Mol Gen Genet (2000) 262: 1047–1051

ORIGINAL PAPER

K. Kikuchi · M. Ueguchi-Tanaka · K. T. Yoshida
Y. Nagato · M. Matsusoka · H.-Y. Hirano

Molecular analysis of the *NAC* gene family in rice

Received: 21 June 1999 / Accepted: 14 October 1999

Abstract Genes that encode products containing a NAC domain, such as *NO APICAL MERISTEM* (*NAM*) in petunia, *CUP-SHAPED COTYLEDON2* (*CUC2*) and *NAP* in *Arabidopsis thaliana*, have crucial functions in plant development. We describe here molecular aspects of the *OsNAC* genes that encode proteins with NAC domains in rice (*Oryza sativa* L.). Sequence analysis revealed that the *NAC* genes in plants can be divided into several subfamilies, such as the NAM, ATAF, and OsNAC3 subfamilies. In rice, *OsNAC1* and *OsNAC2* are classified in the NAM subfamily, which includes *NAM* and *CUC2*, while *OsNAC5* and *OsNAC6* fall into the ATAF subfamily. In addition to the members of these subfamilies, the rice genome contains the NAC genes *OsNAC3, OsNAC4* (both in the OsNAC3 subfamily), *OsNAC7*, and *OsNAC8*. These results and Southern analysis indicate that the *OsNAC* genes constitute a large gene family in the rice genome. Each *OsNAC* gene is expressed in a specific pattern in different organs, suggesting that this family has diverse and important roles in rice development.

Key words NAC domain · *OsNAC* · Rice (*Oryza sativa*) · Gene family · Plant development

Introduction

Plant development is regulated by a large number of genes that encode transcription factors or proteins involved in signal transduction pathways. Transcription factors can be grouped into many different classes on the basis of the domains that bind to specific DNA sequences in the regulatory regions of downstream target genes. Examples of such domains in plants are the homeodomain, the MADS domain, and the AP2 domain.

The formation of the shoot apical meristem (SAM), the function of which is very important not only for vegetative growth but also for reproductive development, is regulated by homeobox genes such as *SHOOT MERISTEMLESS* (*STM*) and *WUSCHEL* (*WUS*) in *Arabidopsis thaliana* (Long et al. 1996; Mayer et al. 1998). Floral organ identity is regulated by ABC genes that encode MADS domains, such as *AGAMOUS* (*AG*) in *Arabidopsis* (for a review, see Riechmann and Meyerowitz 1997) . The AP2 domain was first found in the *APETALA2* gene (Jofuku et al. 1994), one of the genes that encode the A function in flower development in *Arabidopsis*. Subsequent studies have indicated that genes encoding the AP2 domain have significant functions in plant development; e.g., the *AINTEGUMENTA* (*ANT*) gene regulates the formation of ovules in *Arabidopsis* (Elliott et al. 1996) and the *indeterminate spikelet1* (*ids1*) gene is involved in spikelet development in *Zea mays* (Chuck et al. 1998).

A class of genes that encode proteins containing a novel domain has recently been found in petunia and *Arabidopsis*. The mutant *no apical meristem* (*nam*) in petunia fails to develop a SAM and produces aberrant flowers from adventitious shoots (Souer et al. 1996). The *cup-shaped cotyledon* mutant in *Arabidopsis* is associated with mutations in two genes, *CUC1* and *CUC2*. The cotyledons of this mutant are fused to form a cup-shaped structure, and the SAM is lost as in the *nam* mutant (Aida et al. 1997). The *NAM* and *CUC2* genes have been isolated and sequence analysis has revealed that the proteins encoded by them share highly conserved amino acids in the N-terminal regions (Souer et al. 1996; Aida et al. 1997). This conserved region is designated the NAC domain (Aida et al. 1997). The *NAP* gene, which has been isolated as an immediate

Communicated by K. Isono

K. Kikuchi · K. T. Yoshida · Y. Nagato · H.-Y. Hirano (✉)
Laboratory of Plant Breeding and Genetics,
Graduate School of Agricultural and Life Sciences,
University of Tokyo, Yayoi 1-1-1, Bunkyo-ku,
Tokyo 113-8657, Japan
E-mail: ahirano@mail.ecc.u-tokyo.ac.jp
Tel.: +81-3-58415065; Fax: +81-3-58415063

M. Ueguchi-Tanaka · M. Matsusoka
BioScience Center, Nagoya University, Chikusa-ku,
Nagoya 464-8601, Japan

target of the organ-identity genes *AP3* and *PI* in *Arabidopsis*, also encodes a NAC domain (Sablowski and Meyerowitz 1998). Furthermore, CUC2 has been shown to interact with STM to form the SAM and to maintain its function (Aida et al. 1999). Thus, genes that encode products containing a NAC domain have crucial functions in plant development and are required for the formation and maintenance of SAM and for floral morphogenesis.

In this paper, we describe molecular features of the gene family that encodes NAC domains in rice (*OsNAC*; the *NAC* genes from *O. sativa*). The *OsNAC* genes constitute a large gene family in rice and show characteristic expression patterns. The evolutionary relationships of the *NAC* genes in plants and structural features of the NAC domain are also discussed.

Materials and methods

Materials

Eight independent cDNAs encoding NAC domains were found in the rice EST database in September 1996 by a BLAST homology search (Altschul et al. 1990) using the sequence of the NAC domain from the petunia NAM protein. The cDNA clones were obtained from the Rice Genome Project (Japan) and six clones were analyzed here: *OsNAC3* (EST clone C21993), *OsNAC4* (C21888), *OsNAC5* (R1443), *OsNAC6* (R0546), *OsNAC7* (C0564), *OsNAC8* (C10106). The rice cultivar Kamenoo was used for RT-PCR analysis.

Sequencing and analysis of DNA

Nucleotide sequences were determined by the dideoxynucleotide chain-termination method using an automated sequencing system (ABI377, Perkin-Elmer). The cDNA clones were completely sequenced on both strands.

Primers

The following primer pairs were synthesized for PCR amplification for Southern or RT-PCR analysis: 3d1 (5′-TCTCTCTGTGACTGCACC-3′) and 3u1 (5′-GAGTAACACTCAGTATAC-3′) (C-terminal segment specific for *OsNAC3*); 4d1 (5′-GGCTACTTGCAGTCCATC-3′) and 4u1 (5′-TTGCAGATTACCAAATGC-3′) (*OsNAC4*); 5d1 (5′-TTCAAGAACACATCCCTG-3′) and 5u1 (5′-GAAGTGAACTGAAGTACC-3′) (*OsNAC5*); 6d1 (5′-GGCAGCGACCCCCTCCTC-3′) and 6u1 (5′-AAACAGGAAAGCTAGCCC-3′) (*OsNAC6*); 7d1 (5′-CCTCGTCGTTGCAGAGGC-3′) and 7u1 (5′-CAAAGCCACATTTGCATG-3′) (*OsNAC7*); and 8d1 (5′-AGTTGTCGGTGTTGAAGC-3′) and 8u1 (5′-CATAGAAACCCAGTAACC-3′) (*OsNAC8*).

RT-PCR

Total RNA was isolated from various rice organs by the method described in Naito et al. (1988) and cDNA was made using poly(A)$^+$ RNA isolated from total RNA by Oligotex-dT30 (Takara). PCR amplification for 20 cycles was carried out using cDNAs (8 ng) from various organs with primers specific for the C-terminal part of each NAC-containing gene product. The PCR products were detected by Southern analysis using probes that were specific for each cDNA.

Results

Sequence comparison of NAC-domain proteins

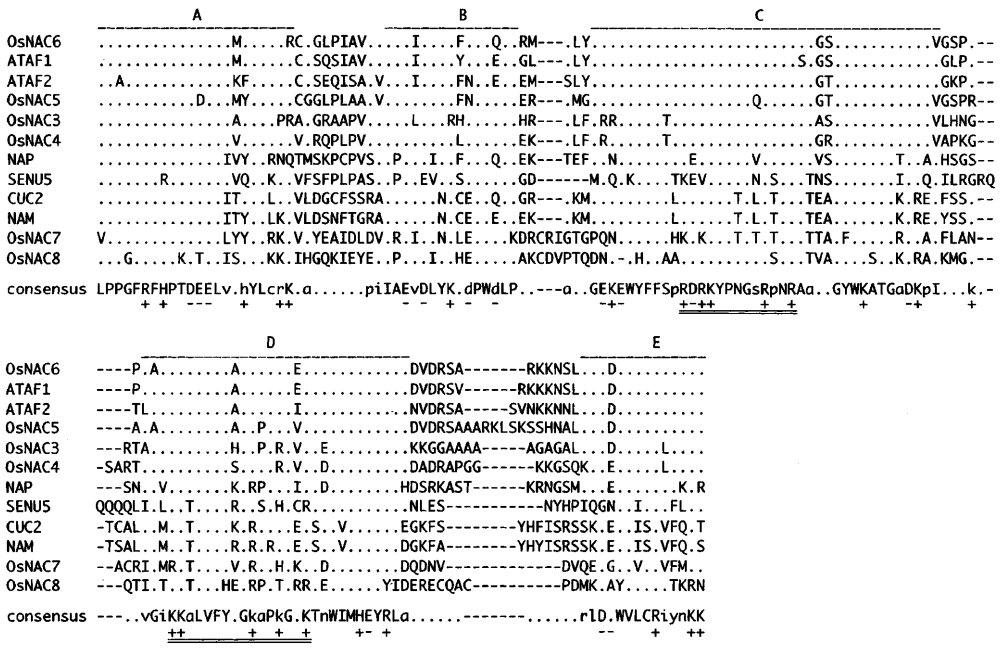
The nucleotide sequences of six cDNAs (*OsNAC3–OsNAC8*) that encode NAC domains from rice were determined (The GenBank Accession Nos. are as follows: *OsNAC3*, AB028182; *OsNAC4*, AB028183; *OsNAC5*, AB028184; *OsNAC6*, AB028185; *OsNAC7*, AB028186; *OsNAC8*, AB028187). The genes *OsNAC3-OsNAC7* encode proteins with 276–329 amino acids, while the *OsNAC8* gene encodes a larger protein of 486 amino acids. All six predicted proteins contain highly conserved regions in their N-terminal segments, which correspond to the NAC domain of NAM and CUC2 (Fig. 1). In contrast, the sequences of their C-terminal portions were highly variable among the six OsNAC proteins and contained simple repeats of amino acids such as GGGGGAAA in OsNAC5 or DDDDLLHHHHH in OsNAC7.

The amino acid sequences of the NAC domains were compared with those found in other NAC proteins (Fig. 1). The NAC domain comprises 151–159 amino acids and is located in the N-terminal region in all NAC proteins analyzed. A consensus sequence for the NAC domain was deduced from the twelve NAC proteins. In addition, we found that the NAC domain could be divided into five subdomains (A–E). Each subdomain was distinguishable by blocks of heterogeneous amino acids or gaps. While the NAC domains were rich in basic amino acids (R, K and H) as a whole, the distributions of positive and negative amino acids in each subdomain were unequal. Subdomains C and D are rich in basic amino acids but poor in acidic amino acids, while subdomain B contains a high proportion of acidic amino acids. Putative nuclear localization signals (NLS) were detected in subdomains C and D.

The *NAC* genes can be classified into several subfamilies

The evolutionary relationships among the *NAC* genes were analyzed using the amino acid sequences of the NAC domains. According to the dendrograms obtained by the neighbor-joining method (Fig. 2) and the maximum-parsimony method (data not shown), which showed similar topologies, the *NAC* genes fall into several subfamilies, i.e., the ATAF, NAM, and OsNAC3 subfamilies.

Although *ATAF1* and *ATAF2* were found in the database, no publication was referenced. However, these two genes were isolated on the basis of their ability to activate the CaMV 35S promoter in yeast (Souer et al. 1996). The two genes are similar and belong to the same subfamily. *OsNAC5* and *OsNAC6* represent this subfamily in rice. The degree of amino acid sequence identity between the NAC domains in the ATAF subfamily



was more than 75%. OsNAC6 and ATAF1 had the highest identity score (90.8%).

The genes *NAM* from petunia and *CUC2* from *Arabidopsis* are very similar to each other (89.0%). The rice genes *OsNAC1* and *OsNAC2* (manuscript in preparation) belong to this subfamily. *OsNAC3* and *OsNAC4* form a cluster but genes belonging to this subfamily have not yet been found in other plants. Although no rice *NAC* gene was observed on the branch containing *Arabidopsis NAP*, it is expected that the corresponding gene in rice will be found by extensive screening. The predicted OsNAC8 protein is exceptionally large and lacks simple amino-acid repeats in the C-terminal part. These distinct features of OsNAC8 are consistent with the evolutionary tree, in which *OsNAC8* is most distant from the other *NAC* genes.

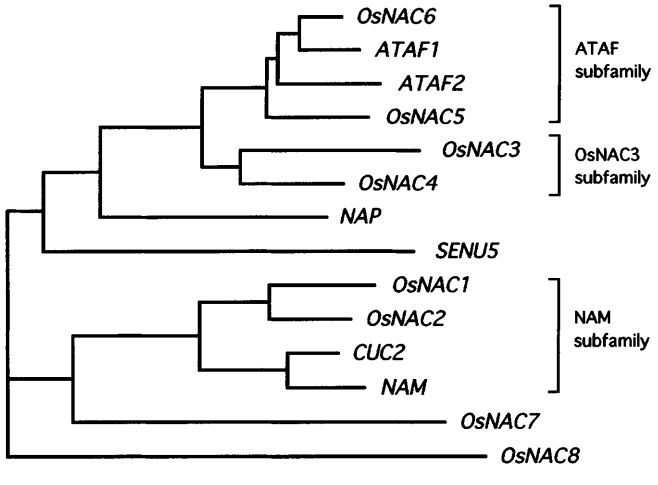


Fig. 2 A dendrogram of *NAC* genes based on the amino-acid sequences of their NAC domains. The tree was made by the neighbor-joining method

Although the sequences of the C-terminal portions of NAC proteins were highly heterogeneous in general, some homologous regions were detected as blocks of identical amino acids among the genes belonging to the same family. In the case of the ATAF subfamily, for example, the sequence SDS-P-LH–DSSCSEQV-SPEF–EVQS-PK was common to OsNAC6 and ATAF1. These semi-conserved regions were not found in the products of genes in the other subfamilies, suggesting that these regions may be associated with a specific function served by the ATAF subfamily.

Southern analysis using NAC-box probes showed that a number of genes related to the NAC genes exist in the rice genome (data not shown). The banding patterns were similar among the genes in the same subfamily, which is consistent with the classification based on the degree of sequence similarity in the NAC domain.

Organ-specific expression

To elucidate the organ-specific expression of *OsNAC* genes, the PCR products amplified from cDNA in a low number of cycles were detected by blot analysis using the gene-specific probes (Fig. 3). The actin gene, a control for constitutive expression, was uniformly expressed in all organs examined.

Each *OsNAC* gene had a unique expression pattern. *OsNAC5* was predominantly expressed in root and embryo. The expression levels of *OsNAC4* and *OsNAC7* varied among organs. *OsNAC4* was expressed strongly in leaf blade and stem, and weakly in root, young panicle and flower. In *OsNAC7*, a higher level of expression was detected in leaf blade, embryo and callus, although the expression level was extremely low or negligible in stem and young panicle. *OsNAC6* and *OsNAC8* were

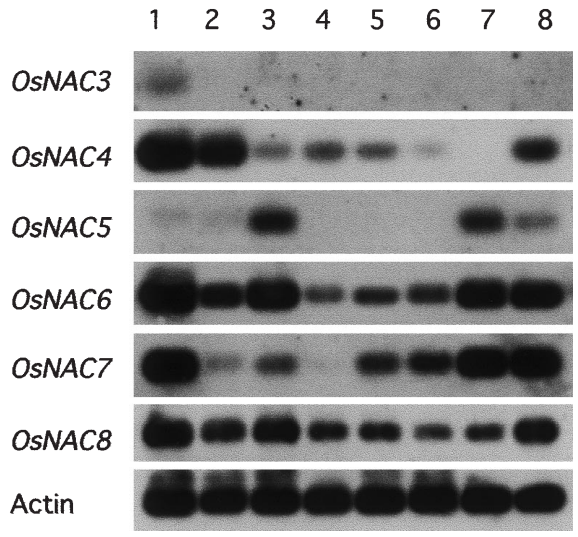


Fig. 3 Organ-specific expression of the *OsNAC* genes. The products of RT-PCR were detected by Southern hybridization. Lanes 1, leaf blade of mature plants; 2, stem of mature plants; 3, roots of seedlings; 4, young panicle including the primordia of primary and secondary rachis branches; 5, flower just before heading; 6, developing seed at 4–6 days after pollination; 7, embryo 4 days after pollination; 8, callus. The actin gene, which is expressed at a similar level in all organs, was used as a control

constitutively expressed in all organs, although slight differences in levels of expression were observed among organs. Expression of *OsNAC3* was only detected in leaf, where the gene was weakly expressed. The different expression patterns of the *OsNAC* genes suggest that their products play functionally diverse roles in rice development.

Discussion

In this paper, we have described the molecular characterization of the gene family that encodes proteins containing the NAC domain in rice. The *OsNAC* genes constitute a large gene family in the rice genome.

By comparing many proteins that have a NAC domain, we found that it could be divided into five subdomains. The function of the NAC domain remains to be understood. Although there have been no publications on ATAF1 and ATAF2, these genes are thought to code for transcription factors (Souer et al. 1996). The NLS-like sequences in subdomains C and D are highly conserved among all twelve proteins, supporting the hypothesis that NAC proteins may operate in the nucleus. If proteins with a NAC domain function as transcriptional regulators, the highly basic region (subdomains C and D) may be involved in DNA binding. We have tried to predict the secondary structure of the NAC domain using computer programs such as COILS (Lupas et al. 1991; Lupas 1996) and PHDsec (Rost and Sander 1993; Rost 1996). Regions that have the potential to form a helical structure are found in subdomains A and D, while the coiled-coil structure, which is predicted to form amphipathic α-helices in the AP2 domain (Okamuro et al. 1997), is not found. A potential helix-turn-helix structure, which is prominent in the NAC domains of the ATAF and OsNAC3 subfamilies, was found in subdomain D, suggesting interaction of this NAC subdomain with DNA (Harrison 1991).

Besides the genes described here, other rice genes encoding a NAC domain have accumulated in the database; the rice EST database now contains more than twenty independent cDNAs. Similarly, information obtained from sequencing ESTs and BACs indicates that the *Arabidopsis* genome has a large number of the *NAC* genes. In petunia, fifteen *NAC* genes were isolated by hybridization (Souer et al. 1998). Recently, GRAB1 and GRAB2, which were isolated from wheat as proteins that interact with a geminivirus protein, were shown to include the NAC domain (Xie et al. 1999). Thus, in plants, *NAC* genes seem to constitute large families. Since no *NAC* gene has been found so far in the genomes of bacteria, fungi or animals, *NAC* genes are specific to plant genomes. Likewise, the genes encoding the AP2 and SBP domains constitute multigene families in plants but are not present in the genomes of members of other kingdoms (Klein et al. 1996; Cardon et al. 1997; Moreno et al. 1997; Okamuro et al. 1997; our unpublished results). Therefore, *NAC*, *AP2* and *SBP* may have evolved after the divergence of the ancestors of plants from those of other organisms. Although genes of the MADS-box family are widely distributed not only in plants but also in yeast and animals, the MADS-box genes, which also encode a K-box, have diverged widely, and constitute many subfamilies, especially in plant genomes (Theissen et al. 1996). Many genes in the NAC, AP2, SBP and MADS families have been shown to be involved in regulating aspects of plant development, such as formation of shoot apical meristem, floral organs, and leaf constitution (Ma et al. 1991; Jofuku et al. 1994; Elliott et al. 1996; Klein et al. 1996; Souer et al. 1996; Aida et al. 1997, 1999; Cardon et al. 1997; Moreno et al. 1997; Riechmann and Meyerowitz 1997; Chuck et al. 1998; Sablowski and Meyerowitz 1998). Therefore, the genes in these families appear to have diversified and evolved during plant evolution, in association with the diversification and organization of plant architecture. The question of when the *NAC* genes arose during plant evolution is therefore of great interest. In the case of the MADS family, ferns have been shown to have the genes with a K-box, which is specific to the plant MADS-box genes (Theissen et al. 1996; Hasebe et al. 1998).

Acknowledgments We thank the Rice Genome Project (Tsukuba, Japan) for providing EST clones. This research was partly supported by a Grant-in-Aid for Scientific Research (09460002) from the Ministry of Education, Science and Culture, and a Grant-in-Aid-(Bio Design Program) from the Ministry of Agriculture, Forestry and Fisheries (Japan).

References

Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M (1997) Genes involved in organ separation in Arabidopsis: an analysis of the *cup-shaped cotyledon* mutant. Plant Cell 9:841–857

Aida M, Ishida T, Tasaka M (1999) Shoot apical meristem and cotyledon formation during *Arabidopsis* embryogenesis: interaction among the *CUP-SHAPED COTYLEDON* and *SHOOT MERISTEMLESS* genes. Development 126:1563–1570

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Cardon GH, Höhmann S, Nettesheim K, Saedler H, Huijser P (1997) Functional analysis of the *Arabidopsis thaliana* SBP-box gene *SPL3*: a novel gene involved in the floral transition. Plant J 12:367–377

Chuck G, Meeley RB, Hake S (1998) The control of maize spikelet meristem fate by the *APETALA2*-like gene *indeterminate spikelet1*. Genes Dev 12:1145–1154

Elliott RC, Betzner AS, Huttner E, Oakes MP, Tucker WQJ, Gerentes D, Perez P, Smyth DR (1996) *AINTEGUMENTA*, an *APETALA2*-like gene of Arabidopsis with pleiotropic roles in ovule development and floral organ growth. Plant Cell 8:155–168

Harrison SC (1991) A structural taxonomy of DNA-binding domains. Nature 353:715–719

Hasebe M, Wen C-K, Kato M, Banks JA (1998) Characterization of MADS homeotic genes in the fern *Ceratopteris richardii*. Proc Natl Acad Sci USA 95:6222–6227

Jofuku KD, den Boer BGW, Van Montagu M, Okamura JK (1994) Control of Arabidopsis flower and seed development by the homeotic gene *APETALA2*. Plant Cell 6:1211–1225

Klein J, Saedler H, Huijser P (1996) A new family of DNA binding proteins includes putative transcriptional regulators of the *Antirrhinum majus* floral meristem identity gene *SQUAMOSA*. Mol Gen Genet 250:7–16

Long JA, Moan EI, Medford JI, Barton MK (1996) A member of the KNOTTED class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. Nature 379:66–69

Lupas A (1996) Prediction and analysis of coiled-coil structures. Methods Enzymol 266:513–525

Lupas A, Van Dyke M, Stock H (1991) Predicting coiled coils from protein sequences. Science 252:1162–1164

Ma H, Yanofsky MF, Meyerowitz EM (1991) *AGL1-AGL6*, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. Genes Dev 5:484–495

Mayer KFX, Schoof H, Haecker A, Lenhard M, Jürgens G, Laux T (1998) Role of *WUSCHEL* in regulating stem cell fate in the *Arabidopsis* shoot meristem. Cell 95:805–815

Moreno MA, Harper LC, Krueger RW, Dellaporta SL, Freeling M (1997) *Liguleless1* encodes a nuclear-localized protein required for induction of ligules and auricles during maize leaf organogenesis. Genes Dev 11:616

Naito S, Dubé PH, Beachy RN (1988) Differential expression of conglycinin α' and β subunit genes in transgenic plants. Plant Mol Biol 11:109–123

Okamuro JK, Caster B, Villarroel R, Montagu MV, Jofuku KD (1997) The AP2 domain of *APETALA2* defines a large new family of DNA binding proteins in *Arabidopsis*. Proc Natl Acad Sci USA 94:7076–7081

Riechmann JL, Meyerowitz EM (1997) MADS domain proteins in plant development. Biol Chem 378:1079–1101

Rost B (1996) PHD: predicting 1D protein structure by profile-based neural networks. Methods Enzymol 266:525–539

Rost B, Sander C (1993) Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol 232:584–599

Sablowski RWM, Meyerowitz EM (1998) A homolog of *NO APICAL MERISTEM* is an immediate target of the floral homeotic genes *APETALA3/PISTILLATA*. Cell 92:93–103

Souer E, van Houwelingen A, Kloos D, Mol J, Koes R (1996) The *no apical meristem gene* of petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. Cell 85:159–170

Souer E, van Houwelingen A, Bliek M, Kloos D, Mol J, Koes R (1998) Co-suppression of *nam* and homologous genes leads to a reduction in axillary mereistem formation and increased leaf and stem size in *Petunia*: a possible role for NAC domain genes in plant development. Flowering Newsl 26:36–46

Theissen G, Kim JT, Saedler H (1996) Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. J Mol Evol 43:484–516

Xie Q, Sanz-Burgos AP, Guo H, García JA, Gutiérrez C (1999) GRAB proteins, novel members of the NAC domain family, isolated by their interaction with a geminivirus protein. Plant Mol Biol 39:647–656