**The Journal of Animal & Plant Sciences, 23(5): 2013, Page: 1360-1365**
**ISSN: 1018-7081**

# MOLECULAR CLONING AND ANALYSIS OF A *CONSTANS* HOMOLOG FROM *NICOTIANA TABACUM*

Y. Lu[*], Y. Liu[**], Y. Sun[**], J. Mu[**], M. Ren[**], X. Zhang[**] and Z. Wang[**]

[*]The Graduate School of Chinese Academy of Agricultural Sciences, Beijing 100081, P.R. China;
[**]Key Laboratory of Tobacco Genetic Improvement and Biotechnology, Tobacco Research Institute, Chinese Academy of Agricultural Sciences, Qingdao 266101, P.R. China
Corresponding author E-mail: wzdycs@tom.com; Tel: 13869835575

## ABSTRACT

Plant development is largely controlled by day length. In *Arabidopsis thaliana*, the *CONSTANS* (*CO*) gene plays a pivotal role in the photoperiodic pathway controlling floral transition. In this research, we employed *Nicotiana tabacum cv. Kutsaga Mammoth 10* as experimental materials and designed hetrologous primers based on the *S. tuberosum* CO gene sequence and tobacco expressed sequence tags (ESTs). By using the approach of RT–PCR combined with rapid amplification of cDNA ends (RACE) technology, a novel cDNA encoding a CO homolog named *NtCO1* (GenBank accession number, JN022535.1) was cloned. The full-length cDNA sequence isolated was 1493 bp with an open reading frame (ORF) of 1211 bp, encoding a protein of 403 amino acids. The predicted NtCO1 protein contains two B-box-type zinc fingers and a CCT domain. Analysis based on amino acid sequence alignment showed that NtCO1 shares high identity with StCO (86.5%) from *Solanum tuberosum*, AtCO (50%) from *A. thaliana* and Hd1 (43.7%) from *Oryza sativa*. Semi-quantitative RT–PCR analysis showed that *NtCO1* was expressed specifically and strongly in leaf tissues but weakly in stem and root tissues.

**Keywords:** *CONSTANS* (*CO*) gene; tobacco; cloning; analysis; *NtCO1*

## INTRODUCTION

Flowering marks an important transition from the vegetative to the reproductive growth phase in the plant life cycle. The photoperiod is an important environmental cue that determines flowering time. Some molecular and genetic studies have showed that the *CONSTANS* (*CO*) gene plays a central role in the photoperiod pathway. The product of *CO* is a transcription factor that contains a zinc finger motif (Imaizumi *et al*., 2005; Kobayashi *et al*., 2007). The *CO* gene was first isolated by map-based cloning from a flowering-delayed mutant of *Arabidopsis* (Putterill *et al*., 1995). The *CO* genes encode proteins with two highly conserved regions: two B-box domains at the amino terminus and a CCT (CO, CO-like, and TOC1) domain near the carboxy terminus. The B-box region has been implicated in protein–protein interactions, and the CCT domain is a nuclear localization signal (Borden, 1998; Strayer *et al*., 2000; Torok *et al*., 2000; Robson *et al*., 2001; Khanna *et al*., 2009). In rice, a major quantitative trait locus (QTL) controlling the response to photoperiod, *Hd1*, was identified by a map-based cloning strategy (Yano *et al*., 2000). Sequencing analysis revealed that it was a homolog of the *CONSTANS* gene in *Arabidopsis*.

Using the sequences of CO and HD1 as models, many *CONSTANS* homologs genes have been identified in different plant species including wheat (*Triticum aestivum*) (Nemoto, 2003), eastern cottonwood (*Populus deltoides*) (Yuceer *et al*., 2002), and potato (*Solanum tuberosum*) (Drobyazina and Khavkin, 2006). COL genes have been identified in several plant species, and each species appears to have a large family of these genes. *Arabidopsis* has 17 COL genes (Robson *et al*., 2001), and rice has at least 16 (Griffiths *et al*., 2003). The *Arabidopsis* CO family can be divided into three groups according to variations in the second B-box. Group I includes CO and COL1–COL5, which contain two B-boxes. Group II includes COL6–COL8 and COL16, which contain two B-boxes, but the second B-box region is less closely related to that of CO. Group III includes COL9–COL15, which has one B-box (Robson *et al*., 2001; Griffiths *et al*., 2003).

*CO* genes play the most crucial roles within the photoperiod pathway. In plants, light signals are first perceived by the photoreceptors containing phytochrome (*PhyA*, *PhyB*, *PhyC*, *PhyD* and *PhyE*) and cryptochrome, which process the physical signals. Expression of CO mRNA and the stability of the CO protein are regulated by the circadian clock and photoreceptors, respectively. Transcriptional regulation of CO results in circadian expression patterns. The CONSTANS protein regulates the expression of floral integrators, namely FLOWERING LOCUS T (FT) and SUPRESSOR OF CONSTANS1 (SOC1) in *Arabidopsis*, ultimately leading to the transition to reproductive growth (Onouchi *et al*., 2000).

In tobacco, most varieties show a neutral response to light. Only multi-leaf varieties are typical of

short-day plants. There has been extensive physiological and biochemical research on photoperiod in tobacco (Wang, 1991; Yan and Zhao, 2001; Duan *et al*., 2011). However, the genetic mechanisms by which photoperiod affects flowering time in tobacco are still unclear. One way to investigate the roles of genes during floral bud development in tobacco is to clone the genes involved in photoperiod pathways. Here, we isolated a COL gene from the multi-leaf tobacco variety Kutsaga Mammoth 10 for the first time, and studied its expression pattern during development.

## MATERIALS AND METHODS

**Plant materials and growth conditions:** *Nicotiana tabacum* cv. Kutsaga Mammoth 10 was obtained from the germplasm resource bank of the Tobacco Research Institute, Chinese Academy of Agricultural Sciences. The seeds were grown in growth chambers under a short-day regime (27°C, 8-h light/22°C, 16-h dark) with light intensity of 700–900 $\mu molm^{-2}s^{-1}$. Tissues and organs were collected at various developmental stages as follows: root and stem (between the 1st and 2nd true leaf) and the 3rd true leaf at the young seedling stage when the 4th leaf had appeared; root and stem (between the 5th and 6th true leaf) and 8th true leaf at the vegetative growth stage when plants had 10–12 leaves; root and stem (between 13th and 14th true leaf) and 15th true leaf at the reproductive stage, when the first flower bud appeared. All samples were frozen in liquid nitrogen immediately after collection and stored at –70°C until use.

**Total RNA extraction and cDNA synthesis:** Total RNA was isolated from each tissue sample using TRIzol® reagent (Invitrogen, CA, USA) followed by DNase 1 treatment to remove any residual DNA (Promega, USA). The quality of total RNA was checked by spectrophotometry and formaldehyde denaturing agarose gel electrophoresis. Reverse transcription was performed using DNase 1 treated total RNA with oligo $(dT)_{18}$ primer and M-MLV Reverse Transcriptase (Promega, Cat. NO.M1705, USA). 5´RACE cDNA and 3´RACE cDNA, which were used to obtain the full-length cDNA sequence of *NtCO1*, were separately synthesized with reverse transcriptase using the SMART™ RACE cDNA Amplification kit (Clontech, Cat. No.634914, CA, USA).

**cDNA cloning of *NtCO1* gene from tobacco:** Blast analysis was performed using *S. tuberosum CO* gene sequence (Genbank accession no. AM888389.1) at the SOL genomics network (http://solgenomics.net/) and seven tobacco expressed sequence tags were obtained. One assembly sequence was obtained by BLAST search and by assembly of the sequences with DNAMAN software.

A specific pair of primer: CO-F1 [5´–TCGACCTTTGATTCTTGGCCGTGTTT–3´] and CO-R1 [5´–GATGCTGCCTCTCTTTGTGCCTCCTG–3´] were designed to amplify the middle fragment of the coding sequence of the gene. The amplified cDNA fragment was independently cloned into the pEASY™-T1 vector (TransGen Biotech Co., Beijing, China), and transformed into competent cells of *Escherichia coli* strain DH51α. The positive clones were screened out by blue-white selection, confirmed by PCR and sequenced. The sequence was submitted to GenBank to identify similar genes. To obtain the full-length cDNA sequence, a set of gene specific primers were designed for 5´- and 3´-RACE. These primers were as follows: GSP1 [5´–CACTACTTCCCCACCAAACAACATCC–3´] for 5´–RACE, and GSP2 [5´–GGATGTTGTTTGGTGGGGAA GTAGTGG–3´] for 3´-RACE. The amplified sequences were cloned, sequenced, and assembled. Synthesis of all primers and sequencing were conducted by Invitrogen (Shanghai, China).

***NtCO1* gene informatics analysis:** The location of the open reading frame (ORF) and DNA translation were predicted using DNAMAN software. The software Expasy (www.expasy.org) was used to analyze the theoretical isoelectric point (pI) and molecular weight of the predicted protein.

For the phylogenetic comparison, amino acid sequences of different CONSTANS-like homologs were obtained from GenBank (http://www. ncbi.nlm.nih.gov/) and aligned with BLAST (http://www.ncbi.nlm.nih.gov /BLAST/). A phylogenetic tree based on complete sequences was generated using MEGA 4 and constructed by the neighbor joining (NJ) method. Bootstrap values were derived from 1,000 replicate runs.

**Semi-quantitative RT-PCR:** Semi-quantitative RT-PCR was carried out to investigate the developmental and tissue-specific expression of *NtCO1*. The gene-specific primers used for RT-PCR were CO-YGF and CO-YGR (5´–GCAGCAACAACTGGGCAAA–3´, 5´–TTACACG CCTCGCAAAC–3´). The cDNA sequence of the actin gene was used as a control and amplified using the primer pair Actin-F (5´–AAGGGATGCGAGGATGGA–3´) and Actin-R (5´–TACAACTTGCATACGACATAGG–3´). The thermocyclere reaction conditions were as follows: initial denaturation at 94°C for 3 min; 30 cycles of 94°C for 30 s, 60°C for 45 s, and 72°C for 1 min; final extension at 72°C for 10 min. The amplified products were separated on a 1.5% agarose gel and analysed under UV-transilluminator. The experiments were repeated three times.

## RESULTS AND DISCUSSION

***NtCO1*gene cloning:** Total RNA was extracted and analyzed using a UV spectrophotometer. The A260:A280 ratios were found in the range of ~1.80. When RNA was separated by agarose gelelectrophoresis, the 28S and 18S

bands were clearly visible (Fig. 1A). These results showed that the extracted RNA samples were of high purity and integrity, with no possible DNA contamination. The primer pair CO-F1 and CO-R1 were used to amplify a middle fragment of 900 bp of the respective gene (Fig. 1B). The 5´cDNA (654 bp) and 3´cDNA (863 bp) ends of the gene were obtained by using the primers GSP1 and GSP2 respectively. (Fig.1C and D). Full length cDNA sequence of 1493 bp was obtained by assembling and designated *NtCO1* (GenBank accession number, JN022535.1).

**Sequence analysis of NtCO1 protein:** The full-length cDNA sequence was 1,493 bp with an open reading frame (ORF) of 1,211bp encoding a protein of 403 amino acids having a molecular weight of 44.8 kDa and a theoretical isoelectric point of 5.28. These characteristics are similar to those reported for CO homologs in other plants. Analysis with Online Smart Server software indicated that NtCO1 protein, without transmembrane and signalpeptides, have water-soluble properties. The predicted NtCO1 protein has two B-box-type zinc fingers at its N-terminus (a.a. 15–57; 58–100) and a CCT domain near the C-terminus (a.a. 334–378) (Fig. 2). Each B-box contained a special structure containing four cysteine residues: C-$X_2$-C-$X_{16}$-C-$X_2$-C (X represents any amino acid). This special arrangement also existed in CO protein of *Arabidopsis* (Putterill *et al*., 1995). It is similar to the zinc finger of GATA-1 transcription factiors (C-X2-C-X17-C-X2-C) and nuclear hormone receptors (C-X2-C-X13-C-X2-C) (Putterill *et al*., 1995; Ramain *et al*.,1993). Alignment of B-box domains and CCT sequences between AtCO and NtCO1 showed high levels of conservation; i.e., 77.9% identity for the B-box domain and 83.7% for the CCT domain. In *Arabidopsis* several CO mutations have been observed in the B-boxes or the CCT domain, revealing that conservation of the region is required for proper CO function (Robson *et al*., 2001). These analyses suggest that this novel protein probably belongs to the Group I of COL family.

**Analysis of NtCO1 and its homologs:** A phylogenetic tree of CONSTANS homologs was constructed using MEGA 4.0 and DNAMAN software (Fig. 3). Homology analysis revealed that NtCO1 has highest similarity (86.5%) to StCO (*S. tuberosum*). Therefore, this homolog showed the smallest genetic distance and the closest genetic relationship of all the CONSTANS homologs. NtCO1 also showed high identity with AtCO (50%) from *A. thaliana* and Hd1 (43.7%) from *Oryza sativa*. Phylogenetic analysis placed NtCO1 and StCO in the same cluster. These analyses suggest that NtCO1 is a CONSTANS homolog.

The phylogenetic tree demonstrates that members of this family could be divided into four divergent groups. NtCO1 was found to have closer relationship with CONSTANS homologs from Solanaceae plants as compared with plants from other groups. TaHd1 (*Triticum aestivum)*, LtCO (*Lolium temulentum*), ZmCO1 (*Zea mays*), and other CONSTANS homologs of Gramineae plants clustered together. CONSTANS homologs from cruciferous plants clustered into one group (AtCO (*A. thaliana*), BjCO (*Brassica juncea*), and BoCO (*Brassica oleracea*)). The evolutionary distance of CO protein basically reflects the proximity of the phylogenetic relationships among species. Other molecular evolutionary analyses showed that some of the COL family evolved before the divergence of gymnosperms and angiosperms, and COL genes within the Brassicaceae family evolve rapidly (Lagercrantz and Axelsson *et al*., 2000).

***NtCO1* gene expression analysis:** Semi-quantitative RT–PCR analysis showed that *NtCO1* was expressed in immature and mature tissues of *N. tabacum*. Transcripts were detected in all parts of the plant analyzed like roots, stems, and leaves throughout development. At the young seedling stage, *NtCO1* was weakly expressed in root, stem, and leaf tissues. In mature seedlings at the vegetative stage, *NtCO1* was more strongly expressed in leaf tissues than in stem and root tissues. At the reproductive stage, *NtCO1* was strongly expressed in leaf and stem tissues but weakly expressed in root tissues (Fig. 4).

Gene expression analysis showed that expression of *NtCO1* varied among different tissues and different developmental stages. It may play an important role in regulating flower development in tobacco, because *CONSTANS* homologs play a central role in controlling flowering time in other plants. The *CONSTANS* gene in *Pharbitis nil* (*PnCO*) was able to complement the *co* mutant of *Arabidopsis* by shortening the time to flowering (Liu *et al*., 2001). The *Brassica napus BnCOa1* gene complemented the *co-2* mutation in *Arabidopsis* in a dose-dependent manner, causing earlier flowering than that of the wild-type in both long- and short-day conditions (Robert *et al*., 1998). Over-expression of *PhalCOL* in tobacco caused an early flowering phenotype (Zhang *et al*., 2010). These results indicate that transcripts of *NtCO1* might play an important role in floral initiation.

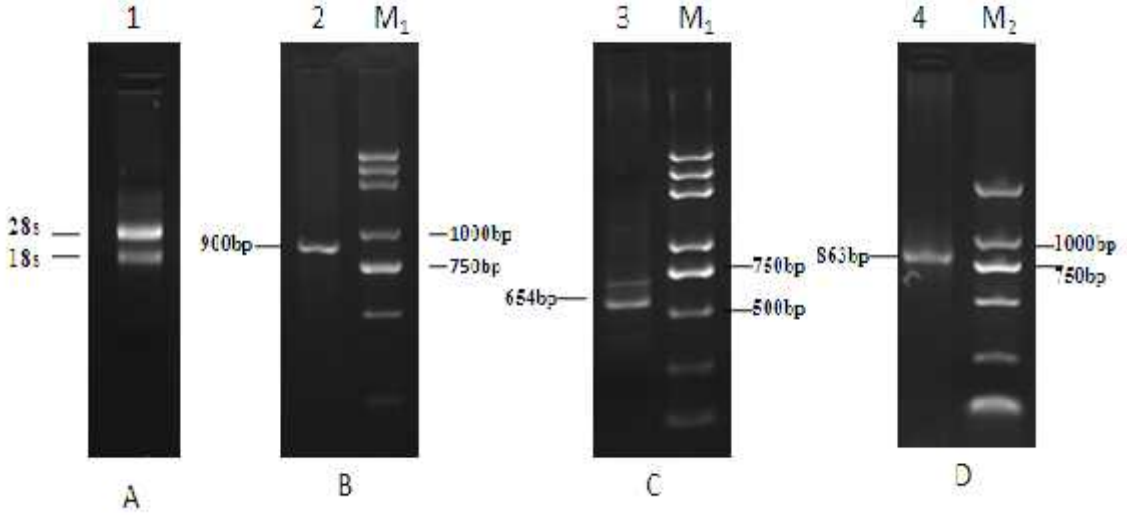


**Figure 1 Total RNA extracted from tobacco, amplification of the middle fragment, PCR products of 5´-RACE and 3´-RACE.**

**A,1: Total RNA, 28S and 18S fragments; B,2: PCR product of middle fragment; C,3: 5´-RACE amplification product; D,4: 3´-RACE amplification product. M1: DL2000 DNA marker; M2: D2000 DNA marker.**

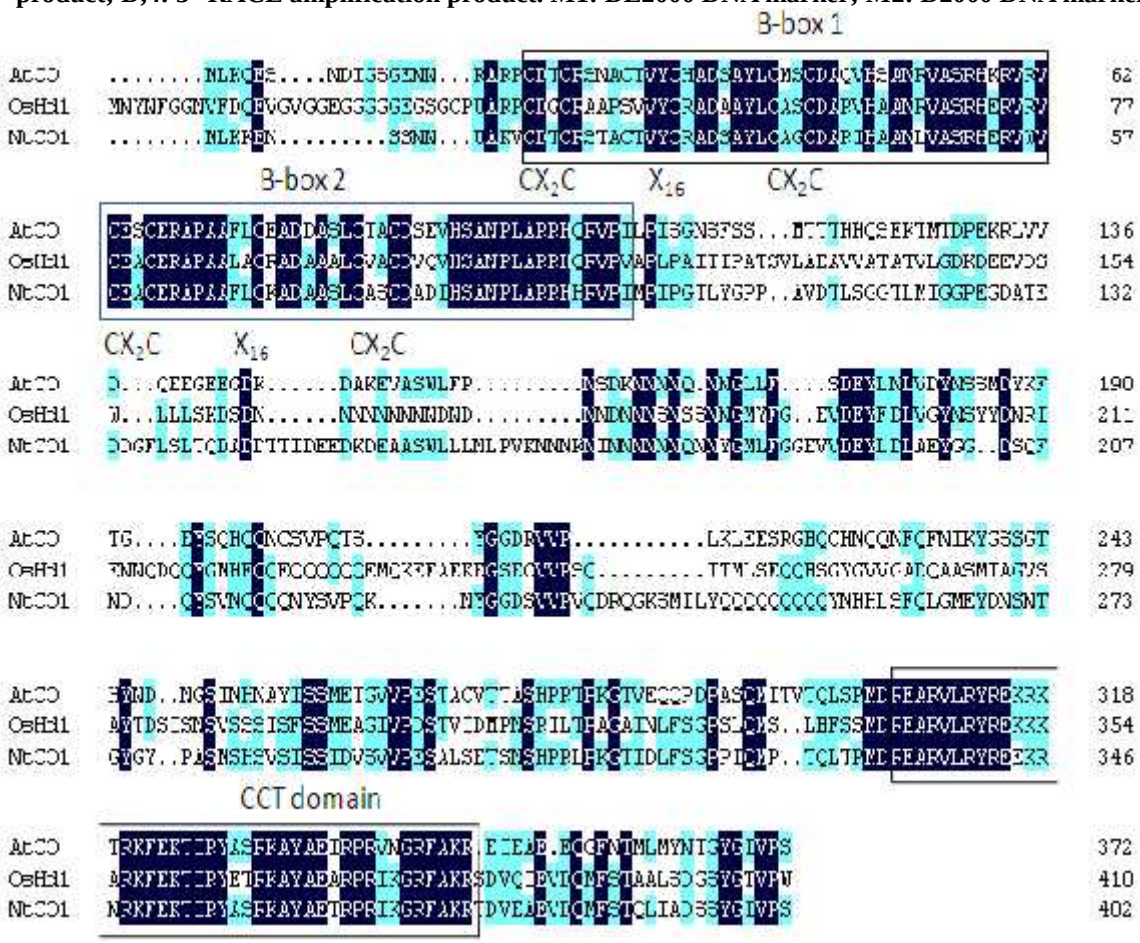


**Figure 2 Amino acid alignment of CONSTANS homologs from *Arabidopsis* (AtCO), rice (OsHd1), and tobacco (NtCO1). Note: The conserved B-Box and CCT domains characteristic of the CO gene families are boxed.**

**Figure 3**

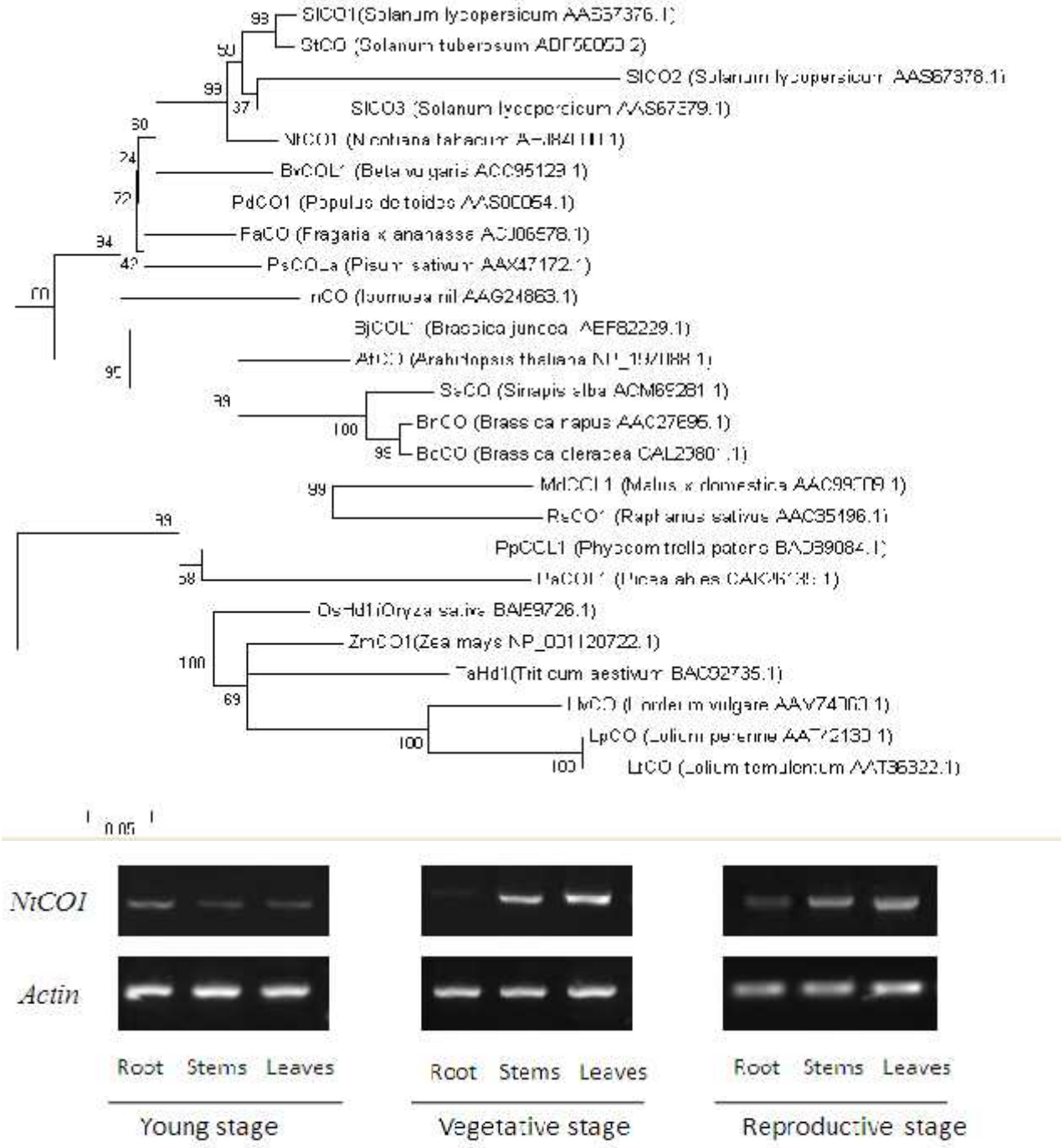


**Figure 4 Expression analysis of *NtCO1* in root, stem and leaf tissues of tobacco at three developmental stages.**

## REFERENCES

Borden, K. L. (1998). Ring fingers and B-boxes: Zinc-binding protein-protein interaction domains. Biochem. Cell Biol., 76: 351-358.

Drobyazina, P. E. and E. E. Khavkin (2006). A structural homolog of *CONSTANS* in potato.Russ. J. Plant Physiol., 53(5): 698-701.

Duan, Y., L. Jin, Y. Yang, C. Huang, Y. Jin, J. Zhou (2011). Effects of photoperiod on floral bud differentiation and growth of flue-cured tobacco. J. Agr. Sci. Technol., 13(3): 108-112.

Griffiths, S., R. P. Dunford, and G. Coupland (2003). The evolution of CONSTANS-Like gene families in

barley, rice and *Arabidopsis*. Plant Physiol., 131: 1855-1867.

Imaizumi, T., T. F. Schultz, F. G. Harmon, L. A. Ho, S. A. Kay (2005). FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in *Arabidopsis*. Sci., 309: 293-297.

Khanna, R., B. Kronmiller, D. R. Maszle, G. Coupland, M. Holm, T. Mizuno and S-H. Wu (2009). The *Arabidopsis* B-box zinc finger family.The Plant Cell, 21: 3416-3420.

Kobayashi, Y. and D. Weigel (2007). Move on up, it's time for change mobile signals controlling photoperiodic-dependent flowering. Genes. Dev., 21: 2371-2384.

Lagercrantz, U., and T. Axelsson (2000). Rapid evolution of the family of CONSTANS LIKE genes in plants. Molecular Biology and Evolution, 17(10): 1499-1507.

Liu, J., J. Yu, L. Mcintosh, H. Kende and J. A. D. Zeevaart (2001). Isolation of a CONSTANS ortholog from *Pharbitis nil* and its role in flowering. Plant Physiol., 125: 1821-1830.

Nemoto, Y., M. Kisaka, T. Fuse, M. Yano and Y. Ogihara (2003). Characterization and functional analysis of three wheat genes with homology to the CONSTANS flowering time gene in transgenic rice. Plant J., 36: 82-93.

Onouchi, H., M. I. Igeno, C. Perilleux, K. Graves and G. Coupland (2000). Mutagenesis of plants overexpressing CONSTANS demonstrates novel interactions among *Arabidopsis* flowering-time genes.Plant Cell, 12: 885-900.

Putterill, J., F. Robson, and K. Lee, R. Simon and G. Coupland (1995). The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors.Cell, 80: 847-857.

Ramain P., P. Heitzler, M. Haenlin and P. Simpson (1993). pannier, a negative regulator of achaete and scute in Drosophila, encodes a zinc finger protein with homology to the vertebrate transcription factor GATA-1. Development, 119(4): 1277-1291.

Robert, L. S., F. Robson, A. Sharpe, D. Lydiate and G. Coupland (1998). Conserved structure and function of the *Arabidopsis* flowering time gene CONSTANS in *Brassica napus*. Plant Mol. Biol., 37: 763-772.

Robson, F., M. M. Costa, S. R. Hepworth, I. V. M Piñeiro, P. H. Reeves, J Putterill and G. Coupland (2001). Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants. Plant J., 28(6): 619-631.

Strayer, C., T. Oyama, T. F. Schultz, R. Raman, D. E. Somers, P. Más, S. Panda, J. A. Kreps, S.A. Kay (2000). Cloning of the *Arabidopsis* clock gene TOC1, an autoregulatory response regulator homolog. Sci., 289: 768-771.

Torok, M. and L. D. Elkin (2000). Two B or not two B? Overview of the rapidly expanding B-box family of proteins.Different. 67: 63-71.

Wang, R. (1991). Effects of Short-day Photoperiod on growth and development of leafy varieties of flue-cured tobacco.China Tobacco. 3: 37-40.

Yan, H. and S. Zhao (2001). Ecological factors on the development characteristics of flue-cured tobacco varieties. China Tobacco Sci.,2: 15-18.

Yano, M., Y. Katayose, and M. Ashikari, U. Yamanouchi, L. Monna, T. Fuse, T. Baba, K. Yamamoto, Y. Umehara, Y. Nagamura and T. Saski (2000). Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene CONSTANS. Plant Cell. 12: 2473-2483.

Yuceer, C., R. L. Harkess, S. B. Land Jr. and D.S. Luthe (2002). Structure and developmental regulation of CONSTANS-LIKE genes isolated from *Populus deltoides*. Plant Sci., 163: 615-625.

Zhang, J. X., K. L. Wu, L. N. Tian, S. J. Zeng, and J. Duan (2010). Cloning and characterization of a novel CONSTANS-like gene from *Phalaenopsis bybrida. Acta*. Physiol. Plant. 33: 409-417.