lane 7 = seed.

DISCUSSION

Comparative genomics determines the relationship of genome structure and function of different species (Hardison, 2003). Several researchers have shown that AN2 or UPA20 proteins from different species are highly conserved (Quattrocchio et al., 1999; Griesbach, 2002; Kay et al., 2007; Zhou and Chai, 2008; van der Hoorn and Kamoun, 2008; Yamagishi et al., 2010). Therefore, tomato can be used as a model organism to study pepper AN2 and UPA20 or pepper can be used as a model organism to study tomato, potato, or tobacco AN2 and UPA20 genes.

In this study, pepper AN2 and UPA20 genes were isolated and characterized. Many genes encoding AN2 and UPA20 from several plant species have been isolated and characterized at the genetic, chemical, and enzymological levels (Quattrocchio et al., 1999; Griesbach, 2002; Kay et al., 2007; Zhou and Chai, 2008; van der Hoorn and Kamoun, 2008; Yamagishi et al., 2010). However, the AN2 and UPA20 genes from pepper have not been previously cloned. To our knowledge, this is the first study verifying the presence of AN2 and UPA20 in pepper.

The isolated AN2 cDNA was 789-bp long and encoded 263 amino acids. Amino acid sequence comparison showed that this protein had high homology with those from tobacco, tomato, potato, and petunia. The isolated UPA20 cDNA was 1298-bp long and encoded 341 amino acids. Sequence analysis revealed that the pepper UPA20 had high homology with the proteins from 3 species-tobacco, petunia, and tomato. The alignment analyses for pepper AN2 and UPA20 proteins revealed that these proteins did not show complete identity with the proteins from eggplant or other plants. This implied that pepper AN2 and UPA20 have some
different functions than those in eggplant and other plants.

In this study, we not only cloned the CDSs of the pepper AN2 and UPA20 genes but also performed a sequence analysis and determined the tissue expression profile. The tissue expression profile analysis indicated that the genes were obviously differentially expressed in various tissues. Since we did not study their functions at the protein level, we did determine the many possible reasons for their differential expression. The possible explanation might be that the biological activities associated with the functions of these genes were required to a different extent in different tissues at the same time.

In summary, we first isolated the pepper AN2 and UPA20 genes and performed the necessary functional and tissue expression profile analyses. The cDNA clone, sequence information, and functional analysis of the pepper AN2 and UPA20 genes might be extremely important in elucidating the molecular mechanisms of their gene expression.

REFERENCES


