BIOINFORMATIC ANALYSIS OF THE ARABIDOPSIS, POPLAR, AND RICE NAC DOMAIN TRANSCRIPTION FACTOR FAMILY AND FUNCTIONAL ANALYSIS OF PTNAC068

by

KIMBERLY DIANE HUNT

(Under the Direction of Sarah Covert)

ABSTRACT

The NAC domain transcription factor family is found only in plants and is involved in important developmental and defense processes. A detailed understanding of the evolutionary relationships within the NAC family would facilitate the development of hypotheses about the function of specific NAC domain proteins. To this end, we conducted several bioinformatic analyses of all Arabidopsis, poplar, and rice NAC domain transcription factors. A phylogenetic analysis of these proteins assembled into 52 clades. Seventeen clades contained sequences from all three species and are predicted to include proteins with conserved functions in plants. Members of the remaining clades are predicted to fall into broad functional groups based on their genome of origin and that of the other proteins with which they clustered (if any). To identify relatively short stretches of conserved amino acids that might activate transcription, convey functional specificity or affect dimerization, the MEME and MAST programs were used to identify conserved motifs outside of the larger NAC domain. In total, 129 conserved motifs were identified as conserved within individual clades. Seventy-seven of these motifs were

present in at least one other sequence in another clade, while 23 of them were present in multiple members of other clades. Proteins that share motifs are predicted to have related functions.

Secondary wall thickening is a hallmark of cells involved in support or water transfer, and is usually seen in vascular cells. Understanding the molecular switches that control secondary wall deposition in plants is important for basic plant biology and understanding the development of vascular tissues. PtNAC068 is a *Populus trichocarpa* NAC domain protein that was predicted to regulate vascular development on the basis of its similarity to vascular related NAC domain proteins in *Arabidopsis* and the presence of its transcription in poplar vascular tissue in EST libraries. Analyses of PtNAC068 over-expression, knockdown, and dominant repression lines of transgenic poplar plants were used to elucidate the role of PtNAC068 in vascular development. PtNAC068 over-expression mutants displayed ectopic lignin deposition in leaf epidermal cells and primary leaf vein pith cells, as well as increased lignin deposition in and around phloem fiber bundles and their associated cells. Dominant repression PtNAC068 mutants displayed a reduction in lignin deposition associated with phloem fiber bundles in primary leaf veins and stems, and a loss of sclereids associated with phloem fiber bundles in stems. Thus we conclude that PtNAC068 is a positive regulator of secondary wall development in phloem fibers and associated cells.

INDEX WORDS: NAC Domain, Transcription Factors, PtNAC068, Vascular Development, Phylogenetic Analysis

BIOINFORMATIC ANALYSIS OF THE *ARABIDOPSIS*, POPLAR, AND RICE NAC DOMAIN TRANSCRIPTION FACTOR FAMILY AND FUNCTIONAL ANALYSIS OF PTNAC068

by

KIMBERLY DIANE HUNT

B.S., University of Florida, 2001

A Thesis Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2008

BIOINFORMATIC ANALYSIS OF THE *ARABIDOPSIS*, POPLAR, AND RICE NAC DOMAIN TRANSCRIPTION FACTOR FAMILY AND FUNCTIONAL ANALYSIS OF PTNAC068

by

KIMBERLY DIANE HUNT

Major Professor: Sarah Covert

Committee: Jeffrey Dean
Sidney Kushner
Russell Malmberg
Joe Nairn

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2008

DEDICATION

To my parents, Lee Sr. and Pamela Hunt, and my brothers, Lee Jr., Chris, and Brian Hunt, for their constant faith in me and my abilities.

ACKNOWLEDGEMENTS

I want to begin by thanking my advisor, Dr. Sarah Covert for challenging me to do things I never thought possible and furthering my scientific development. I also want to thank each member of my committee, Dr. Jeffrey Dean, Dr. Sidney Kushner, Dr. Russell Malmberg, and Dr. Joe Nairn: first, for taking the time to serve on my committee, and secondly for their tremendous support and advice over the years. I want to thank Janice Lunsford for her tireless efforts on behalf of me and all of the graduate students in the Department of Genetics. I would like to thank the members of the Covert lab, both past and present. I would also like to thank Dr. Hazel Wetzstein and her lab for all the microscopy help. I want to thank the Meagher lab, especially Daniel Ruzicka, for all the help with the Real-time PCR and associated experiments. I want to thank everyone involved in our weekly group lab meetings. I want to thank all the members of Scott Gold's lab for all the troubleshooting help and for their friendship over the years. Finally, I want to thank all my friends, especially, Rebecca Tomlinson, for all the support and friendship through the years.

TABLE OF CONTENTS

Page

Chapter 1

Introduction and Literature Review

NAC domain Family

NAC domain proteins are a large transcription factor family involved in plant development. The 'NAC' name was coined following the identification of the first four family members and stands for petunia No Apical Meristem (NAM), *Arabidopsis* ATAF1/2, and Cup-Shaped Cotyledon 2 (CUC2) [1]. All NAC domain proteins have two structural regions, a conserved NAC domain and a variable region [2-4]. The conserved NAC domain is 151-159 amino acids long and is near the N-terminus of each NAC protein. It consists of five subdomains designated A, B, C, D, and E (Figure 1.1) [2-4]. The five subdomains are distinguished by intervening blocks of heterogeneous amino acids or gaps [2]. Of these, subdomains C and D are rich in basic amino acids, contain putative nuclear localization signals (NLS), a helix-turn-helix structure, and are predicted to be involved in DNA binding [2-4]. In one family member, AtNAM, subdomain E is required for DNA binding and can form a helix-turn-helix motif [3, 4]. Subdomains D and E in AtNAM also function as a transcriptional activator [3]. All NAC domain proteins contain a "variable region" after conserved subdomain E that is highly heterogeneous and along with the NAC domain may be involved in determining the function of family members [4].

The first NAC domain protein, No Apical Meristem (NAM), was identified in petunia by screening an insertional plant library for seedlings that were arrested in development [5]. Since the initial discovery of NAM, many more NAC domain family members have been found in plants. The most defining phenotypic features of NAM, CUC 1, and CUC2 mutants were the complete loss of the shoot apical meristem (SAM), fusion of the cotyledons, and aberrant flowers on escaped shoots [1, 5-9]. NAM, CUC1, CUC2, and CUC3 are expressed at the boundaries of a variety of organ primordia and meristems, and are involved in the development and maintenance

of the embryonic shoot apical meristem as well as in the development of floral organs , [1, 5-9]. NAM is thought to function in the specification of primordia and meristems by influencing the rate and or direction of cell division and expansion in the regions between the primordia [5]. CUC mutants are thought to function in the specification of primordia and meristems by influencing cell division and differentiation [6, 8, 9]. Other NAC domain proteins, NAC1 and AtNAC2, are expressed in *A. thaliana* root tip meristems and at lateral root initiation sites [10, 11]. NAC1 and AtNAC2 initiate lateral root formation through auxin and ethylene signaling pathways (Xie et al, 2000; He et al 2005). RING HC proteins have been shown to bind NAC family proteins [12, 13]. SINAT5, a RING HC protein, down regulates auxin signals in plants by targeting NAC1 for ubiquitin-mediated proteolysis [13].

Recently, the NAC domain family has been implicated in vascular development in *Arabidopsis*. Seven vascular-related NAC domain (VND) proteins were identified using an *Arabidopsis in vitro* xylem vessel element formation system in association with microarray and promoter analyses [14]. Two of the seven VND proteins, VND6 and VND7, function as transcriptional switches for plant metaxylem and protoxylem vessel formation [14]. Other NAC domain proteins involved in vascular development are NST1, NST2, and NST3/SND1/ANAC012 [15-19]. The functions of NST1 and NST2 were elucidated using a chimeric repressor derived from NAC secondary wall thickening promoting factor 1 (NST1) and NST2 [17]. NST1 and NST2 act redundantly to regulate the secondary cell wall thickenings in anther walls and tracheary elements [17]. NST3/SND1/ANAC012 is expressed in vascular tissues of *Arabidopsis* stems and roots and regulates secondary wall thickening in fiber cells [15, 16, 18, 19].

Besides developmental processes, members of the NAC family are involved in plant defense and stress responses. The NAC domain proteins first identified in viral or pathogen defense are GRAB1 and GRAB2 (Geminivirus Rep A-binding). GRAB1/2 expression was shown to inhibit wheat dwarf geminivirus DNA replication in cultured wheat cells by interacting with RepA protein [20]. Following the identification of the GRAB proteins, a NAC protein designated TIP (TCV-interacting protein) was implicated as an essential component in the turnip crinkle virus (TCV) resistance response pathway [21]. Since then, other NAC proteins associated with pathogen defense were identified. These NACs include Stprx2 and StNAC, OsNAC19, CaNAC1, and ATAF2 [22-25]. In addition to induction by pathogens, some NAC proteins are induced in response to environmental stresses. OsNAC6 and TaNAC69 are two NAC proteins whose expression is induced in response to cold stress, drought stress and herbivory [26, 27]. ANAC019, ANAC055, and ANAC072 are NAC proteins induced by drought, high salinity, and abscisic acid [28].

A phylogenetic study of *A. thaliana* and *Oryza sativa* NAC domain proteins identified candidates that may be responsible for the morphological and physiological differences between monocotyledonous and dicotyledonous plants [4]. In this study, 75 putative rice NAC domains (identified from a cDNA database) were compared to 105 putative *A. thaliana* NAC domains (identified from the *Arabidopsis* genomic database) [4]. From this comparison, the NAC domain family was classified into two main groups and 18 subfamilies [4]. Two subfamilies contain only rice proteins (total of at least 6 proteins), while 4 subfamilies contain only *A. thaliana* proteins (total of 24 proteins). Since NAC domain family members are known to be involved in regulating plant developmental processes, it is possible that NAC domain proteins unique to either rice or *A. thaliana* might be involved in determining the morphological differences

between monocotyledonous and dicotyledonous plants [4]. Of the eighteen NAC domain subfamilies, twelve contain identifiable motifs in the variable region, demonstrating that some of the variable domains have common motifs despite their divergence [4].

Secondary Vascular Development

Secondary vascular tissues in plants are produced by a ring of meristematic tissue, known as the vascular cambium, which is located underneath the bark of trees or shrubs (Figure 1.2). The vascular cambium produces two types of cells: fusiform initials and ray initials (Figure 1.3). Fusiform initials are long tapered cells that undergo longitudinal cell division to produce two elongated cells [29]. One of these cells differentiates into secondary phloem or xylem while the other cell stays as a fusiform initial [29]. Phloem cells are produced toward the bark while xylem cells are produced toward the center of the tree, expanding its girth. Phloem produced from the fusiform initial functions in the transport of sugars resulting from photosynthesis, while xylem transports water and minerals absorbed by the roots [30]. Ray initial cells are short cubiodal cells and, like fusiform initials, undergo longitudinal cell division to produce two cells, one cell remains as a cambial ray initial and the other cell differentiates into either phloem or xylem parenchyma [29]. Phloem parenchyma and xylem parenchyma cells help to move water and food substances horizontally in plants. These cells are alive when mature and also function in regeneration and wound healing [31, 32].

Phloem is composed of multiple cell types which function in support or transportation of photosynthetic sugars horizontally and vertically throughout plants. Secondary phloem is typically composed of four cell types: companion cells, sieve tube members, parenchyma, and fibers [33]. Sieve tube members and associated companion cells are produced from the same

mother cell and remain in contact with each other for the rest of their existence [33]. Companion cells, unlike sieve elements, have a dense protoplasm containing the cell's nucleus, mitochondria and ribosomal machinery, and they provide the associated sieve tube member with proteins, signal molecules, ribosomal proteins and ATP [33]. Phloem parenchyma cells are usually responsible for transporting photosynthates into sieve tube members [33]. Phloem fiber cells are usually elongated cells with thick, lignified walls and they occur in bands or bundles [33]. These cells are flexible with great tensile strength and usually function in protection and support [33].

Several genes are known to affect normal phloem development. WOODEN LEG (WOL) is responsible for proper proliferation of the procambial cells in *Arabidopsis* roots that lead to phloem cell development and xylem organization [34]. ALTERED PHLOEM DEVELOPMENT (APL) encodes a MYB coiled-coil type of transcription factor that is required for identity and development of phloem tissues in *Arabidopsis* plants [35].

Terrestrial, seed-producing plants are divided into angiosperms and gymnosperms [31]. Both angiosperms and gymnosperms produce xylem that contains tracheids, the elongated, spindle-shaped cells that transport water and minerals [31, 32]. However, angiosperm xylem is distinct from gymnosperm xylem because angiosperm xylem contains vessel elements, while gymnosperm xylem contains only tracheids [31, 32]. Vessel elements are shorter and wider than tracheids and have perforated ends walls that form perforation plates [31, 32]. Vessel elements are considered more efficient conductors of water than tracheids due to the unobstructed flow of water through the perforation plates of the vessel elements [31, 32].

Although the morphological stages of secondary phloem and xylem development have been well documented, our understanding of the pathways involved in these processes is far from

complete. For example, xylogenesis is driven by the coordinate expression of numerous genes involved in differentiation, elongation, and programmed cell death [36]. However, little is known about the genes that regulate and coordinate these stages of development in xylem or phloem. Identification of these genes will open the door for more applied research, such as using developmental genes to improve wood strength and quality, or to improve wood utility for bioenergy.

Overview

At the outset of this research, we predicted that NAC domain proteins were likely to regulate vascular development in poplar because NAC domain proteins were known to play important roles in plant apical meristems [37-39], and it seemed likely that they also would play roles in vascular development. In support of this idea, preliminary blast searches identified several NAC domain proteins in wood and vascular cambium EST libraries from various *Populus* species (data not shown). In the last two years, several papers have reported that NAC domain proteins regulate aspects of vascular development in *Arabidopsis*, thus these findings further strengthened the underlying rationale for this dissertation project.

In chapter 2 we did a bioinformatic analysis of the *Arabidopsis*, poplar, and rice NAC domain family. A Bayesian analysis of the above NAC domain protein sequences produced 52 well-supported clades. Seventeen of these clades contained sequences from all three species and are predicted to include proteins with conserved functions in plants. Members of the remaining clades are predicted to fall into broad functional groups based on their genome of origin and that of the other proteins with which they cluster (if any). Relatively short stretches of conserved amino acids that might activate transcription, convey functional specificity, or affect

dimerization were identified using the MEME and MAST algorithms. In total 129 conserved motifs were identified. Seventy-seven of these motifs were present in at least one other NAC domain protein, while 23 of them were present in multiple members of other clades. NAC domain proteins in clades that share motifs are predicted to have related functions. This work was significant as it is the first study of the NAC domain family to include a complete set of poplar and rice NAC domain proteins, and it is the first to determine if variable region motifs in NAC domain proteins are shared between phylogenetic clades, as well as within them.

In chapter 3 we elucidated the role of a *Populus* NAC domain transcription factor, PtNAC068, in vascular development by analyzing PtNAC068 over-expression, knockdown, and dominant repression lines. PtNAC068 over-expression mutants displayed ectopic lignin deposition in leaf epidermal cells and primary vein pith cells, as well as an increase in lignin deposition in phloem fiber bundles and associated cells in leaves and stems. PtNAC068 dominant repressor mutants had a reduction in the size and number of phloem fiber bundles in primary leaf veins and stems, as well as delayed development of the sclereids associated with phloem fibers in stems. These findings indicated that PtNAC068 was a positive regulator of secondary wall development in phloem fibers and their associated sclereid cells in poplar. This was significant as it represents the first mutational analysis of a poplar NAC domain protein, the first use of the strong repression technology (EAR domain) in poplar, and represents the first transcription factor reported to be involved in phloem fiber development in trees.

References

1. Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M: **Genes Involved in Organ Separation in Arabidopsis: An Analysis of the cup-shaped cotyledon Mutant**. *The Plant Cell* 1997, **9**:841-857.
2. Kikuchi K, Ueguchi-Tanaka M, Yoshida KT, Nagato Y, Matsusoka M, Hirano H-Y: **Molecular Analysis of the NAC gene family in Rice**. *Mol Gen Genet* 2000, **262**:1047-1051.
3. Duval M, Hsieh T-F, Kim SY, Thomas TL: **Molecular characterization of AtNAM: a member of the *Arabidopsis* NAC domain superfamily**. *Plant Molecular Biology* 2002, **50**:237-248.
4. Ooka H, Satoh K, Doi K, Nagata T, Otomo Y, Murakami K, Matsubara K, Osato N, Kawai J, Piero C *et al*: **Comprehensive Analysis of NAC Family Genes in Oryza sativa and Arabidopsis thaliana**. *DNA Research* 2003, **10**:239-247.
5. Souer E, van Houwelingen A, Kloos D, Mol J, Koes R: **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* 1996, **85**:159-170.
6. Aida M, Ishida T, Tasaka M: **Shoot apical meristem and cotyledon formation during Arabidopsis embryogenesis: interaction among the CUP-SHAPED COTYLEDON and SHOOT MERISTEMLESS genes**. *Development* 1999, **126**:1563-1570.
7. Sablowski RWM, Meyerowitz EM: **A Homolog of NO APICAL MERISTEM Is and Immediate Target of the Floral Homeotic Genes APETALA3/PISTILLATA**. *Cell* 1998, **92**:93-103.

8. Takada S, Hibara K-i, Ishida T, Tasaka M: **The CUP-SHAPED COTYLEDON1 gene of Arabidopsis regulates shoot apical meristem formation**. *Development* 2001, **128**:1127-1135.

9. Vroemen CW, Mordhorst AP, Albrecht C, Kwaaitaal MACJ, de Vries SC: **The CUP-SHAPED COTYLEDON3 Gene Is Required for Boundary and Shoot Meristem Formation in Arabidopsis**. *The Plant Cell* 2003, **15**:1563-1577.

10. Xie Q, Frugis G, Colgan D, Chua N-H: **Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development**. *Genes and Development* 2000, **14**:3024-3036.

11. He XJ, Mu RL, Cao WH, Zhang ZG, Zhang JS, Chen SY: **AtNAC2, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development**. *Plant Journal* 2005, **44**(6):903-916.

12. Greve K, La Cour T, Jensen M, Poulsen F, Skriver K: **Interactions between plant RING-H2 and plant-specific NAC (NAM/ATAF1/2/CUC2) proteins: RING-H@ molecular specificity and cellular localization**. *Biochemistry Journal* 2003, **371**:97-108.

13. Xie Q, Guo H-S, Dallman G, Shengyun F, Weissman A, Chua N-H: **SINAT5 promotes ubiquitin-related degradation of NAC1 to attenuate auxin signals**. *Nature* 2002, **419**:167-170.

14. Kubo M, Udagawa M, Nishikubo N, Horiguchi G, Yamaguchi M, Ito J, Mimura T, Fukuda H, Demura T: **Transcription switches for protoxylem and metaxylem vessel formation**. *Genes & Development* 2005, **19**(16):1855-1860.

15. Ko JH, Yang SH, Park AH, Lerouxel O, Han KH: **ANAC012, a member of the plant-specific NAC transcription factor family, negatively regulates xylary fiber development in Arabidopsis thaliana**. *Plant Journal* 2007, **50**(6):1035-1048.

16. Mitsuda N, Iwase A, Yamamoto H, Yoshida M, Seki M, Shinozaki K, Ohme-Takagi M: **NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of Arabidopsis**. *Plant Cell* 2007, **19**(1):270-280.

17. Mitsuda N, Seki M, Shinozaki K, Ohme-Takagi M: **The NAC transcription factors NST1 and NST2 of Arabidopsis regulate secondary wall thickenings and are required for anther dehiscence**. *Plant Cell* 2005, **17**(11):2993-3006.

18. Zhong RQ, Demura T, Ye ZH: **SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of Arabidopsis**. *Plant Cell* 2006, **18**(11):3158-3170.

19. Zhong RQ, Richardson EA, Ye ZH: **Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of Arabidopsis**. *Planta* 2007, **225**(6):1603-1611.

20. Xie Q: **GRAB proteins, novel members of the NAC domain family, isolated by their interaction with a geminivirus protein**. *Plant Molecular Biology* 1999, **39**(4):647-656.

21. Ren T: **HRT gene function requires interaction between a NAC protein and viral capsid protein to confer resistance to turnip crinkle virus**. *Plant Cell* 2000, **12**(10):1917-1925.

22. Collinge M, Boller T: **Differential induction of two potato genes, Stprx2 and StNAC, in response to infection by Phytophthora infestans and to wounding**. *Plant Molecular Biology* 2001, **46**(5):521-529.

23. Delessert C, Kazan K, Wilson IW, Van Der Straeten D, Manners J, Dennis ES, Dolferus R: **The transcription factor ATAF2 represses the expression of pathogenesis-related genes in Arabidopsis**. *Plant Journal* 2005, **43**(5):745-757.

24. Lin RM, Zhao WS, Meng XB, Wang M, Peng YL: **Rice gene OsNAC19 encodes a novel NAC-domain transcription factor and responds to infection by Magnaporthe grisea**. *Plant Science* 2007, **172**(1):120-130.

25. Oh SK, Lee S, Yu S, Choi D: **Expression of a novel NAC domain-containing transcription factor (CaNAC1) is preferentially associated with incompatible interactions between chili pepper and pathogens**. *Planta* 2005, **222**(5):876-887.

26. Ohnishi T, Sugahara S, Yamada T, Kikuchi K, Yoshiba Y, Hirano HY, Tsutsumi N: **OsNAC6, a member of the NAC gene family, is induced by various stresses in rice**. *Genes & Genetic Systems* 2005, **80**(2):135-139.

27. Xue GP, Bower NI, McIntyre CL, Riding GA, Kazan K, Shorter R: **TaNAC69 from the NAC superfamily of transcription factors is up-regulated by abiotic stresses in wheat and recognises two consensus DNA-binding sequences**. *Functional Plant Biology* 2006, **33**(1):43-57.

28. Tran LSP, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Fujita M, Seki M, Shinozaki K, Yamaguchi-Shinozaki K: **Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter**. *Plant Cell* 2004, **16**(9):2481-2498.

29. Mauseth JD: **Botany an Introduction to Plant Biology**, 2nd edn. Philadelphia: Saunders College Publishing; 1995.

30. Berleth T, Mattsson J, Hardtke CS: **Vascular continuity and auxin signals**. *Trends in Plant Sciences* 2000a, **5**:387-393.

31. Taiz L, Zeiger E: **Plant Physiology 2nd Edition**. Sunderland, Massachusetts: Sinauer Associates Inc; 1998.

32. Raven P, Evert R, Eichhorn S: **Biology of Plants**, 6th edn. New York: Freeman Co.; 1999.

33. Beck C: **An introduction to Plant Structure and Development**: Cambridge University Press; 2005.

34. Scheres B, Dilaurenzio L, Willemsen V, Hauser MT, Janmaat K, Weisbeek P, Benfey PN: **MUTATIONS AFFECTING THE RADIAL ORGANIZATION OF THE ARABIDOPSIS ROOT DISPLAY SPECIFIC DEFECTS THROUGHOUT THE EMBRYONIC AXIS**. *Development* 1995, **121**(1):53-62.

35. Bonke M, Thitamadee S, Mahonen AP, Hauser MT, Helariutta Y: **APL regulates vascular tissue identity in Arabidopsis**. *Nature* 2003, **426**(6963):181-186.

36. Plomion C, Leprovost G, Stokes A: **Wood Formation in Trees**. *Plant Physiology* 2001, **127**:1513-1523.

37. Souer E, vanHouwelingen A, Kloos D, Mol J, Koes R: **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* 1996, **85**(2):159-170.

38. Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M: **Genes involved in organ separation in Arabidopsis: An analysis of the cup-shaped cotyledon mutant**. *Plant Cell* 1997, **9**(6):841-857.

39. Xie Q: **Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development**. *Genes & Development* 2000, **14**(23):3024-3036.

40. Kikuchi K, Ueguchi-Tanaka M, Yoshida KT, Nagato Y, Matsusoka M, Hirano HY: **Molecular analysis of the NAC gene family in rice**. *Molecular and General Genetics* 2000, **262**(6):1047-1051.

Figure 1.1. Alignment of NAC domain proteins from petunia (NAM.pro), *A. thaliana* (AtNAC1.pro), pine (pineNAC1.pro) and poplar (popNAC1.pro). NAC subdomains A-E, as defined by Kikuchi et al.[40], are denoted by labeled over-lines.

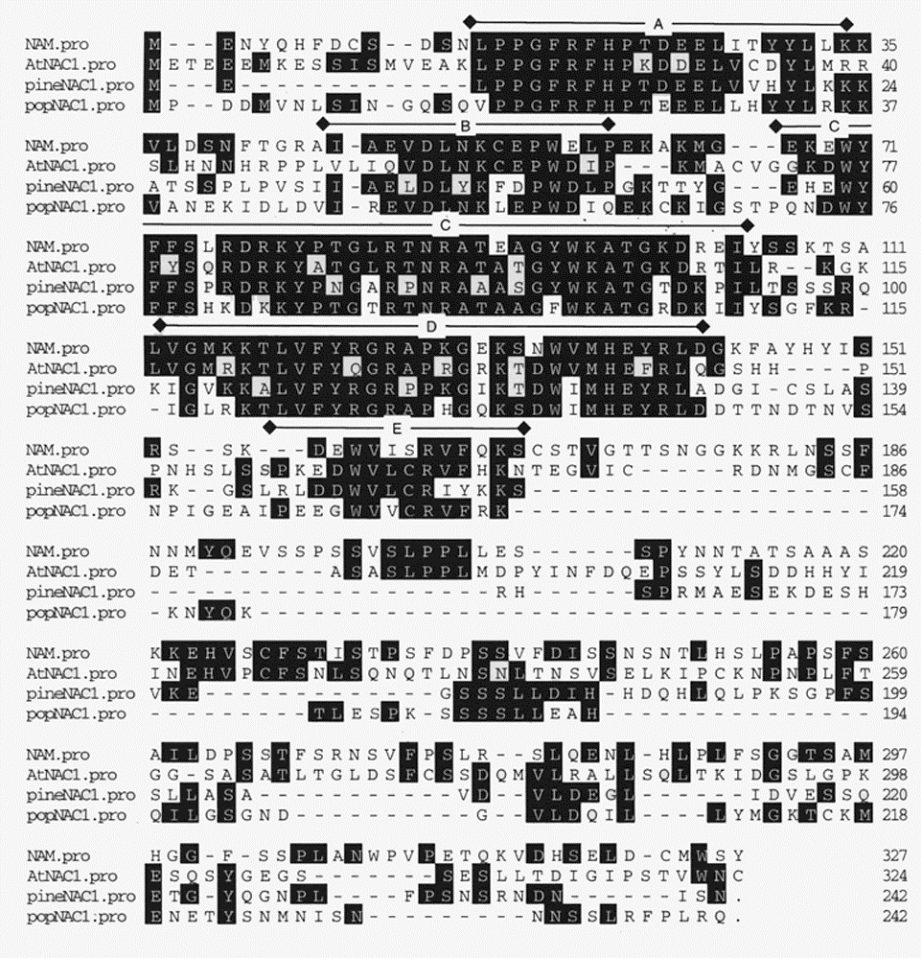


Figure 1.2. Cross-section of a woody stem. P=Phloem, VC=Vascular Cambium, Pi=Pith, B=Bark, R=Ray, and X=Xylem. The vascular cambium produces phloem to the outside and xylem to the inside.

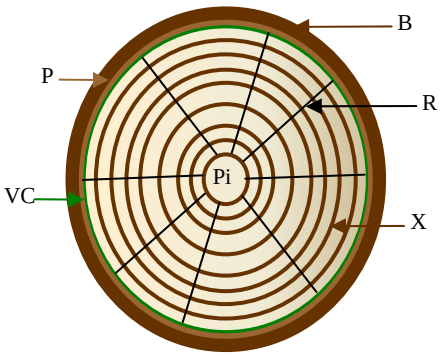
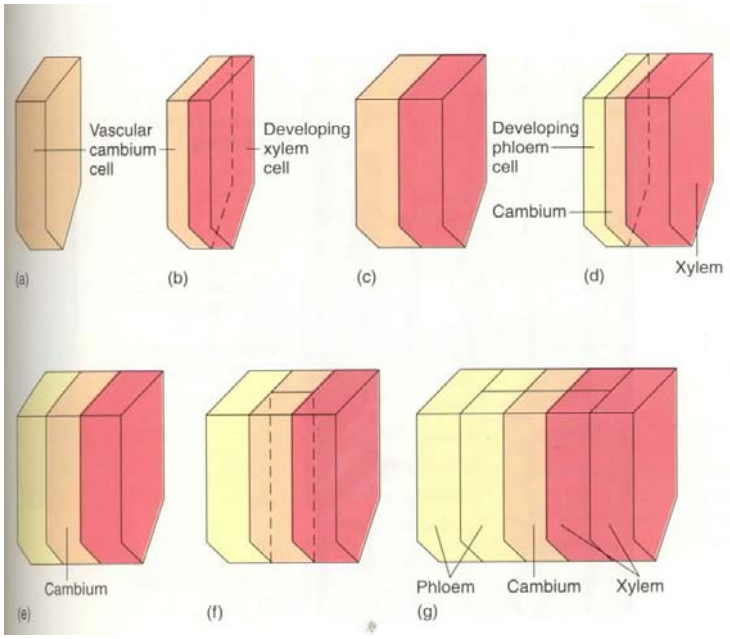


Figure 1.3. Cells produced from the vascular cambium. This figure is borrowed and adapted from [29]. (A) The lower half of a fusiform initial before division. (B) Division by a periclinal wall results in two thin cells. The outer cell remains a fusiform initial while the inner cell develops into secondary xylem. (C) Both cells enlarge to the size of the original cell. (D) The fusiform initial divides again; this time the outer cell matures as secondary phloem while the inner one remains a fusiform initial. (E) The cells grow back to the original size. (F) The fusiform initial divides with an anticlinal wall, resulting in two fusiform initials. (G) After the radial division in (F), a new row of cells is initiated in the secondary xylem and phloem.



Chapter 2

Bioinformatic Analysis of the *Arabidopsis*, Poplar, and Rice NAC Domain Transcription Factor Family [1]

---

[1]Kimberly Diane Hunt, Sarah Covert, and Russell Malmberg, to be submitted to BMC Evolutionary Biology.

Abstract

The NAC domain transcription factor family is found only in plants and is involved in important developmental and defense processes. A detailed understanding of the evolutionary relationships within the NAC domain family would facilitate the development of hypotheses about the functions of specific family members. To this end, we conducted several phylogenetic analyses of all *Arabidopsis*, poplar, and rice NAC domains. A Bayesian analysis of these sequences assembled them into 52 well-supported clades. Seventeen clades contained sequences from all three species and are predicted to include proteins with conserved functions in plants. Members of the remaining clades are predicted to fall into broad functional groups based on their genome of origin and that of the other proteins with which they cluster (if any). To identify relatively short stretches of conserved amino acids that might activate transcription, convey functional specificity, or affect dimerization, the MEME and MAST algorithms were used to identify motifs conserved within and between the clades defined by the phylogenetic analysis. In total, 129 conserved motifs were identified. Seventy-seven of them were present in at least one other NAC domain protein, while 23 of them were present in multiple members of other clades. NAC domain proteins in clades that share motifs are predicted to have related functions.

Introduction

NAC domain proteins are plant-specific transcription factors that regulate a variety of developmental and defense processes [1]. The "NAC domain" name was coined after the identification of a highly conserved sequence in the N-terminus of NO APICAL MERISTEM (NAM), ATAF1, ATAF2, and CUP-SHAPED COTYLEDON2 (CUC2) [2]. This conserved domain is typically 151-159 amino acids long and can be divided into 5 subdomains, A-E

(Figure 2.1) [3]. Two potential nuclear localization signals (NLS) were predicted to be present in subdomains C and D [3] and consistent with this, several NAC domain proteins have been shown to localize to the nucleus [4-10]. A DNA-binding domain is located within subdomains D and E [3, 11], and accordingly several NAC domain proteins have been shown to bind to DNA [4, 11-13]. This DNA-binding depends on homo- or heterodimer formation, and much of the dimerization ability has been mapped to subdomain A [4, 13, 14].

Unlike the N-terminal region, the C-terminal portion of NAC domain proteins is not well conserved. These "variable regions" can activate transcription [4, 5, 7-11, 13, 15-17], and like other plant activation domains, they frequently contain simple amino acid repeats and regions rich in serine and threonine, proline and glutamine, or acidic residues [1]. The variable region has also been reported to influence dimerization [13]. Because they are not conserved among all NAC domain proteins, the variable regions also are likely to account for the specific activities of different NAC domain family members.

NAC domain proteins are encoded by large gene families (i.e. $\geq$ 100 members/genome) in the higher plants. Functional analysis of such large protein families is impaired not only by their size, but also by the possibility of functional redundancy between paralogous proteins. However, inter-species comparisons of large protein families from whole genome sequences can provide a valuable framework for detailed functional studies by predicting evolutionary relationships and identifying conserved motifs of potential functional significance. A previous phylogenetic analysis of NAC domain proteins compared 75 sequences from a rice (*Oryza sativa*) cDNA database to 105 sequences from the *Arabidopsis* genome [18]. That study divided the NAC domain family into 2 main groups and 18 subgroups, six of which contained sequences from only one species or the other [18]. In addition, it identified 13 motifs in the variable

regions of certain subgroups [18]. Since the publication of this previous study, the rice genome and the poplar (*Populus trichocarpa*) genome have both been sequenced, allowing identification of 44 additional rice NAC domain proteins and 157 poplar NAC domain proteins. To provide a more detailed understanding of this family, we completed a phylogenetic analysis of all *Arabidopsis*, poplar, and rice NAC domain proteins. Doing so allowed us to identify orthologous and paralogous proteins, as well as family members that may direct functions specific to herbaceous annuals, woody perennials or monocots. In addition, a motif search identified many short, conserved sequences in different variable regions. These motifs may be used to predict the functions of uncharacterized NAC domain proteins and point towards portions of the variable regions that may be responsible for transcriptional activation or functional specificity.

Methods and Materials

Identifying NAC Domain Transcription Factors in *Arabidopsis*, Rice, and Poplar

Arabidopsis and rice NAC domain transcription factors were downloaded from the Ohio State University *Arabidopsis* Transcription Factor Database (http://*Arabidopsis*.med.ohio-state.edu/AtTFDB/), the Peking University Center for Bioinformatics Database of *Arabidopsis* Transcription Factors (http://datf.cbi.pku.edu.cn/browsefamily.php?familyname=NAC), the University of Potsdam Rice Transcription Factor Database (http://ricetfdb.bio.uni-potsdam.de/v2.0/) and the TIGR Rice Genome Annotation Database (http://www.tigr.org/tdb/e2k1/osa1/). Poplar NAC domain proteins were identified by using a NAC domain consensus sequence as the query in a blastp search of the *Populus trichocarpa* genome (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html), and by a keyword search of

the same database for all entries annotated as having the appropriate Pfam domain (#PF02365). After merging all gene lists by species, duplicates were eliminated and specific protein models were selected, if alternatives existed, on the basis of any available EST data and the presence of a predicted variable domain.

Sequence Manipulation and Tree Construction

Arabidopsis, rice, and poplar NAC domain proteins were initially aligned in Clustalx [19-21], then the variable domain was removed from each one using Seaview [22]. The trimmed NAC domain sequences were re-aligned in Clustalx [19-21] and the result analyzed by Bayesian analysis (prset aamodelpr=mixed; mcmcp Nrun=2 ngen=1500000 printfreq=400 samplefreq=200 nchains=4 burnin=500) in MrBayes [23, 24], maximum likelihood analysis (JTT matrix maximum likelihood estimates for transition/transversion and invariable sites) in PHYM [25], distance (neighbor joining and bootstrapping out of 100 and 1000 replicates) in PAUP [26], and maximum parsimony (heuristic search with random sequence addition and 100 tree limit hold at each stage) in PAUP [26].

Identification of Conserved Motifs in the Variable Regions of NAC domain proteins

All proteins in clades I-LII of the Mr.Bayes tree (Fig. 2) were aligned in ClustalX, and their NAC domains were removed using Seaview. All variable regions less than 18 amino acids long, or clades with a majority of variable regions less than 18 amino acids long were not analyzed further. To identify conserved motifs within members of a clade, the variable regions from each clade were submitted to MEME (http://meme.sdsc.edu/meme/intro.html) [27]. Motifs present in at least 50% of the sequences in each clade, or present in more than one species and

having an E value $\leq$ 0 were recorded. To determine if the identified motifs were present in any other NAC domain proteins, MAST (http://meme.sdsc.edu/meme/intro.html) [28] searched a database composed of all NAC domain proteins in *Arabidopsis*, poplar and rice with the MEME output from each clade.

Results

The structure of NAC domain proteins in *Arabidopsis*, Poplar, and Rice

One hundred *Arabidopsis* NAC domain proteins were collected from two different transcription factor databases. All members of this list were included in a previous study of *Arabidopsis* NAC domain proteins [18], except for 2 (ANAC106, ANAC107). The NAC domain proteins from *Arabidopsis* are identified herein by the ANAC### names assigned in the earlier paper, as well as by their *Arabidopsis* Gene Index codes (Table S2.1). Five of the 105 NAC domain proteins reported previously to be in *Arabidopsis* [18] represent duplicated gene models (ANAC022, ANAC035, ANAC039, ANAC080, AND ANAC091), thus we chose not to include them in this study.

One hundred thirty unique rice NAC domain proteins were collected from two transcription factor databases. Each database identified the Rice NAC domain proteins using LOC_Os identifiers (Table S2.3). However, the locus identifiers were derived from two different genome assemblies (TIGR and RAP-DB). All Rice Annotation Project Database (RAP-DB) identifiers were converted to TIGR locus OS_ identifiers. Duplicate protein models and protein models with incomplete NAC subdomains were removed. A subset of rice NACs were identified by ONAC### names assigned in an earlier paper [18]. We chose to build upon

this naming scheme for the rice NAC domain proteins, and thus assigned new ONAC### names to the 44 additional rice sequences included in this study (Table S2.3).

One hundred sixty-five potential NAC domain proteins were identified in the poplar genome after combining the results of a blastp search (maximum E value = 9e-06) with the results of a keyword search. All of the corresponding gene models, which were generated by automated annotation of the poplar genome, were inspected manually for the presence of a complete NAC domain, a potential C-terminal variable region, and the availability of supporting EST sequences. Eighty-eight of the gene models selected by the automated annotation were replaced by a human annotator with alternative models that were judged to be better, on the basis of their length and/or EST data. Of these, 157 sequences had complete NAC domains, four had partial NAC domains, and four were partial NAC domain proteins fused to other types of proteins. The 157 sequences containing complete NAC domains were analyzed further. All were arbitrarily assigned a name in the PtNAC### format to follow the paradigm established by [18] (Table S2.2).

The identified members of the NAC domain family vary considerably in length; in each species examined here the shortest proteins were 128-148 amino acids long, while the longest proteins were 649-726 amino acids long (Tables S2.1-S2.3). The variable region of NAC domain proteins ranged from less than 8 amino acids to 575 amino acids in length (Tables S2.1-S2.3). In some cases, the variable region constituted up to 79% of the total length of the NAC domain protein. Therefore, the overall variation in NAC domain protein length is due primarily to differences in the length of the C-terminal, variable region of each protein.

Phylogenetic Analysis of NAC Domains

To infer the evolutionary relationships among the NAC domain proteins in *Arabidopsis*, poplar, and rice, and to develop hypotheses about the functions of specific NAC domain proteins, several phylogenetic analyses were performed. These analyses were limited to the NAC domain of each protein because the high level of sequence divergence within the variable regions prevented their alignment. A phylogenetic tree of all NAC domains in *Arabidopsis,* poplar, and rice was produced with the program MrBayes [23, 24]. It included 52 clades, denoted hereafter with roman numerals that were supported by at least 0.51 credibility (Figure 2.2). The phylogenetic relationships amoung many of these clades could not be resolved with a significant level of statistical support (i.e. $\geq$ 0.50 credibility), so much of the tree contains a comb-like architecture when displayed as an unrooted phylogram. The composition of the 52 clades in the MrBayes tree was confirmed by maximum parsimony (PAUP) [26] and maximum likelihood (PHYM) [25] analysis. Although the architecture of each consensus tree was slightly different, the same proteins were clustered together in each one (data not shown).

Seventeen of the clades in Figure 2.2 include sequences from all three plant species; thus, they are likely to encode conserved functions in many plants. Examples of proteins with known functions in these three-species clades include VND6 and VND7 (clade XXIV), which are involved in *Arabidopsis* xylem development [29], NST1, NST2, and SND1/NST3/ANAC012 (clade XXV), which are involved in *Arabidopsis* secondary wall thickening [7, 15, 30, 31], NAC1 (clade XXVII), which is involved in *Arabidopsis* lateral root formation [4], and ATAF1, ANAC019, ANAC055, ANAC072, and ATAF2 (clade XXXVI), which are involved in drought stress or defense responses [10, 17, 32] (Figure 2.2).

The variable numbers of paralogous sequences in most of the three-species clades indicates that there has been considerable gene duplication and/or loss within these subfamilies since these plant species evolved as distinct lineages. For example, in clade XXVII, NAC1, a regulator of lateral root formation [4], has two orthologs in poplar and five in rice, and in clade XXXVIII, NAP, a regulator of floral development [33], has one ortholog in poplar and five in rice (Figure 2.2). Consistent with the idea that paralogous genes can have over lapping, partitioned or novel functions [34], some of the NAC domain paralogs in these clades are known to share redundant or partially redundant functions [17, 31, 35] (e.g., members of clades XXV and XXXVI), while others encode proteins with subtle differences in function [29] (e.g., members of clade XXIV).

Fifty-six of the NAC domains in Figure 2.2 are in clades composed of sequences from two plant species. All of these can be predicted to fall into broad functional groups on the basis of their genome of origin and that of the other proteins with which they clustered. Among the two-species clades, seven contain sequences from *Arabidopsis* and poplar (clades III-V, XVII, and XXX-XXXII), two contain sequences from poplar and rice (clades XI and LI), and 1 contains sequences from *Arabidopsis* and rice (clade XLVII). The 37 proteins in the *Arabidopsis* + poplar clades are predicted to encode dicot-specific functions. CUC1 and CUC2, which are both required for embryonic apical meristem function and cotyledon separation in *Arabidopsis* [2] are in one such clade (XXXI)(Figure 2.2). The seven proteins in the two poplar + rice clades are predicted to encode functions lost from *Arabidopsis*, while the 12 proteins in the *Arabidopsis* + rice clade are predicted to encode herbaceous annual-specific functions. To date, none of these 19 proteins have been characterized functionally. Of the 12 sequences in the *Arabidopsis* + rice

clade (XLVII), only one is from *Arabidopsis*, so this is a subfamily that has undergone considerable divergence in rice.

The remaining sequences, which are in single-species clades or are not clustered with other proteins, represent almost one-third (109/388) of all the NAC domains included in the analysis. As proteins that lack orthologs, they are likely to encode highly specialized functions, or functions lost from the other two species [36]. Among these species-specific NAC domains, 11 *Arabidopsis*, 43 poplar, and 28 rice sequences had a least one identifiable paralog, whereas eight *Arabidopsis*, seven poplar, and 12 rice NAC domains did not cluster with any other family members (Figure 2.2). Thus, some of these species-specific NAC domain sequences have undergone diversification via gene duplication, while others have not retained any duplicates that arose since each of these lineages evolved. Except for ANAC092 (clade XXIX), which is upregulated by $NaCl^-$ stress and induces lateral roots when over-expressed [9], none of these species-specific NAC domain proteins have been characterized.

After the phylogenetic analysis and the variable region motif analysis (see below) were completed, 11 additional NAC domain gene models were identified in the Universitaet-Postdam rice transcription factor database (http://ricetfdb.bio.uni-potsdam.de/v2.1/) (Table 2.1). To infer the evolutionary relationships among these proteins and those included in Figure 2.2, a maximum parsimony tree was made with the NAC domains from all three plant species (data not shown). Of the newly added rice gene models, two did not join any clade, three fell into the same rice-only cluster, and six were added to five different three-species clades (Table 2.1).

Classification of Orthologous Proteins

To gain further insight into the evolution of the NAC domain family and to facilitate future analyses of NAC domain family function, we categorized all of the *Arabidopsis*, poplar, and rice NAC domain proteins from clades I-LII into four ortholog classes, based on the number of potential orthologs in each genome (Figure 2.2, Tables S2.4-S2.6). Proteins with a 1:1 orthologous relationship (i.e., one *Arabidopsis* ortholog: one poplar ortholog) are predicted to have conserved functions across species boundaries, thus allowing inferences to be made about the function of any uncharacterized partners in these pairings. There were a total of seven 1:1 orthologous relationships between *Arabidopsis*:poplar, *Arabidopsis*:rice, and poplar:rice NAC domain proteins (Tables S2.4-S2.6). This classification applies to clade XXII in which NTL8, a mediator of salt responsive flowering in *Arabidopsis* [37], has a single uncharacterized ortholog in poplar, PtNAC084.

Proteins with a 1:n or n:1 relationship are predicted to have undergone functional diversification in one or the other species. This diversification could manifest itself as the acquisition of novel, useful functions by individual paralogs (neofunctionalization), the division of functions between paralogs (subfunctionalization), or the loss of function by one or more paralogs (nonfunctionalization) [34]. Clade XVI, in which the single *Arabidopsis* gene XND1 (ANAC104) has four poplar orthologs presents an interesting example of a 1:n relationship. XND1 is a regulator of vascular cell development in xylem [38], a physiological process that is more complex in a woody species with annual growth rings than it is in *Arabidopsis*.

Proteins with a n:n orthologous relationship have diversified unequally in both species since they evolved as distinct lineages. In these instances it can be difficult to identify orthologous pairs with confidence because gene duplications may have occurred after the two

plant species diverged from their common ancestor. Clades XXIV and XXV provide several examples of n:n relationships (Figure 2.2); both of these clades include *Arabidopsis* proteins that function in secondary cell wall development [7, 15, 29, 31, 35, 39] , but their orthologs in poplar and rice have not been characterized.

Analysis of Motifs in the Variable Region

Although the variable-region of NAC domain proteins (Figure 2.1) is not well conserved among all members of the family, short motifs in this region that are conserved within subfamilies might influence transcriptional activation, dimerization, or functional specificity. To identify such sequences, the variable regions from all clades with five or more members (Figure 2.2) were analyzed one clade at a time using MEME [27]. All motifs that were present in 50% or more of the sequences within a clade, or that were present in more than one species and had an E value ≤ 0 are reported in Table 2.2. Sequences meeting these criteria were judged likely to be functionally significant because of their presence in many members of a clade, or because of their conservation across species boundaries. In total, 129 conserved motifs were identified from the variable regions of 25 clades (Table 2.2).

In a preliminary MEME analysis with all of the clades in Figure 2.2, approximately 300 additional motifs were conserved in sequences present in clades composed of two, three, or four sequences. Most of these motifs appeared to be present in more than one sequence only because of a relatively recent gene duplication event (data not shown). They were rarely present in sequences from a second species (data not shown). As a consequence of these distribution patterns, it seemed unlikely that the conservation of these motifs was highly suggestive of functional significance and, therefore, they were not included in Table 2.2.

To determine if the motifs conserved within clades were present in other NAC domain subfamilies, the MAST algorithm [28] was used to search a database composed of all NAC domain proteins in *Arabidopsis*, poplar and rice with the MEME output from each clade. At least 77 of the motifs were present in at least one other sequence, and 23 were present in multiple members of another clade (Table 2.2). For example, clades XXIV, XXV and XXVI share two motifs (XXIV-M1/XXV-M1/XXVI-M1 and XXIV-M2/XXV-M3/XXVI-M6) that are present in 55-100% of the members in these three clades, but are present in very few sequences outside of these clades (Table 2.2). The sharing of conserved motifs between clades XXIV and XXV is supportive of the idea that variable region motifs contribute to the specific function of NAC domain proteins, because multiple members in each of these clades are involved in secondary wall development in *Arabidopsis* vascular tissues [7, 29, 31, 35, 39]. Further support for this idea comes from clade XVI, which is the only other clade of NAC domain proteins previously linked to vascular development. One of its members, XND1, negatively regulates lignocellulose synthesis and programmed cell death in xylem [38]. All members of clade XVI contain motif XVI-M1 (Table 2.2). The only other NAC domain proteins containing this motif are SND1/NST3/ANAC012, which is a member of clade XXV and is a regulator of secondary walls in phloem and xylem fibers [15, 35, 39, 40], and ANAC014, which is of unknown function.

In other cases, the motif analysis detected potential relationships among clades for which no functional information is available. For example, motif II-M1 is present in all eight members of clade II as well as all four members of clade III, while motif IV-M1 is present in all eight members of clade IV as well as all members of clades XVIII, XIX and XXXIV, and half the members of clade XVII (Table 2.2). Similarly, 54% and 77% of the 13 sequences in clade XXVIII contain motifs XXVIII-M1 and XXVIII-M2, respectively, as do 75% of the four

sequences in clade XXX (Table 2.2). It is possible that there are interactions or functional redundancies between members of the clades that share these motifs.

The distributions of other motifs in the NAC domain family are nearly clade-specific. Both of the clades that have been linked to defense and stress responses fit into this category. Clade XXI contains seven proteins of unknown function and TIP (Figure 2.2), which regulates the defense response of *Arabidopsis* to turnip crinkle virus [16, 41]. Of the seven potentially significant motifs detected in members of clade XXI, only XXI-M4 is found in a protein outside of the clade (Table 2.2). Clade XXXVI contains 15 sequences of unknown function and four *Arabidopsis* proteins involved in drought tolerance or other environmental stress responses [10, 17] (Figure 2.2). Of the nine potentially significant motifs detected in members of clade XXXVI, only XXXVI-M1 and XXXVI-M7 are detected in one or two other proteins outside of the clade (Table 2.2). Four other clades, VI, IX, XI, and XIII, contain motifs that are nearly clade-specific and/or nearly species-specific, but all four of these clades are composed entirely, or almost entirely, of sequences from a single species (Table 2.2). Thus, the distribution and apparent conservation of the motifs in these four clades may be reflective of recent subfamily expansion within a single species.

Of the 25 clades submitted to MEME, clades XXVII, XXXVIII, and XL lack motifs that are conserved in a majority of the sequences in each clade. Sequences in clade XXVII and XXXVIII are linked with apical and floral meristem development [4, 33, 42], but the motif analysis did not reveal any candidate motifs for these processes. All three of these clades are relatively large (11-26 sequences/clade) and diversification of the sequences within each clade may account for this lack of motif conservation. In support of this idea, clades XXVII and XL each contain well supported subfamilies (Figure 2.2).

Discussion

The NAC domain family encodes plant-specific transcription factors that are involved in many important developmental and defense responses. In order to provide a solid foundation for future functional and evolutionary studies of the NAC domain family we have applied a standard nomenclature to each member of the family and performed a genome wide phylogenetic analysis of the NAC domain protein family in *Arabidopsis*, poplar, and rice. The *Arabidopsis*, poplar, and rice Bayesian tree produced a comb-like architecture on which a majority of the clades are found. Several transcription factor family analyses [43-46] including a previous NAC domain analysis [31] have resulted in phylogenetic trees with a comb-like pattern. A comb usually means that the internal relationships of the clades on the comb were impossible to determine at the level of statistical significance employed by a given tree-building program. The addition of an appropriate outgroup might help to root the tree and add more resolution to the comb. Since the completion of this analysis, 32 NAC domain proteins have been identified in the *Physcomitrella patens* (moss) genome by PlnTFDB. Redoing this analysis with the *P. patens* NAC domain proteins could help to root the tree and add more resolution to the comb, but would require all subsequent motif analyses to be redone. Overall, the motif analysis was supportive of the evolutionary relationships detected by the phylogenetic study of NAC domains, because the distribution of the motifs largely followed the clade boundaries established by the phylogenetic analysis.

An earlier phylogenetic analysis of the NAC domain family was done using a complete set of *Arabidopsis* and a partial set of EST-supported rice NAC domain proteins [18]. This study coined the ANAC and ONAC naming scheme and divided the NAC family into 18 subgroups [18]. Twelve of these (SENU5, ANAC001, TIP, OsNAC8, ANAC011, NAC1, ONAC022,

TERN, ONAC001, ANAC001, ANAC063, and ONAC001) corresponded to strongly supported clades in our phylogenetic analysis of this protein family. Among these, clades VI and L remained composed of only *Arabidopsis* sequences, while clade XLIII remained composed of only rice sequences. This strengthens the earlier prediction that these proteins perform *Arabidopsis*-specific or monocot-specific functions [18], but none of these proteins have been functionally characterized to date. With the addition of 44 rice and 157 poplar NAC domain proteins to this analysis, five of the subgroups defined by Ooka et al. [18] (NAP, NAC2, OsNAC7, NAM, and ONAC003) were divided among multiple clades (IV, XVIII, XXIV–XXXI, XXXIII–XXXIV, XXXVII–XXXVIII), which contained proteins from 1-3 species. Other analyses involving a subset of NAC domain proteins [7, 15, 29, 30] reported the same phylogenetic relationships found herein for clades XXIV and XXV.

The *Arabidopsis*, poplar, and rice NAC domain protein family has undergone several rounds of whole genome duplications [47, 48]. The first genome duplication event predated the monocot-dicot divergence more than 300 million years ago [47]. The next genome duplication event occurred after the monocot dicot divergence 170 – 235 million years ago, and is referred to as the beta or eurosid duplication event. It covered approximately 59% of the genome [47, 48]. The most recent duplication events occurred after the divergence of *Arabidopsis* and poplar. This event in *Arabidopsis*, referred to as the alpha duplication event, covered 89% of the genome and occurred about 86 million years ago [47], while the most recent poplar duplication occurred 100-120 million years ago and covered 92% of the genome [48]. Even though *Arabidopsis* and poplar have gone through the same number of duplication events, poplar has retained more of the duplicated NAC domain proteins than *Arabidopsis* as there are 57 more poplar NAC domain proteins than *Arabidopsis* NAC domain proteins.

Following duplication events, duplicated genes undergo one of several evolutionary fates, including nonfunctionalization, neofunctionalization, or subfunctionalization [34, 36]. Clade XXV (NST) is likely to have been retained by subfunctionalization. Clade XXV contains three *Arabidopsis* (ANAC012, ANAC043, and ANAC066), four poplar (PtNAC061, PtNAC063, PtNAC065, and PtNAC068), and two rice (ONAC007/079 and ONAC029) NAC domain proteins. The three *Arabidopsis* genes (NST1/2/3) in clade XXV share some redundant function in secondary wall thickening [15, 30, 31, 35, 40].

The MEME analysis identified 129 conserved motifs in the variable regions of 25 clades (Table 2.2) that might activate transcription or convey functional specificity. A majority of the motifs (10/13) identified by Ooka et al. [18] also were found in this MEME analysis. Two of these motifs (motif ii and xii) were further refined by the inclusion of additional sequences reducing the length of the identified conserved amino acids to two or three which are possible targets for future transcriptional activation experiments.

Other motifs identified in this study were previously identified from the variable C terminal region of multiple NAC domain proteins. For example, clade XXXVI which comes from *Arabidopsis*, poplar or rice contains seven motifs (Clade XXXVI motif 2-6, 13, and 16) that are not shared with members in any other clade and one motif (Clade XXVI motif 1) that is shared with only ONAC153. Two of these clade XXXVI motifs (motif 1 and motif 4) have been identified previously in ATAF NAC domain proteins from *Arabidopsis*, rice, canola, and sugarcane [18, 49]. Currently the function of these motifs remains unknown.

None of the nine proteins in clade XXVI have been functionally characterized, but the majority of the sequences in this clade contain two motifs (XXIV-M1 and XXIV-M3) that are present in 82-100% of the sequences in clades XXIV and XXV (Table 2.2), both of which are

associated with vascular development in *Arabidopsis*. Thus, members of clade XXVI are predicted to regulate some aspect of vascular development.

Motifs responsible for transcriptional activation of associated NAC domain proteins could be determined by yeast transactivation analyses. Yeast transactivation analysis has been used to confirm the function of ANAC012 as a transcriptional activator and to identify an activation domain in the c-terminus of ANAC012 [15]. An activation domain was identified as the WQ-box motif which corresponds to the NST clade (clade XXV motif 1). The WQ box in clade XXV is shared with three other clades (XXIV, XXVI, XXXVI). Due to the absence of gene replacement techniques in plants, it will be difficult to assess directly the functional significance of the motifs present in the variable regions of NAC domain proteins. However, it may be possible to determine if certain motifs influence the functional specificity of individual proteins by fusing deletion constructs lacking individual motifs to the EAR repressor domain [39, 50]. The EAR motif has been used previously to create dominant, loss-of-function mutants in transcription factors, including several NAC domain proteins [7, 29, 39, 50]. If individual variable region motifs specify which gene targets are regulated by a given NAC domain protein, then the phenotypes of transgenic plants expressing a motif-deletion+EAR construct should differ from those expressing an intact protein+EAR construct.

References

1. Olsen AN, Ernst HA, Lo Leggio L, Skriver K: **NAC transcription factors: structurally distinct, functionally diverse**. *Trends in Plant Science* 2005, **10**(2):79-87.
2. Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M: **Genes involved in organ separation in *Arabidopsis*: An analysis of the cup-shaped cotyledon mutant**. *Plant Cell* 1997, **9**(6):841-857.
3. Kikuchi K, Ueguchi-Tanaka M, Yoshida KT, Nagato Y, Matsusoka M, Hirano HY: **Molecular analysis of the NAC gene family in rice**. *Molecular and General Genetics* 2000, **262**(6):1047-1051.
4. Xie Q: ***Arabidopsis* NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development**. *Genes & Development* 2000, **14**(23):3024-3036.
5. Fujita M, Fujita Y, Maruyama K, Seki M, Hiratsu K, Ohme-Takagi M, Tran LSP, Yamaguchi-Shinozaki K, Shinozaki K: **A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway**. *Plant Journal* 2004, **39**(6):863-876.
6. Xie Q, Guo HS, Dallman G, Fang SY, Weissman AM, Chua NH: **SINAT5 promotes ubiquitin-related degradation of NAC1 to attenuate auxin signals**. *Nature* 2002, **419**(6903):167-170.
7. Zhong RQ, Demura T, Ye ZH: **SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of *Arabidopsis***. *Plant Cell* 2006, **18**(11):3158-3170.

8. Greve K, La Cour T, Jensen MK, Poulsen FM, Skriver K: **Interactions between plant RING-H2 and plant-specific NAC (NAM/ATAF1/2/CUC2) proteins: RING-H2 molecular specificity and cellular localization**. *Biochemical Journal* 2003, **371**:97-108.

9. He XJ, Mu RL, Cao WH, Zhang ZG, Zhang JS, Chen SY: **AtNAC2, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development**. *Plant Journal* 2005, **44**(6):903-916.

10. Lu PL, Chen NZ, An R, Su Z, Qi BS, Ren F, Chen J, Wang XC: **A novel drought-inducible gene, ATAF1, encodes a NAC family protein that negatively regulates the expression of stress-responsive genes in *Arabidopsis***. *Plant Molecular Biology* 2007, **63**(2):289-305.

11. Duval M, Hsieh TF, Kim SY, Thomas TL: **Molecular characterization of AtNAM: a member of the *Arabidopsis* NAC domain superfamily**. *Plant Molecular Biology* 2002, **50**(2):237-248.

12. Ernst HA, Olsen AN, Skriver K, Larsen S, Lo Leggio L: **Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors**. *Embo Reports* 2004, **5**(3):297-303.

13. Hegedus D, Yu M, Baldwin D, Gruber M, Sharpe A, Parkin I, Whitwill S, Lydiate D: **Molecular characterization of Brassica napus NAC domain transcriptional activators induced in response to biotic and abiotic stress**. *Plant Molecular Biology* 2003, **53**(3):383-397.

14. Olsen AN, Ernst HA, Lo Leggio L, Skriver K: **DNA-binding specificity and molecular functions of NAC transcription factors**. *Plant Science* 2005, **169**(4):785-797.

15. Ko JH, Yang SH, Park AH, Lerouxel O, Han KH: **ANAC012, a member of the plant-specific NAC transcription factor family, negatively regulates xylary fiber development in *Arabidopsis* thaliana**. *Plant Journal* 2007, **50**(6):1035-1048.

16. Ren T: **HRT gene function requires interaction between a NAC protein and viral capsid protein to confer resistance to turnip crinkle virus**. *Plant Cell* 2000, **12**(10):1917-1925.

17. Tran LSP, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Fujita M, Seki M, Shinozaki K, Yamaguchi-Shinozaki K: **Isolation and functional analysis of *Arabidopsis* stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter**. *Plant Cell* 2004, **16**(9):2481-2498.

18. Ooka H, Satoh K, Doi K, Nagata T, Otomo Y, Murakami K, Matsubara K, Osato N, Kawai J, Carninci P *et al*: **Comprehensive analysis of NAC family genes in Oryza sativa and *Arabidopsis* thaliana**. *DNA Research* 2003, **10**(6):239-247.

19. Higgins DG, Sharp PM: **CLUSTAL - A PACKAGE FOR PERFORMING MULTIPLE SEQUENCE ALIGNMENT ON A MICROCOMPUTER**. *Gene* 1988, **73**(1):237-244.

20. Higgins DG, Sharp PM: **FAST AND SENSITIVE MULTIPLE SEQUENCE ALIGNMENTS ON A MICROCOMPUTER**. *Computer Applications in the Biosciences* 1989, **5**(2):151-153.

21. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools**. *Nucleic Acids Research* 1997, **25**(24):4876-4882.

22. Galtier N, Gouy M, Gautier C: **SEAVIEW and PHYLO_WIN: Two graphic tools for sequence alignment and molecular phylogeny**. *Computer Applications in the Biosciences* 1996, **12**(6):543-548.

23. Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F: **Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference**. *Bioinformatics* 2004, **20**(3):407-415.

24. Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP: **Evolution - Bayesian inference of phylogeny and its impact on evolutionary biology**. *Science* 2001, **294**(5550):2310-2314.

25. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood**. *Systematic Biology* 2003, **52**(5):696-704.

26. Swofford DL: **PAUP*. Phylogenetic Analysis Using Parsimony (*and other Methods).** In., 4 edn. Sunderland, Massachusetts: Sinauer Associates; 2003.

27. Bailey TL, Elkan C: **Fitting a mixture model by expectation maximization to discover motifs in biopolymers**. In: *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology: 1994; Menlo Park, California*: AAAI Press 1994: 28-36.

28. Bailey TL, Gribskov M: **Combining evidence using p-values: application to sequence homology searches**. *Bioinformatics* 1998, **14**(1):48-54.

29. Kubo M, Udagawa M, Nishikubo N, Horiguchi G, Yamaguchi M, Ito J, Mimura T, Fukuda H, Demura T: **Transcription switches for protoxylem and metaxylem vessel formation**. *Genes & Development* 2005, **19**(16):1855-1860.

30. Mitsuda N: **Functional analysis of NST transcription factors regulating secondary wall synthesis**. *Plant and Cell Physiology* 2007, **48**:S209-S209.

31. Mitsuda N, Seki M, Shinozaki K, Ohme-Takagi M: **The NAC transcription factors NST1 and NST2 of *Arabidopsis* regulate secondary wall thickenings and are required for anther dehiscence**. *Plant Cell* 2005, **17**(11):2993-3006.

32. Delessert C, Kazan K, Wilson IW, Van Der Straeten D, Manners J, Dennis ES, Dolferus R: **The transcription factor ATAF2 represses the expression of pathogenesis-related genes in *Arabidopsis***. *Plant Journal* 2005, **43**(5):745-757.

33. Sablowski RWM, Meyerowitz EM: **A homolog of NO APICAL MERISTEM is an immediate target of the floral homeotic genes APETALA3/PISTILLATA**. *Cell* 1998, **92**(1):93-103.

34. Prince VE, Pickett FB: **Splitting pairs: the diverging fates of duplicated genes**. *Nature Reviews Genetics* 2002, **3**(11):827-837.

35. Zhong RQ, Richardson EA, Ye ZH: **Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of *Arabidopsis***. *Planta* 2007, **225**(6):1603-1611.

36. Tian CG, Wan P, Sun SH, Li JY, Chen MS: **Genome-wide analysis of the GRAS gene family in rice and *Arabidopsis***. *Plant Molecular Biology* 2004, **54**(4):519-532.

37. Kim YS, Kim SG, Park JE, Park HY, Lim MH, Chua NH, Park CM: **A membrane-bound NAC transcription factor regulates cell division in *Arabidopsis***. *Plant Cell* 2006, **18**(11):3132-3144.

38. Zhao CS, Avci U, Grant EH, Haigler CH, Beers EP: **XND1, a member of the NAC domain family in *Arabidopsis* thaliana, negatively regulates lignocellulose synthesis and programmed cell death in xylem**. *Plant Journal* 2008, **53**(3):425-436.

39. Mitsuda N, Iwase A, Yamamoto H, Yoshida M, Seki M, Shinozaki K, Ohme-Takagi M: **NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of *Arabidopsis***. *Plant Cell* 2007, **19**(1):270-280.

40. Zhong RQ, Demura T, Ye ZH: ***Arabidopsis* NAC domain protein, SND1, is a key transcriptional activator regulating secondary cell wall biosynthesis in fibers**. *Plant and Cell Physiology* 2007, **48**:S111-S111.

41. Ren T, Qu F, Morris TJ: **The nuclear localization of the *Arabidopsis* transcription factor TIP is blocked by its interaction with the coat protein of Turnip crinkle virus**. *Virology* 2005, **331**(2):316-324.

42. Vroemen CW, Mordhorst AP, Albrecht C, Kwaaitaal M, de Vries SC: **The CUP-SHAPED COTYLEDON3 gene is required for boundary and shoot meristem formation in *Arabidopsis***. *Plant Cell* 2003, **15**(7):1563-1577.

43. Baum DA, Yoon HS, Oldham RL: **Molecular evolution of the transcription factor LEAFY in Brassicaceae**. *Molecular Phylogenetics and Evolution* 2005, **37**(1):1-14.

44. Parenicova L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B *et al*: **Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: New openings to the MADS world**. *Plant Cell* 2003, **15**(7):1538-1551.

45. Reyes JC, Muro-Pastor MI, Florencio FJ: **The GATA family of transcription factors in *Arabidopsis* and rice**. *Plant Physiology* 2004, **134**(4):1718-1732.

46. Zhao YH, Wang GY, Zhang JP, Yang JB, Peng S, Gao LM, Li CY, Hu JY, Li DZ, Gao LZ: **Expressed sequence tags (ESTs) and phylogenetic analysis of floral genes from a paleoherb species, Asarum caudigerum**. *Annals of Botany* 2006, **98**(1):157-163.

47. Bowers JE, Chapman BA, Rong JK, Paterson AH: **Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events**. *Nature* 2003, **422**(6930):433-438.

48. Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A *et al*: **The genome of black cottonwood, Populus trichocarpa (Torr. & Gray)**. *Science* 2006, **313**:1596-1604.

49. Nogueira FTS, Schlogl PS, Camargo SR, Fernandez JH, De Rosa VE, Pompermayer P, Arruda P: **SsNAC23, a member of the NAC domain protein family, is associated with cold, herbivory and water stress in sugarcane**. *Plant Science* 2005, **169**(1):93-106.

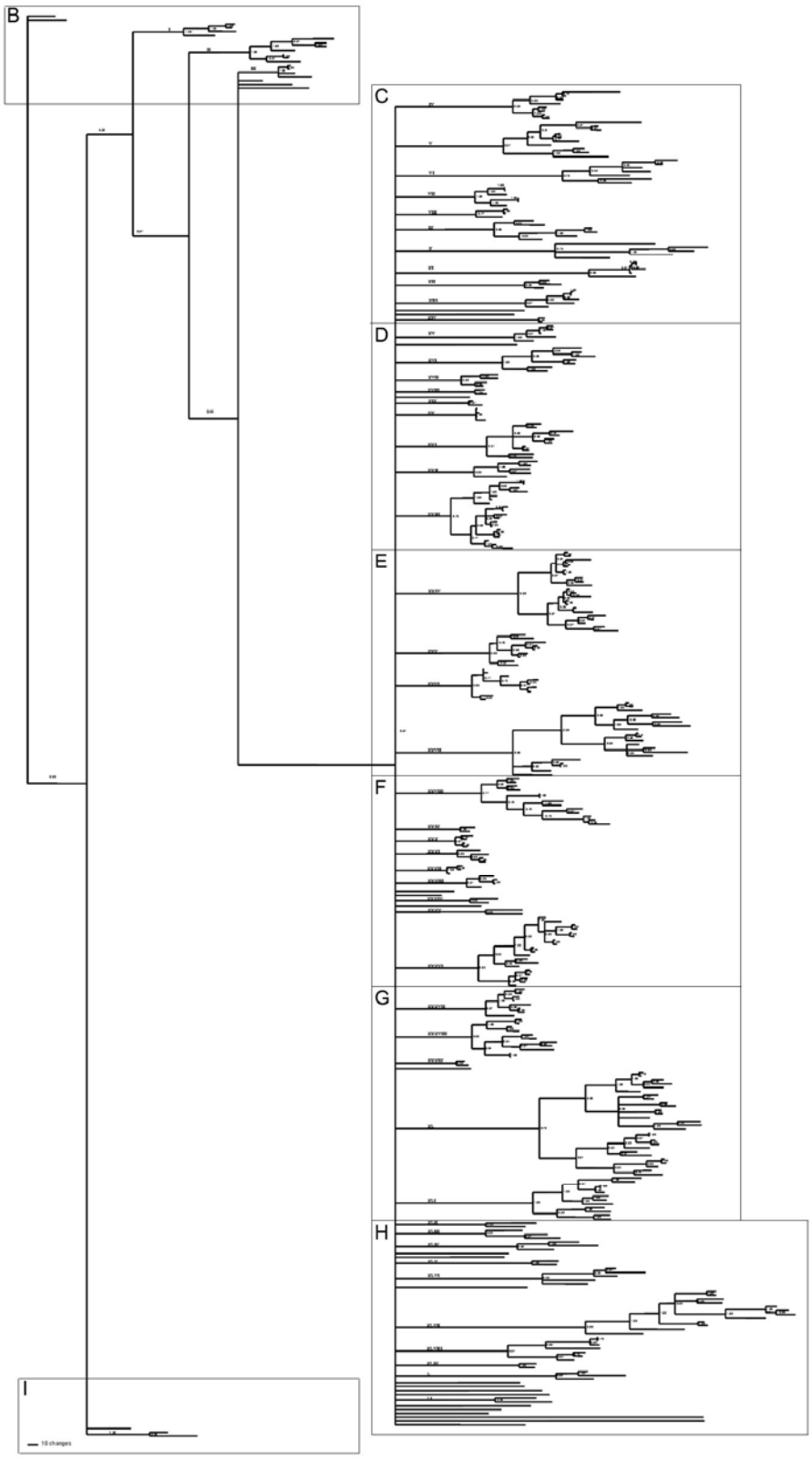
50. Hiratsu K, Matsui K, Koyama T, Ohme-Takagi M: **Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis***. *Plant Journal* 2003, **34**(5):733-739.

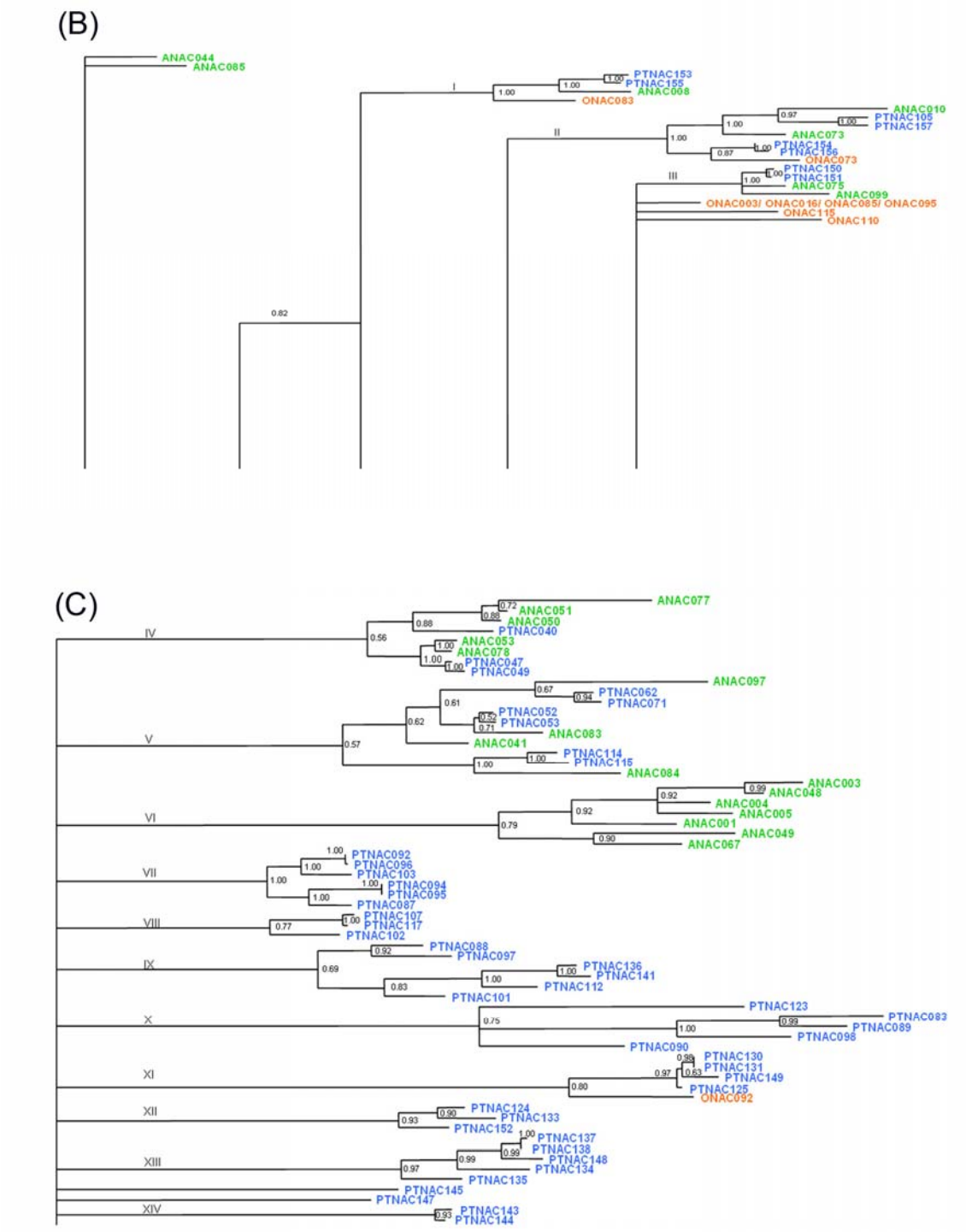
51. Bailey TL, Williams N, Misleh C, Li WW: **MEME: discovering and analyzing DNA and protein sequence motifs**. *Nucleic Acids Research* 2006, **34**:W369-W373.

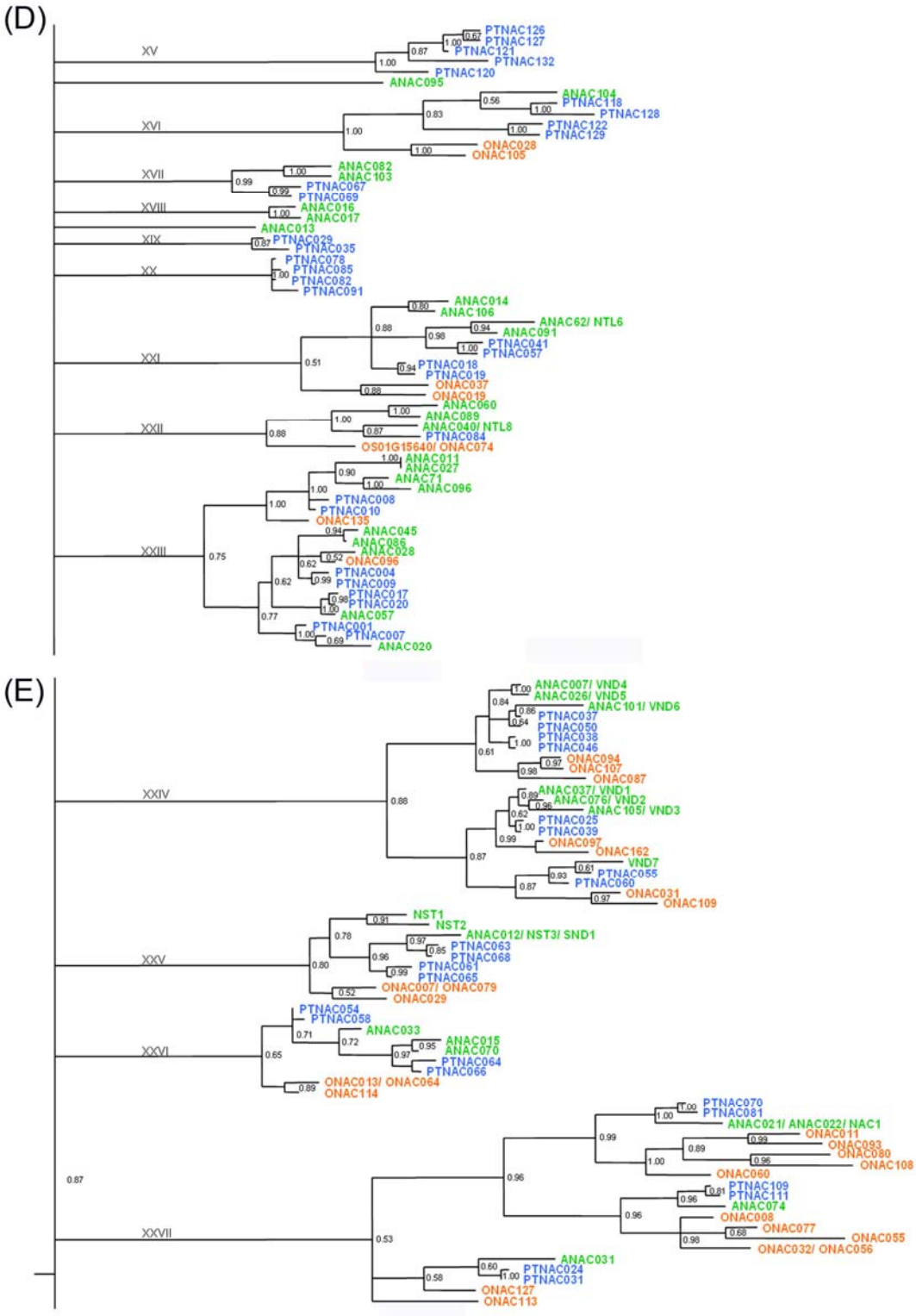
Figure 2.1. Typical NAC domain protein structure. Lettered boxes represent conserved sub-domains A-E.

Figure 2.2. Bayesian phylogenetic tree of *Arabidopsis*, poplar, and rice NAC domain proteins. (A) View of the whole consensus tree. Boxes with letters indicate areas of the tree enlarged in subsequent panels (B-I). Roman numerals indicate clade numbers. Arabic numerals indicate the frequency of each division in all trees used to form the consensus, with 1.00 equaling 100%. *Arabidopsis* proteins are green. Poplar proteins are blue. Rice proteins are orange.

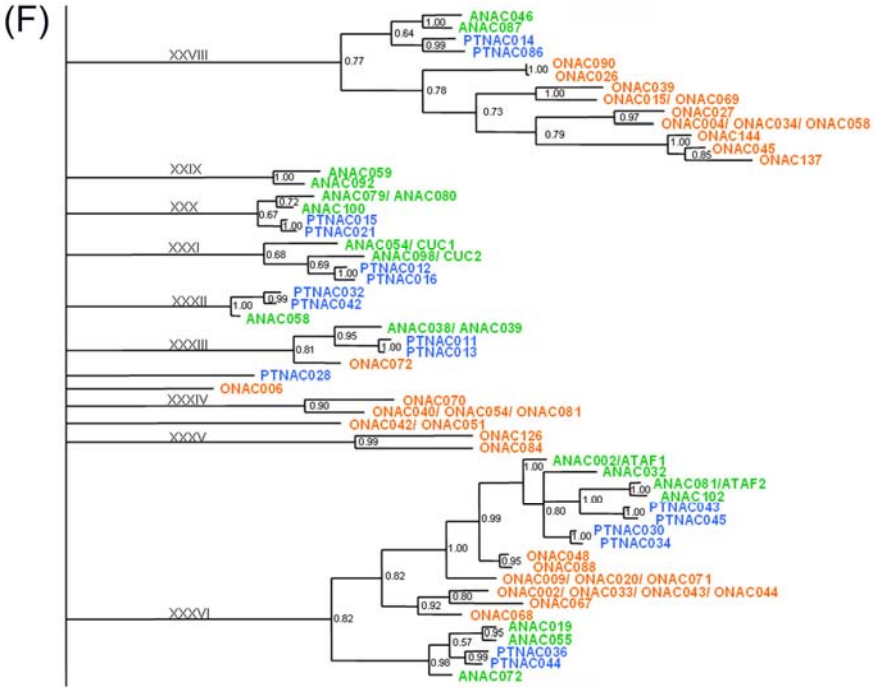
(A)



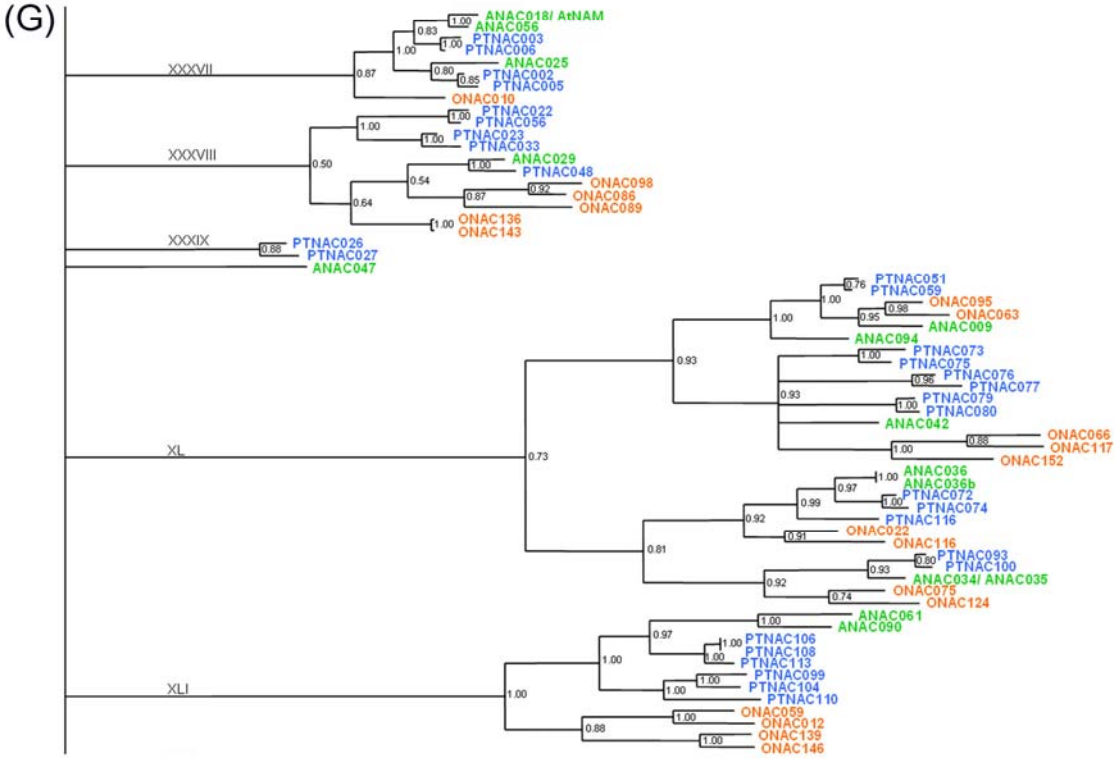

(B)
ANAC044
ANAC085
PTNAC153
PTNAC155
ANAC008
ONAC083
ANAC010
PTNAC105
PTNAC157
ANAC073
PTNAC154
PTNAC156
ONAC073
PTNAC150
PTNAC151
ANAC075
ANAC099
ONAC003/ ONAC016/ ONAC085/ ONAC095
ONAC115
ONAC110
I
II
III
0.82
(C)
ANAC077
ANAC051
ANAC050
PTNAC040
ANAC053
ANAC078
PTNAC047
PTNAC049
ANAC097
PTNAC062
PTNAC071
PTNAC052
PTNAC053
ANAC083
ANAC041
PTNAC114
PTNAC115
ANAC084
ANAC003
ANAC048
ANAC004
ANAC005
ANAC001
ANAC049
ANAC067
PTNAC092
PTNAC096
PTNAC103
PTNAC094
PTNAC095
PTNAC087
PTNAC107
PTNAC117
PTNAC102
PTNAC088
PTNAC097
PTNAC136
PTNAC141
PTNAC112
PTNAC101
PTNAC123
PTNAC083
PTNAC089
PTNAC098
PTNAC090
PTNAC130
PTNAC131
PTNAC149
PTNAC125
ONAC092
PTNAC124
PTNAC133
PTNAC152
PTNAC137
PTNAC138
PTNAC148
PTNAC134
PTNAC135
PTNAC145
PTNAC147
PTNAC143
PTNAC144
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
XIV


(D)
XV
XVI
XVII
XVIII
XIX
XX
XXI
XXII
XXIII
PTNAC126
PTNAC127
PTNAC121
PTNAC132
PTNAC120
ANAC095
ANAC104
PTNAC118
PTNAC128
PTNAC122
PTNAC129
ONAC028
ONAC105
ANAC082
ANAC103
PTNAC067
PTNAC069
ANAC016
ANAC017
ANAC013
PTNAC029
PTNAC035
PTNAC078
PTNAC085
PTNAC082
PTNAC091
ANAC014
ANAC106
ANAC62/ NTL6
ANAC091
PTNAC041
PTNAC057
PTNAC018
PTNAC019
ONAC037
ONAC019
ANAC060
ANAC089
ANAC040/ NTL8
PTNAC084
OS01G15640/ ONAC074
ANAC011
ANAC027
ANAC71
ANAC096
PTNAC008
PTNAC010
ONAC135
ANAC045
ANAC086
ANAC028
ONAC096
PTNAC004
PTNAC009
PTNAC017
PTNAC020
ANAC057
PTNAC001
PTNAC007
ANAC020
(E)
XXIV
XXV
XXVI
XXVII
ANAC007/ VND4
ANAC026/ VND5
ANAC101/ VND6
PTNAC037
PTNAC050
PTNAC038
PTNAC046
ONAC094
ONAC107
ONAC087
ANAC037/ VND1
ANAC076/ VND2
ANAC105/ VND3
PTNAC025
PTNAC039
ONAC097
ONAC162
VND7
PTNAC055
PTNAC060
ONAC031
ONAC109
NST1
NST2
ANAC012/ NST3/ SND1
PTNAC063
PTNAC068
PTNAC051
PTNAC065
ONAC007/ ONAC079
ONAC029
PTNAC054
PTNAC058
ANAC033
ANAC015
ANAC070
PTNAC064
PTNAC066
ONAC013/ ONAC064
ONAC114
PTNAC070
PTNAC081
ANAC021/ ANAC022/ NAC1
ONAC011
ONAC093
ONAC080
ONAC108
ONAC060
PTNAC109
PTNAC111
ANAC074
ONAC008
ONAC077
ONAC055
ONAC032/ ONAC056
ANAC031
PTNAC024
PTNAC031
ONAC127
ONAC113

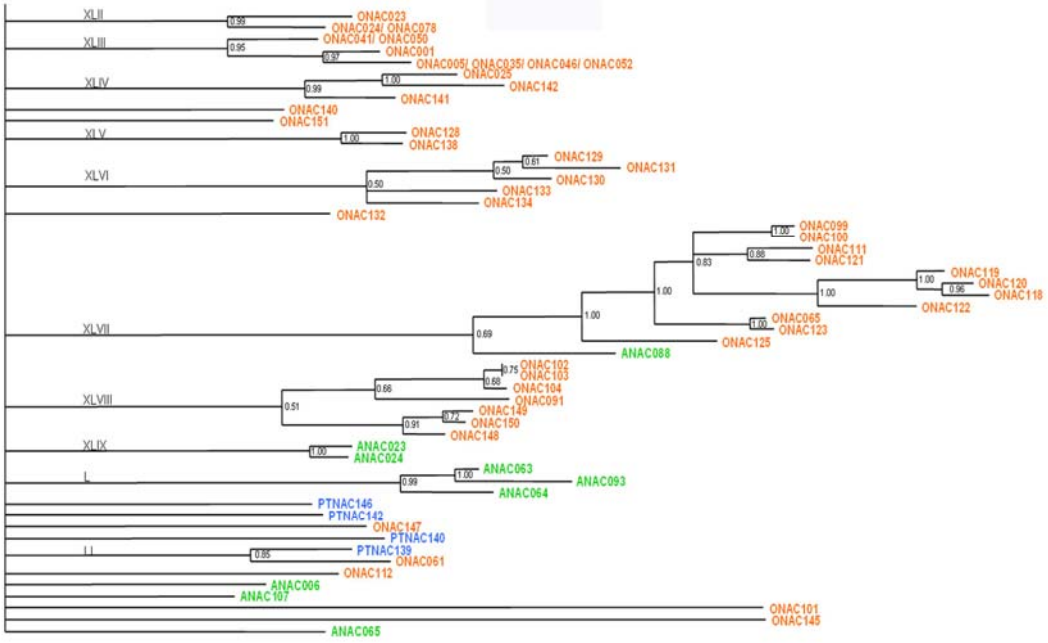
(F)


XXVIII
ANAC046
ANAC087
PTNAC014
PTNAC086
ONAC090
ONAC026
ONAC039
ONAC015/ ONAC069
ONAC027
ONAC004/ ONAC034/ ONAC058
ONAC144
ONAC045
ONAC137
XXIX
ANAC059
ANAC092
XXX
ANAC079/ ANAC080
ANAC100
PTNAC015
PTNAC021
XXXI
ANAC054/ CUC1
ANAC098/ CUC2
PTNAC012
PTNAC016
XXXII
PTNAC032
PTNAC042
ANAC058
XXXIII
ANAC038/ ANAC039
PTNAC011
PTNAC013
ONAC072
PTNAC028
ONAC006
XXXIV
ONAC070
ONAC040/ ONAC054/ ONAC081
ONAC042/ ONAC051
XXXV
ONAC126
ONAC084
XXXVI
ANAC002/ATAF1
ANAC032
ANAC081/ATAF2
ANAC102
PTNAC043
PTNAC045
PTNAC030
PTNAC034
ONAC048
ONAC088
ONAC009/ ONAC020/ ONAC071
ONAC002/ ONAC033/ ONAC043/ ONAC044
ONAC067
ONAC068
ANAC019
ANAC055
PTNAC036
PTNAC044
ANAC072

(G)


XXXVII
ANAC018/ AtNAM
ANAC056
PTNAC003
PTNAC006
ANAC025
PTNAC002
PTNAC005
ONAC010
XXXVIII
PTNAC022
PTNAC056
PTNAC023
PTNAC033
ANAC029
PTNAC048
ONAC098
ONAC086
ONAC089
ONAC136
ONAC143
XXXIX
PTNAC026
PTNAC027
ANAC047
XL
PTNAC051
PTNAC059
ONAC095
ONAC063
ANAC009
ANAC094
PTNAC073
PTNAC075
PTNAC076
PTNAC077
PTNAC079
PTNAC080
ANAC042
ONAC066
ONAC117
ONAC152
ANAC036
ANAC036b
PTNAC072
PTNAC074
PTNAC116
ONAC022
ONAC116
PTNAC093
PTNAC100
ANAC034/ ANAC035
ONAC075
ONAC124
XLI
ANAC061
ANAC090
PTNAC106
PTNAC108
PTNAC113
PTNAC099
PTNAC104
PTNAC110
ONAC059
ONAC012
ONAC139
ONAC146

(H)


XLII
ONAC023
ONAC024/ ONAC078
XLIII
ONAC041/ ONAC050
ONAC001
ONAC005/ ONAC035/ ONAC046/ ONAC052
XLIV
ONAC025
ONAC142
ONAC141
ONAC140
ONAC151
XLV
ONAC128
ONAC138
XLVI
ONAC129
ONAC131
ONAC130
ONAC133
ONAC134
ONAC132
ONAC099
ONAC100
ONAC111
ONAC121
ONAC119
ONAC120
ONAC118
ONAC122
ONAC065
ONAC123
ONAC125
XLVII
ANAC088
ONAC102
ONAC103
ONAC104
ONAC091
XLVIII
ONAC149
ONAC150
ONAC148
XLIX
ANAC023
ANAC024
L
ANAC063
ANAC093
ANAC064
PTNAC146
PTNAC142
ONAC147
PTNAC140
LI
PTNAC139
ONAC061
ONAC112
ANAC006
ANAC107
ONAC101
ONAC145
ANAC065

(I)

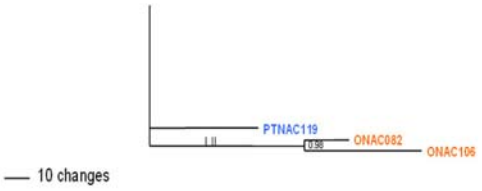

PTNAC119
LII
ONAC082
ONAC106
10 changes

Table 2.1. Recently Identified Rice NACs.

TIGR Locus Name[1]	ONAC Name[2]	Clade No.	Clade Composition[3]	Predicted Length (amino acids)	Predicted Gene Structure[4]
Os03g12120	ONAC153	None	N/A	887	
Os03g61249	ONAC154	XLVIII	R	317	
Os03g61319	ONAC155	XLVIII	R	317	
Os05g34600	ONAC156	XL	A, P, R	418	
Os05g35170	ONAC157	None	N/A	450	
				450	
				407	
				329	
				303	
Os05g48850	ONAC158	II	A, P, R	315	
Os09g38000	ONAC159	XVII	A, P, R	367	
Os09g38010	ONAC160	XVII	A, P, R	217	
Os09g12380	ONAC161	XLVIII	R	496	
Os10g38834	ONAC162	XXIV	A, P, R	342	
Os11g05614	ONAC163	XLI	A, P, R	885	

1 TIGR uses a similar convention for loci names as the one used for the *Arabidopsis* genome, with minor modification for the larger size of the rice genome. Each nuclear gene is labeled LOC_OsXXg##### with LOC_Os referring to Oryza sativa locus, XX referring to chromosome 01-12, g referring to gene, and a 5-digit number referring to the gene order on the chromosome (http://www.tigr.org/tdb/e2k1/osa1/tigr_gene_nomenclature.shtml).

2 ONAC naming scheme coined from Ooka et al 2003 paper. Red ONAC names represent NACs who were assigned ONAC names in this study.

3 A=*Arabidopsis*, P=Poplar, R=Rice

4 The predicted gene structures are from TIGR, are not to scale, and are found on the Rice Transcription Factor Database version 2.1 (http://ricetfdb.bio.uni-potsdam.de/v2.1/fam_mem.php?family_id=NAC). The red boxes indicate exons, the grey lines indicate introns, and the black boxes indicate UTR sequences.

Table 2.2. Conserved motifs in the variable regions of *Arabidopsis*, poplar and rice NAC domain proteins[a].

Clade # (members[b])	Motif name	# Seqs with motif[c]	Motif sequence[d]	Other motif locations[e]
II (2A 2R 4P)	II-M1	2A 2R 4P	VFYQTQPRQC[GS][SG][SL]I[KA][DA]	III ONAC097 ONAC003 ONAC115
IV (5A 3P)	IV-M1	5A 3P	[EG][QN][RY][GY]APF[ILM]EEEW[DEA][DE]D[EG][GV][AP]L[ILV]P	XVII XVIII XIX XXXIV ANAC013 ONAC042 ONAC115 ONAC157
	IV-M2	2A 2P	[RW][AG][MW]Lx[FL]M[CF][FL]WVL[IL]LSVS[FY]K[IV]	XXXIV ONAC042
	IV-M3	2A 2P	KYP[FL][IL]K[QK][AT]SH[MR]LGxIP[AT]Px[AS]xAS[EQ]F[PQ][ST]KDA	ANAC013 ONAC070
	IV-M4	2A 2P	[MT][EG]WS[FY][GD]KN[GE]N[LV][DN][VIL][IV]LS[FL]G[LV][PV]Q[QG]	ONAC070
	IV-M5	2A 2P	SS[GS]S[IV][HN]VTAG[MV][IM][RT]I[ES][ND][MS]	XXXIV ANAC016
	IV-M6	2A 2P	NYL[FIP][DG]E[SAP][FY][LM]D[AP][LN][NGS][ND]L[PL]x[NS][DE]G[LF][FY]LE[AT]NDLS	None
	IV-M7	2A 2P	[FV]EQPx[EDN][FST][SFY][ED]xDxKP[IMV][IL]R	None
	IV-M8	2A 1P	LC[VI][LV]N[KT]E[AT]P[LFS]PL[LIK][QY][YM]	None

	IV-M9	2A 1P	[IM][SV]V[EF]FE[TM][FLQ][KE][LR]E[MS]M[SK]A[EH][AM][ME]I[SN][IF]L[QE][SA][RQ][IV]D[AF]L[DNR][KQR][EK][NI][ED][DEH][LP][KH][KR]	None
	IV-M11	1A 1P	R[NT][ST]QDxCSSTxTT[AV]xxT[ST]xx[IM]xxxxxxAT[NT]TAISALLEFSLM	ONAC003 ONAC115 ONAC161
	IV-M12	2A 1P	D[DH][VA]Y[VL][DEH][MI][DN][DE][IL][DG][QE][KV]x[ET][NC] [FLI]V	None
	IV-M18	2A 1P	KR[RK][RK]x[NI]	ANAC064 ONAC115 ONAC151 ONAC153
	IV-M23	2A 2P	G[TS]GPKN	XVIII XIX ANAC013 ONAC042 ONAC070
V (4A 6P)	V-M1	2A 6P	PxF[FYM]DFMx[KRE]xx[TS]xxL	ANAC075 PtNAC072
	V-M2	1A 6P	S[GS][ILV]TE[VI]S[SC]NESD[DQ][EH][EQ]	ONAC045
VI (7A)	VI-M1	4A	KLG[QE]E[TS]x[KE][KN]KRA[GS]F[FV][HY]RMI[QH]x[FL]VKKIH	None
	VI-M2	5A	[IR][MVT][FSN]M[QKH][DQ]x[RY][SN][DN][HY][RT]Px[KN][SP][LV][TS]GV[FL]xD[DH][SV][ST][DS]	ONAC128
	VI-M3	4A	DSD[LP][IL][ST][PA][KT]R[NI]SIx[TN][SL]STCxSF[GA]SS[ND]	None
	VI-M5	4A	[IV][NQK]LVSL[TA]QEVS	None
VII (5P)	VII-M1	4P	[NY]Q[NFL][PQR][TA][QE]L[DEN][SH][PLF][RQ][GT][FY][DE]EG[CDE]Y[GS][LA][DNW][ST]A	PtNAC070

	VII-M2	3P	P[DH][AE][IN][MK]PT[YF]EE[GS]E[AS]S[FS]NV	XXIV ANAC078 PtNAC141
IX (6P)	IX-M1	5P	[YL]M[AP][SF]D[FS][EG]NQN[PL][NC][KE]	None
	IX-M3	6P	[VSG][EGS]K[SN]HHMA[FS][DNV]SEN	None
	IX-M4	4P	GYSSYF[NT]SSSSDN[DN]LADVALP	None
	IX-M5	5P	[FYL][IW]NQ[SP]T[VY][DG]	ONAC128
	IX-M6	4P	[GD]EG[VE]WS[NS][TL]	None
	IX-M9	4P	[QS][PI]EGE[CT]GPS[MV][EV][MI]P	None
X (5P)	X-M7	3P	Q[VI]PxLxET[FY][GE]xP[AV]P[MV]	PtNAC009
	X-M9	3P	DD[DEI][DEI]ED[YQ]	ONAC128 ONAC148
	X-M45	3P	PANS[QS]	Removed[f]
XI (1R 4P)	XI-M1	4P	KPS[VL]QEE[MI]ES[MI][RK][EK]Q[YH]SSRNDFEAGSSTN[FV]	None
	XI-M2	4P	WNNMQQ[LS]PPSPY[DH]Px[LY][PL][AP][PA][PL][SC]T[SG]SGHYYV[EN]QQE	PtNAC088
	XI-M3	4P	[HQ]PFPSLWSWTN	PtNAC120
	XI-M4	3P	GGQQQQEQT[SI]LPTNYEGYD[HQ]H	PtNAC079
XIII (5P)	XIII-M1	5P	[AV]S[YN][NE][GD][NGD][GE]G[VF][SC][PA][AT]GD[NT][AF][VY][TS][IP][EQ][ND][AM][AG][PL][AE][VA][SP]AI	ONAC077 PtNAC147
	XIII-M2	5P	[EPT]S[FIA][SN]E[NH][QP]Q[MI][AS][FP]	PtNAC145 PtNAC147 PtNAC152

	XIII-M3	5P	[SN][NIK][IV]I[EDK]N[QE][GE][IT][MT]GLA[DT][IF][SP]T[EQ]D[YS]FENAAA[NIT][VI]N	None
	XIII-M4	5P	[LM][KIL]E[STY][RL][KM][RN][KDE][ST][MNY][WD]	PtNAC147
	XIII-M5	4P	SSEVIQR[NT][IV]N[IL]DA[VA]INSGPTMAQ[SL][AT]H	PtNAC064 PtNAC128
	XIII-M6	5P	[IS]E[PA][VA][LF]xx[QY][PG][QR][QS]xN[FIN][IV]	None
	XIII-M7	4P	[PQ][IA]M[LP][LR][AE]EE	XLVIII
XVI (1A 2R 4P)	XVI-M1	1A 2R 4P	GTELSCLDE[VM]FLS[LM]DD	ANAC012 ANAC014
	XVI-M2	1A 2R 4P	[DS][EDQ][IV]S[FL][PH]	Removed
XXI (4A 2R 4P)	XXI-M1	4A 2R 2P	[FL][IVP][ANQ]QG[TS][AG]PRR[IL]RLQ	None
	XXI-M2	2A 2P	DS[PIM]Y[AS][SG]DF[GS][NY][DCN][QED][NIY]G[FL]xF[QL]D[GV][[TAS][SI]E[QP]D[VA]S[LI]T[ED][LV]L[DE][EDR][VFY][FL][NH]N[HP][DN]	None
	XXI-M3	3A 1R 2P	G[DNS][DG][AV][ES][GS][TR]GI[KT]I[RL][AR]R[QR]	None
	XXI-M4	2A 2P	[KR]xSPD[DN][AT][SS[DE][LM]VQET[AP][TS]S	None
	XXI-M6	2A 2P	[LD][DS][YFP][KET][LI]F[SP]P[LV]H[SEV]QVQ[SE][AE][LMV]GSSS[SFY][MN][FGT]x[QHP]	None
	XXI-M7	2A 2P	[GQY][IL]Q[TFS][QHN][NY][GV][TAG]N[ED][TAV]D[ADE][YD][TIM]S[DEK]F[LIV]DS[IF]L[KDQ]	None
	XXI-M8	1A 2P	L[GQ][VY]GSE[TG][QG][LA]xG[QW]x[PS][PD][GT]N[SF][YH]xxDx[GV][IQ][YQ]xM	None
XXII (3A 1R 1P)	XXII-M1	3A 1P	[QD][VAS][DP][NST][DE][ED]D[FC][YF]A[DE]IL[RKN]D[DE]I[EV][KN]LD[ED]	ONAC066 ONAC162

XXIII (9A 2R 8P)	XXIII-M1	3A 1R 2P	[FE]EVVE[EK][IV][KVNHG[LM][FL][VI]STRQ[AT]T[EK][TI][FL][FY]HQ[ILV]VPS QT[VL]K[IV][HY][IL]NPA	ONAC003 ONAC148
	XXIII-M2	2A 1R 4P	EC[AS][RMS]LQ[HY][QR][F[TP]LPPL[ER][VL]E[DN][FS]PQ	ONAC124
	XXIII-M5	4A 1R 8P	[AH][WT][IAV]S[PQ][DE][LF]I[LN][DG]SS[KFS][QC][GY]	None
	XXIII-M6	2A 1R 4P	D[KS][DE]D[SF][WS][ML][QG]FI[TN]D[DE][PA][WN]D	XVII ONAC125
	XXIII-M9	1A 1R 2P	[FK][KD][EAT][EK][RK[MLR]VENLR[WG]V[GK][MLV][SV][NS][NDK][ED]L[EG] [KE][SI][FV][VMT]E[EG][TDH]	ONAC097
	XXIII-M15	1R 2P	[IG]PKI[GI]EHYx[SV][TV][TH]xx[MV][PE][CQ][EP][HQ][SW]	ONAC161
	XXIII-M19	1A 2P	[YH][SG]T[IV][PS]YPPSKV[DN][IM]	None
	XXIII-M20	1A 2P	NGICSE[IL]E[ES][QE][GR]Q[CL][TQ]	ONAC126 ONAC159
	XXIII-M23	2A 2P	[QHK][IT][PW]K[KRT]K[EGK]E[KE][NE]x[GVA][DEN]E[EQ]	ANAC097 ONAC115
XXIV (7A 7R 8P)	XXIV-M1	7A 4R 8P	[DA]Q[VT]TDWR[VA]LDK[FL]VASQLSH[ES][ED]	XXV XXVI
	XXIV-M2	7A 3R 8P	[FL][LV][QE]LPQLESP[SK]L[PL][QL]S[AKS]R[PT]AS[SC]	XXV XXVI PtNAC044
	XXIV-M4	1A 1R 4P	[SK]K[QR][SE]ES[AV]QE[YN]ASTSTSSCQIDLWK	ANAC040 ONAC108 PtNAC144
	XXIV-M5	2A 2P	GVxSVx[DE]P[ILM][DN]Y[IV]x[KR]Qx[QH][NG][FIV][FL][AG]XX[FL][LM][CF]K QE[IL]EG	None
	XXIV-M7	2A 1R 2P	DxD[IL][GY][IR][CD][IVL][FH][EDV]	None

	XXIV-M8	3A 2P	[RP]xT[GI][QP]xKxx[EQ]xW	II
	XXIV-M10	2A 2R	CKxE[LV]E[LY]H[HY]	Removed
	XXIV-M11	2A 3R	W[YF]x[DE]xx[AS][FPV][MG]A	Removed
	XXIV-M12	2A 3R	[RL]xAx[VK][RAQ][RK][MA][AG][DG][DYG]	Removed
	XXIV-M14	1A 2P	[VI][SH][H[FS]L[DG]CFPD	None
	XXIV-M22	1R 4P	[LF][QLP][YN][NQT][MI]PHD	ANAC064 ANAC073 ANAC092
XXV (3A 2R 4P)	XXV-M1	3A 1R 4P	[DNS]W[AV][AT]LDRLVASQLNG	XXIV XXVI XXXVI
	XXV-M2	3A 2R 4P	[IL]D[QH]IL[EHLQ]YMGR[TS][CG]K	ONAC070 ONAC106
	XXV-M3	3A 2R 4P	[RGP]FM[KH]LP[SNPR]LESP	XXIV XXVI ONAC126
	XXV-M5	1A 4P	[TS][QH][DVE]Y[NTS][NSP]E[INM][DE]LW[NGS][FLT][TS][KTS][SR]S	ONAC070 ONAC156
	XXV-M6	2A 4P	KN[LY][QN]KTL[DE]SPK	ONAC040
	XXV-M11	1R 1P	RYLRP[IV]DT	ONAC083
	XXV-M12	3A 4P	D[PN][LN]CH[LV]S	None
XXVI (3A 2R 4P)	XXVI-M1	2A 2R 3P	G[DE]W[SA][FIM]LD[KR]L[LV][AT]SHQ[NG][LN][DE][QDH]	XXIV XXV
	XXVI-M2	1A 1R 2P	D[IV]EC[SP]QNL[LM][RK]LT	None

	XXVI-M3	2A 2P	[CQ][EP][PS]G[LV][ED]V[GC]TC[EK]	None
	XXVI-M5	2A 4P	[NL]Q[AST]R[GST]F[ML][HP][RE][DV]S[QP][YE]Q[HL][FR][SQ][HN][HM][KN]	None
	XXVI-M6	1A 2R 2P	[FLS][DEH][GAP]SM[HQ]LPQL[FAM]S[PA][DE][SAQ][AP][PVA][ACP][PAG]	XXIV XXV ONAC041 ONAC068 ONAC092 ONAC095
	XXVI-M7	2A 2P	[KV]N[LC][FN][KM]V[SGV][NGH][ED][GDV][GS][ST]x[ISA][MNS]	None
	XXVI-M8	1A 2P	[EM]MDFWG	XXV
XXVII (3A 11R 6P)	XXVII-M1	1A 2P	[RE][EG]V[VI][AC][KR][PD][SN][MI]xSCx[DN][DE]T[GA]S[SA]SLP[AP]L[ML]D[SP][YI[TN][FY][DE]Q[TE][QP]	ONAC128 PtNAC096
	XXVII-M5	1A 2P	[DH]HQT[LI][AP][CR][G[YW][QE]QM	None
XXVIII (2A 9R 2P)	XXVIII-M1	2A 3R 2P	E[MQ][FI][SGK][GS][EN]QS[LM][VL]S[ALV]SQ[DE]TGL[ST][ST]DVN[TA]T[AG][ET]I[SE]	XXX ANAC098 PtNAC028 ONAC042
	XXVIII-M2	2A 6R 2P	L[PD]P[LI][IM]DP[SP]x[YA]	XXX ANAC098 PtNAC028
	XXVIII-M5	1A 4R	[CDF][DNH][DML][DER][AG]x[LI]Wx[YF]	None
	XXVIII-M8	2R 2P	[IL][REG][QFY]CK	XX
	XXVIII-M16	1A 2P	[VP]V[SM]KQE[IM]G	None
XXXVI (7A 6R 6P)	XXXVI-M1	3A 3R 4P	[PK]R[LH][HT][TA]DSS[CG]SE[HQ]V[LV]SP[ES][EV][TS][CS][ES]	ONAC153
	XXXVI-M2	1A 3R 4P	Q[MDL][SP][PL]LQD[MI][FL]MY[LW][QG]KPF	None

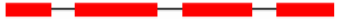
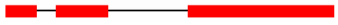
	XXXVI-M3	4A 6R 4P	KG[ST]IEKY	None
	XXXVI-M4	3A 2R 5P	VQS[EQ]PK[WI]	None
	XXXVI-M5	2A 4P	A[LF]D[FN][PN][FY]NY[IL]D[AD][TG][FM][DQ][DIN]	None
	XXXVI-M6	3A 2P	FGWA[STN][LF][AI]G[LNQ][NV][SEP]x[PN][ENQ][LS]	None
	XXXVI-M7	3A 2P	[STC][ST][SF]SS[SH][HQF]x[DE]DVL[DE][SP][LF][HPT][EQ][IE][DIK][DN][RNQ]	PtNAC029 ONAC153
	XXXVI-M13	1A 1R 2P	[YAN][GE][LF][RG][HY][SG][GIT][QC][PGQ][GS][SG]G[FY]G[FD]	None
	XXXVI-M16	2A 1R 2P	[LI]PRxN[SN]	None
XXXVII (3A 1R 4P)	XXXVII-M1	3A 1R 4P	[VIMN][LF][RP][QP][PK][FYR][QT][LI][PS][SG][LM][NY]W[YNT]S	PtNAC042 PtNAC156
	XXXVII-M2	4P	[HKQ][NQ][NP][GK][THP][LP][HS][LAV][AS]K[GHP]T[NAG][YF][AEG][SAT]LL[DEV]N[DE][DH][NT][FL]FEGI[IL]	None
	XXXVII-M3	4P	[NP]SS[SAG]KRF[HQ]G[DEG][IL]N[GS][DG][EI][TIS][GV][TV][[QGR][ENT]D[GN][NT]S[FS][VI][SAV][TLM]L	ONAC093
	XXXVII-M4	2A 1R 4P	DK[ED][DSM[ED][DG]M[LF]	ONAC120 ONAC151
	XXXVII-M5	2A 2P	[KT]R[ST]L[PC]x[HPQ][QY][YWR][WMP]	PtNAC038
	XXXVII-M6	4P	[QE][LF]PQ[ST][TSP][PQ]L[HV][HQ][PQ][ST][MT]L	ONAC075
	XXXVII-M7	1A 4P	N[NS][STP][QH]R[PQ][MIT][DM]	XXIV PtNAC067
XXXVIII (1A 5R 5P)	XXXVIII-M1	2R 2P	[HG][AN][FL][EL][NS][RLS]L[DK][HR][HK][DHP][GAV][EY][VRT][ND][NQ][QY][NR][VEG]N[CG][FL][PR][RP][KGS][KR][MK][LM][AR][CT][KS]A[DT][SIM][EF][DE][DE][GV][SV]	ONAC125
	XXXVIII-M10	1A 1P	MK[IL][PS][GR]TCSL[AS][HR]LLEM[DE]Y[LM]G	ONAC106

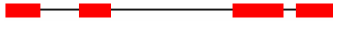
XL (5A 9R 12P)	XL-M1	1A 6P	[NY]GD[QD][FIL]F[DGT][DHN]WDELR[SPT][VM]V[ED][CFL]A	ONAC106
	XL-M3	1A 2R 2P	[TV][SP]LK[VE]LEQR[AV][AES][MA][EM][EA[EKL][MA][KNQR]	PtNAC116
	XL-M5	1R 2P	[FI][PN]F[SN]L[PQ][SG][NT]LxD[AD]W[KR][SM][NT]L[PA]W[DE]S[PL][SP]Cx[ST]E[MV]ST[TN][YF][SQ][TS][NT]KC	None
	XL-M6	1A 2R 2P	[NS][ST][MT][AT][QL]RAL[SC][HQ][ST][WF][VG]	None
	XL-M12	4R 2P	[QH]QQQQQFY	ANAC075 ANAC107 ONAC003 ONAC013 ONAC083 ONAC098 ONAC115 ONAC125 ONAC138 PtNAC151
	XL-M13	1A 2P	G[VI]EDHPS[LV]RPS[LI][PS][ST]R[AH]	None
	XL-M14	1A 2P	DDL[HQ]RL[ILV]NYQ[QI]	ONAC067
	XL-M22	7R 1P	AAAAAA	XXXVI ANAC075 ONAC002 ONAC003 ONAC013 ONAC073 ONAC092 ONAC098 ONAC106 ONAC112 ONAC115 ONAC125 ONAC156

XLI (2A 4R 6P)	XLI-M1	2A 4R 6P	[ST]GSFRAFDRRPL	ANAC076 ONAC073
	XLI-M3	2A 4P	[LD][ENS][EP]P[AI]W[ED]W[EP][QG][LQ]	XXXVI XL
XLVII (11R)	XLVII-M1	10R	D[AD][AS]DGADQ[SG][SC]SGV[IVM]	None
	XLVII-M2	10R	F[ED][FL]PES[IL]D[ED][VM][LV][SG][YC][IF]DF[ATS][TA][DMG][DA]	ANAC084
	XLVII-M3	11R	FSG[HY]GKKRKREP[EQD][SCR]	II ONAC136 PtNAC150
	XLVII-M4	7R	[VA]F[RH]DL[PA]D[LM]IVL[PQ][AP]E[EQ]	None
	XLVII-M5	8R	PQ[SI]A[VGM][SAT]ETA[LM][LFI]E[EQD]L[VAG][PL]PP[QA]PV	None
	XLVII-M7	6R	D[LQ][PT][GD][SR]I[DY][DE]D[ED]L[QS][SRV][FS][VLQ]	ONAC136
	XLVII-M11	7R	[PL]P[PT]AA[VLM]V[ND]	ANAC010 ONAC011
XLVIII (9R)	XLVIII-M6	7R	R[AK]A[DPK]D[SG][AV][AD]H[QP]E	ANAC056 ONAC145 PtNAC117

Table S2.1. *Arabidopsis* NAC domain genes.

Clade	Locus Name	ANAC Name[1]	Predicted Length (Amino Acids)	No. Exons	No. Introns	Predicted Gene Structure[2]
I	At1G25580	ANAC008	450	6	5	
II	At1G28470	ANAC010	315	3	2	
II	At4G28500	ANAC073	306	3	2	
III	At4G29230	ANAC075	499	6	5	
III	At5G56620	ANAC099	387	5	4	
IV	At3G10480	ANAC050	448	4	3	
IV	At3G10490	ANAC051/ ANAC052	239	4	3	
IV	At3G10500	ANAC053	550	6	5	
IV	At5G04400	ANAC077	396	4	3	
IV	At5G04410	ANAC078	568	6	5	
V	At2G33480	ANAC041	269	3	2	
V	At5G13180	ANAC083	253	3	2	
V	At5G14000	ANAC084	207	3	2	
V	At5G50820	ANAC097	185	3	2	
VI	At1G01010	ANAC001	430	6	5	
VI	At1G02220	ANAC003	395	6	5	
VI	At1G02230	ANAC004	580	7	6	
VI	At1G02250	ANAC005	351	6	5	
VI	At3G04420	ANAC048	343	6	5	
VI	At3G04430	ANAC049	199	4	3	
VI	At4G01520	ANAC067	303	4	3	
XVI	At5G64530	ANAC104	188	3	2	
XVII	At5G09330	ANAC082	490	4	3	
XVII	At5G64060	ANAC103	357	4	3	
XVIII	At1G34180	ANAC016	565	5	4	
XVIII	At1G34190	ANAC017	558	4	3	
XXI	At1G33060	ANAC014	649	7	6	
XXI	At3G49530	ANAC062	470	6	5	
XXI	At4G35580	ANAC106	513	6	5	
XXI	At5G24590	ANAC091	452	5	4	
XXII	At2G27300	ANAC040	336	4	3	

Clade	Locus Name	ANAC Name[1]	Predicted Length (Amino Acids)	No. Exons	No. Introns	Predicted Gene Structure[2]
XXII	At3G44290	ANAC060	336	4	3	
XXII	At5G22290	ANAC089	341	4	3	
XXIII	At1G32510	ANAC011	284	4	3	
XXIII	At1G54330	ANAC020	299	3	2	
XXIII	At1G64105	ANAC027	162	1	0	
XXIII	At1G65910	ANAC028	632	6	5	
XXIII	At3G03200	ANAC045	480	6	5	
XXIII	At3G17730	ANAC057	247	3	2	
XXIII	At4G17980	ANAC071	263	5	4	
XXIII	At5G17260	ANAC086	477	6	5	
XXIII	At5G46590	ANAC096	293	4	3	
XXIV	At1G12260	ANAC007	396	3	2	
XXIV	At1G62700	ANAC026	395	3	2	
XXIV	At1G71930	ANAC030	325	3	2	
XXIV	At2G18060	ANAC037	366	3	2	
XXIV	At4G36160	ANAC076	378	3	2	
XXIV	At5G62380	ANAC101	349	2	1	
XXIV	At5G66300	ANAC105	293	3	2	
XXV	At1G32770	ANAC012	359	3	2	
XXV	At2G46770	ANAC043	366	3	2	
XXV	At3G61910	ANAC066	335	2	1	
XXVI	At1G33280	ANAC015	306	3	2	
XXVI	At1G79580	ANAC033	372	3	2	
XXVI	At4G10350	ANAC070	342	3	2	
XXVII	At1G56010	ANAC021/ ANAC022	325	3	2	
XXVII	At1G76420	ANAC031	335	3	2	
XXVII	At4G28530	ANAC074	353	3	2	
XXVIII	At3G04060	ANAC046	339	3	2	
XXVIII	At5G18270	ANAC087	336	3	2	
XXIX	At3G29035	ANAC059	225	2	1	
XXIX	At5G39610	ANAC092	286	3	2	
XXX	At5G07680	ANAC079/ ANAC080	316	3	2	

Clade	Locus Name	ANAC Name[1]	Predicted Length (Amino Acids)	No. Exons	No. Introns	Predicted Gene Structure[2]
XXX	At5G61430	ANAC100	337	3	2	
XXXI	At3G15170	ANAC054	311	3	2	
XXXI	At5G53950	ANAC098	376	3	2	
XXXII	At3G18400	ANAC058	315	3	2	
XXXIII	At2G24430	ANAC038/ ANAC039	317	3	2	
XXXVI	At1G01720	ANAC002	290	3	2	
XXXVI	At1G52890	ANAC019	318	3	2	
XXXVI	At1G77450	ANAC032	254	3	2	
XXXVI	At3G15500	ANAC055	318	3	2	
XXXVI	At4G27410	ANAC072	298	3	2	
XXXVI	At5G08790	ANAC081	284	3	2	
XXXVI	At5G63790	ANAC102	313	3	2	
XXXVII	At1G52880	ANAC018	321	3	2	
XXXVII	At1G61110	ANAC025	324	3	2	
XXXVII	At3G15510	ANAC056	365	3	2	
XXXVIII	At1G69490	ANAC029	269	3	2	
XL	At1G26870	ANAC009	426	4	3	
XL	At2G02450	ANAC034/ ANAC035	380	4	3	
XL	At2G17040	ANAC036	251	3	2	
XL	At2G43000	ANAC042	276	3	2	
XL	At5G39820	ANAC094	335	3	2	
XLI	At3G44350	ANAC061	229	3	2	
XLI	At5G22380	ANAC090	236	3	2	
XLVII	At5G18300	ANAC088	148	1	0	
XLIX	At1G60280	ANAC023	348	1	0	
XLIX	At1G60350	ANAC024	321	1	0	
L	At3G55210	ANAC063	281	3	2	
L	At3G56530	ANAC064	320	3	2	
L	At5G39690	ANAC093	295	4	3	
	At1G03490	ANAC006	254	3	2	
	At1G32870	ANAC013	529	3	2	
	At3G01600	ANAC044	371	5	4	

Clade	Locus Name	ANAC Name[1]	Predicted Length (Amino Acids)	No. Exons	No. Introns	Predicted Gene Structure[2]
	At3G04070	ANAC047	360	3	2	
	At3G56560	ANAC065	229	3	2	
	At5G14490	ANAC085	351	4	3	
	At5G39540	ANAC107	197	2	1	
	At5G41090	ANAC095	213	2	1	

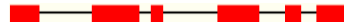
1 ANAC naming scheme coined from Ooka et al 2003 paper. Red ANAC names represent NACs who were assigned ANAC names in this study.

2 Predicted gene structures came from (http://*Arabidopsis*.med.ohio-state.edu/AtTFDB/ and http://datf.cbi.pku.edu.cn/browsefamily.php?familyname=NAC) and are not to scale. The red boxes indicate exons, the grey lines indicate introns, and the black boxes indicate UTR sequences.

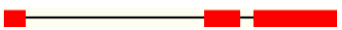
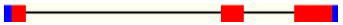
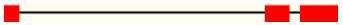
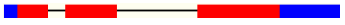
Table S2.2. Poplar NAC domain genes.

Clade	PtNAC Name	JGI Gene Model Name	Predicted Length (Amino Acids)	No. Exons	No. Introns	Predicted Gene Structure[1]
I	PtNAC153	eugene3.00101229	423	6	5	
I	PtNAC155	eugene3.00081085	429	6	5	
II	PtNAC105	eugene3.01240095	310	3	2	
II	PtNAC154	estExt_fgenesh4_pg.C_640203	299	3	2	
II	PtNAC156	estExt_fgenesh4_pg.C_LG_VII0110	299	3	2	
II	PtNAC157	fgenesh4_pg.C_LG_IV000476	313	3	2	
III	PtNAC150	estExt_fgenesh4_pm.C_1450025	453	6	5	
III	PtNAC151	eugene3.00061060	456	6	5	
IV	PtNAC040	estExt_Genewise1_v1.C_LG_X2865	441	4	2	
IV	PtNAC047	gw1.VIII.1502.1	565	8	7	
IV	PtNAC049	gw1.X.2890.1	551	6	5	
V	PtNAC052	estExt_fgenesh4_pm.C_LG_III0561	255	3	2	
V	PtNAC053	estExt_Genewise1_v1.C_LG_I7833	257	3	2	
V	PtNAC062	fgenesh4_pg.C_LG_XII001011	252	3	2	
V	PtNAC071	eugene3.00150808	227	3	2	
V	PtNAC114	estExt_Genewise1_v1.C_LG_I3228	244	3	2	
V	PtNAC115	eugene3.00880064	248	3	2	
VII	PtNAC092	eugene3.27120002	246	5	4	
VII	PtNAC094	eugene3.00140219	191	3	2	
VII	PtNAC095	eugene3.00140220	191	3	2	
VII	PtNAC096	eugene3.00140528	246	4	3	
VII	PtNAC103	eugene3.00021693	249	4	3	
VIII	PtNAC087	eugene3.00021434	208	4	3	

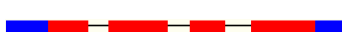
Clade	PtNAC Name	JGI Gene Model Name	Predicted Length (Amino Acids)	No. Exons	No. Introns	Predicted Gene Structure[1]
VIII	PtNAC102	gw1.II.957.1	128	2	1	
VIII	PtNAC107	fgenesh4_pg.C_scaffold_2712000001	165	3	2	
VIII	PtNAC117	eugene3.00140527	464	6	5	
IX	PtNAC088	eugene3.00021695	563	5	4	
IX	PtNAC097	eugene3.00140529	329	4	3	
IX	PtNAC101	eugene3.00021432	453	4	3	
IX	PtNAC112	eugene3.00021433	584	3	2	
IX	PtNAC136	estExt_fgenesh4_pg.C_LG_XIV0205	619	7	6	
IX	PtNAC141	eugene3.00140221	649	6	5	
X	PtNAC083	eugene3.00011678	442	3	2	
X	PtNAC089	eugene3.00091123	514	3	2	
X	PtNAC090	fgenesh4_pg.C_LG_XVIII000824	433	3	2	
X	PtNAC098	fgenesh4_pg.C_LG_IX001064	387	3	2	
X	PtNAC123	fgenesh4_pg.C_LG_XVI000432	646	4	3	
XI	PtNAC125	fgenesh4_pg.C_scaffold_64000081	254	3	2	
XI	PtNAC130	eugene3.00640092	254	3	2	
XI	PtNAC131	fgenesh4_pg.C_scaffold_64000087	254	3	2	
XI	PtNAC149	fgenesh4_pg.C_LG_VII000185	259	3	2	
XII	PtNAC124	fgenesh4_pg.C_scaffold_120000085	350	5	4	
XII	PtNAC133	fgenesh4_pg.C_scaffold_120000084	346	5	4	
XII	PtNAC152	eugene3.00130793	173	4	3	
XIII	PtNAC134	fgenesh4_pg.C_scaffold_379000002	343	1	0	
XIII	PtNAC135	eugene3.00060270	417	1	0	

Clade	PtNAC Name	JGI Gene Model Name	Predicted Length (Amino Acids)	No. Exons	No. Introns	Predicted Gene Structure[1]
XIII	PtNAC137	eugene3.42540002	402	1	0	
XIII	PtNAC138	fgenesh4_pg.C_scaffold_379000005	432	1	0	
XIII	PtNAC148	eugene3.23360001	425	1	0	
XIV	PtNAC143	eugene3.00040900	317	1	0	
XIV	PtNAC144	fgenesh4_pg.C_LG_V000226	392	1	0	
XV	PtNAC120	fgenesh4_pg.C_LG_XV000068	335	6	5	
XV	PtNAC121	fgenesh4_pg.C_LG_XII000282	173	3	2	
XV	PtNAC126	fgenesh4_pg.C_LG_XII000280	172	3	2	
XV	PtNAC127	fgenesh4_pg.C_LG_XII000281	171	3	2	
XV	PtNAC132	eugene3.00041119	149	2	1	
XVI	PtNAC118	fgenesh4_pg.C_LG_III000116	192	3	2	
XVI	PtNAC122	fgenesh4_pm.C_LG_V000020	195	3	2	
XVI	PtNAC128	fgenesh4_pm.C_LG_I000514	192	3	2	
XVI	PtNAC129	fgenesh4_pm.C_LG_VII000145	197	3	2	
XVII	PtNAC067	eugene3.00700014	403	2	1	
XVII	PtNAC069	eugene3.00070370	432	2	1	
XIX	PtNAC029	eugene3.00020577	609	5	4	
XIX	PtNAC035	estExt_Genewise1_v1.C_LG_V0549	570	6	5	
XX	PtNAC078	gw1.212.127.1	151	1	0	
XX	PtNAC082	eugene3.02120017	157	3	2	
XX	PtNAC085	gw1.212.47.1	151	1	0	
XX	PtNAC091	eugene3.19920001	141	4	3	
XXI	PtNAC018	gw1.I.5633.1	537	5	4	
XXI	PtNAC019	gw1.XI.837.1	429	5	4	
XXI	PtNAC041	estExt_Genewise1_v1.C_LG_XII1196	316	4	3	

Clade	PtNAC Name	JGI Gene Model Name	Predicted Length (Amino Acids)	No. Exons	No. Introns	Predicted Gene Structure[1]
XXI	PtNAC057	eugene3.00150047	382	4	3	
XXII	PtNAC084	eugene3.00090087	313	4	3	
XXIII	PtNAC001	fgenesh4_pg.C_LG_X001573	344	4	3	
XXIII	PtNAC004	fgenesh4_pg.C_LG_IV000766	680	6	5	
XXIII	PtNAC007	fgenesh4_pg.C_LG_VIII000707	340	4	3	
XXIII	PtNAC008	fgenesh4_pm.C_scaffold_29000078	322	4	3	
XXIII	PtNAC009	fgenesh4_pg.C_scaffold_44000152	677	6	5	
XXIII	PtNAC010	fgenesh4_pm.C_scaffold_118000015	~322	4	3	
XXIII	PtNAC017	gw1.122.151.1	230	3	2	
XXIII	PtNAC020	gw1.XII.1246.1	230	3	2	
XXIV	PtNAC025	fgenesh4_pm.C_LG_VII000469	169	3	2	
XXIV	PtNAC037	fgenesh4_pg.C_LG_XII001280	370	3	2	
XXIV	PtNAC038	fgenesh4_pg.C_scaffold_127000044	356	3	2	
XXIV	PtNAC039	gw1.57.170.1	346	3	2	
XXIV	PtNAC046	fgenesh4_pm.C_LG_III000335	357	3	2	
XXIV	PtNAC050	fgenesh4_pg.C_LG_XV000986	364	3	2	
XXIV	PtNAC055	eugene3.00410078	319	3	2	
XXIV	PtNAC060	gw1.XIX.2092.1	317	3	2	
XXV	PtNAC061	fgenesh4_pg.C_LG_II001646	371	4	3	
XXV	PtNAC063	estExt_fgenesh4_pg.C_LG_I3128	419	3	2	
XXV	PtNAC065	gw1.XIV.1905.1	379	3	2	
XXV	PtNAC068	eugene3.00111299	423	3	2	
XXVI	PtNAC054	eugene3.00101697	318	3	2	

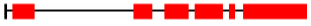
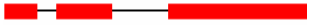
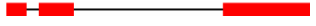
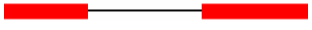
Clade	PtNAC Name	JGI Gene Model Name	Predicted Length (Amino Acids)	No. Exons	No. Introns	Predicted Gene Structure[1]
XXVI	PtNAC058	estExt_fgenesh4_pg.C_LG_VIII0685	319	3	2	
XXVI	PtNAC064	gw1.XIII.3189.1	337	4	3	
XXVI	PtNAC066	eugene3.01250035	368	3	2	
XXVII	PtNAC024	gw1.V.2219.1	179	3	2	
XXVII	PtNAC031	eugene3.00020050	404	3	2	
XXVII	PtNAC070	estExt_fgenesh4_pg.C_LG_VII0796	302	3	2	
XXVII	PtNAC081	estExt_fgenesh4_pg.C_LG_V0378	304	3	2	
XXVII	PtNAC109	fgenesh4_pm.C_LG_II000178	281	3	2	
XXVII	PtNAC111	eugene3.00051326	286	3	2	
XXVIII	PtNAC014	grail3.0065001201	358	3	2	
XXVIII	PtNAC086	gw1.XIII.1922.1	262	4	3	
XXX	PtNAC015	fgenesh4_pm.C_LG_XII000069	359	3	2	
XXX	PtNAC021	eugene3.00150202	359	3	2	
XXXI	PtNAC012	gw1.XI.3770.1	301	3	2	
XXXI	PtNAC016	eugene3.01070008	324	3	2	
XXXII	PtNAC032	eugene3.00150297	322	3	2	
XXXII	PtNAC042	eugene3.00120476	366	4	3	
XXXIII	PtNAC011	grail3.0193000901	312	3	2	
XXXIII	PtNAC013	fgenesh4_pm.C_scaffold_147000020	317	4	3	
XXXVI	PtNAC030	estExt_fgenesh4_pg.C_LG_II0726	~292	3	2	
XXXVI	PtNAC034	estExt_fgenesh4_pm.C_LG_V0368	279	3	2	
XXXVI	PtNAC036	estExt_Genewise1_v1.C_LG_XI3994	342	3	2	
XXXVI	PtNAC043	eugene3.00050086	307	3	2	
XXXVI	PtNAC044	eugene3.01070092	344	3	2	
XXXVI	PtNAC045	grail3.0011008901	304	3	2	

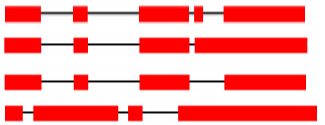
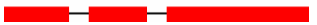
Clade	PtNAC Name	JGI Gene Model Name	Predicted Length (Amino Acids)	No. Exons	No. Introns	Predicted Gene Structure[1]
XXXVII	PtNAC002	estExt_fgenesh4_pg.C_LG_XI0289	335	4	3	
XXXVII	PtNAC003	grail3.0014014001	355	3	2	
XXXVII	PtNAC005	grail3.0093008801	351	3	2	
XXXVII	PtNAC006	grail3.0107006501	360	3	2	
XXXVII	PtNAC022	eugene3.01870003	337	3	2	
XXXVIII	PtNAC023	eugene3.00280154	338	3	2	
XXXVIII	PtNAC033	eugene3.00160810	314	3	2	
XXXVIII	PtNAC048	eugene3.00101577	283	3	2	
XXXVIII	PtNAC056	eugene3.00040458	342	3	2	
XXXIX	PtNAC026	gw1.XIII.1914.1	367	3	2	
XXXIX	PtNAC027	eugene3.00190359	368	3	2	
XL	PtNAC051	fgenesh4_pg.C_LG_XVII000368	401	3	2	
XL	PtNAC059	eugene3.01670013	401	3	2	
XL	PtNAC072	fgenesh4_pg.C_LG_IX000258	284	3	2	
XL	PtNAC073	eugene3.00051123	307	3	2	
XL	PtNAC074	estExt_Genewise1_v1.C_LG_IV1433	278	3	2	
XL	PtNAC075	fgenesh4_pg.C_LG_II000536	293	3	2	
XL	PtNAC076	fgenesh4_pg.C_LG_VII000785	303	3	2	
XL	PtNAC077	fgenesh4_pg.C_LG_V000378	308	3	2	
XL	PtNAC079	fgenesh4_pg.C_LG_I000708	248	4	3	
XL	PtNAC080	fgenesh4_pg.C_LG_III001261	281	3	2	
XL	PtNAC093	fgenesh4_pg.C_LG_III000226	313	4	3	
XL	PtNAC100	fgenesh4_pg.C_scaffold_86000102	383	3	2	
XL	PtNAC116	eugene3.00050415	254	3	2	
XLI	PtNAC099	fgenesh4_pg.C_LG_VI001372	269	3	2	
XLI	PtNAC104	eugene3.00160698	256	3	2	
XLI	PtNAC106	fgenesh4_pg.C_LG_I001572	246	3	2	

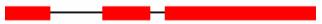
Clade	PtNAC Name	JGI Gene Model Name	Predicted Length (Amino Acids)	No. Exons	No. Introns	Predicted Gene Structure[1]
XLI	PtNAC108	fgenesh4_pg.C_scaffold_5651000001	246	3	2	
XLI	PtNAC110	eugene3.00160699	255	3	2	
XLI	PtNAC113	estExt_fgenesh4_pg.C_LG_IX1366	227	4	3	
LI	PtNAC139	eugene3.00400178	386	5	4	
	PtNAC028	estExt_fgenesh4_pm.C_LG_XVII0129	309	4	3	
	PtNAC119	eugene3.00012003	375	5	4	
	PtNAC140	fgenesh4_pg.C_LG_VI001273	281	3	2	
	PtNAC142	eugene3.00011461	322	1	0	
	PtNAC145	fgenesh4_pg.C_LG_XVI000191	340	1	0	
	PtNAC146	eugene3.00400326	288	3	2	
	PtNAC147	eugene3.00150032	418	1	0	

1 Predicted gene structures came from JGI Populus trichocarpa version 1.1 (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html) and are not to scale. The red boxes indicate exons, the grey lines indicate introns, and the blue boxes indicate UTR sequences.

Table S2.3. 119 rice NAC domain gene.

Clade	TIGR Locus Name[1]	RAP-DB Locus Name[2]	ONAC Name[3]	Predicted Length (amino acids)	No. Exons	No. Introns	Predicted Gene Structure[4]
I	Os06g15690	Os06g0267500	ONAC083	419	6	5	
II	Os01g48130	Os01g0672100	ONAC073	334	3	2	
XI	Os01g59640	Os01g0811500	ONAC092	402	3	2	
XVI	Os02g34970	Os02g0555300	ONAC028	205	3	2	
XVI	Os04G35660	Os04g0437000	ONAC105	201	3	2	
XXI	Os06g01230	Os06g0101800	ONAC019	393	4	3	
				360	3	2	
XXI	Os08g06140	Os08g0157900	ONAC037	731	7	2	
				730	7	2	
				726	7	2	
XXII	Os01g15640	Os01g0261200	ONAC074	490	4	3	
XXIII	Os03g02800	Not Found	ONAC096	651	6	5	
XXIII	Os10g42130	Os10g0571600	ONAC135	330	3	2	
XXIV	Os02g42970	Os02g0643600	ONAC094	371	3	2	
XXIV	Os03g03540	Not Found	ONAC097	367	3	2	
XXIV	Os04g45340	Os04g0536500	ONAC107	N/A	N/A	N/A	N/A
XXIV	Os04g59470	Os04g0691300	ONAC109	286	1	0	
XXIV	Os06g01480	Os06g0104200	ONAC087	365	3	2	
XXIV	Os08g01330	Os08g0103900	ONAC031	325	3	2	
XXV	Os06g04090	Os06g0131700	ONAC007, ONAC079	396	3	2	
XXV	Os08g02300	Os08g0115800	ONAC029	401	3	2	
XXVI	Os02g15340	Os02g0252200	ONAC013, ONAC064	360	3	2	
XXVI	Os06g33940	Os06g0530400	ONAC114	N/A	N/A	N/A	N/A
XXVII	Os02g06950	Os02g0165400	ONAC093	504	6	5	
XXVII	Os02g56600	Os02g0810900	ONAC032, ONAC056	332	3	2	
XXVII	Os03g01870	Os03g0109000	ONAC055	298	3	2	
XXVII	Os04g43560	Os04g0515900	ONAC008	279	3	2	
XXVII	Os04g52810	Os04g0619000	ONAC108	327	3	2	
XXVII	Os06g23650	Os06g0344900	ONAC113	374	2	1	

Clade	TIGR Locus Name[1]	RAP-DB Locus Name[2]	ONAC Name[3]	Predicted Length (amino acids)	No. Exons	No. Introns	Predicted Gene Structure[4]
XXVII	Os06g46270	Os06g0675600	ONAC011	305	3	2	
XXVII	Os08g10080	Os08g0200600	ONAC080	294	3	2	
XXVII	Os08g40030	Os08g0511200	ONAC127	341	2	1	
XXVII	Os10g33760	Os10g0477600	ONAC077	325	3	2	
XXVII	Os12g41680	Os12g0610600	ONAC060	334	3	2	
XXVIII	Os01g01470	Os01g0104500	ONAC090	321	2	1	
XXVIII	Os01g29840	Os01g0393100	ONAC026	334	2	1	
XXVIII	Os02g36880	Os02g0579000	ONAC027	375	3	2	
XXVIII	Os03g21030	Os03g0327100	ONAC039	359	3	2	
XXVIII	Os04g38720	Os04g0460600	ONAC004, ONAC034, ONAC058	344	3	2	
XXVIII	Os07g48550	Os07g0684800	ONAC015, ONAC069	302	3	2	
XXVIII	Os11g03310	Os11g0127000	ONAC137	352	2	1	
XXVIII	Os11g03370	Os11g0127600	ONAC045	360	3	2	
XXVIII	Os12g03050	Os12g0123800	ONAC144	397	2	1	
XXXIII	Os09g32260	Os09g0497900	ONAC072	353 353	3 3	2 2	
XXXIV	Os02g57650	Os02g0822400	ONAC070	633	5	4	
XXXIV	Os08g44820	Os08g0562200	ONAC040, ONAC054, ONAC081	657 619 492 434	5 4 4 4	4 3 3 3	
XXXV	Os08g33670	Os08g0433500	ONAC126	335	3	2	
XXXV	Os09g24560	Os09g0411900	ONAC084	301	3	2	
XXXVI	Os01g60020	Os01g0816100	ONAC068	319	3	2	
XXXVI	Os01g66120	Os01g0884300	ONAC048	304	3	2	
XXXVI	Os03g60080	Os03g0815100	ONAC002, ONAC033, ONAC043, ONAC044	317	2	1	
XXXVI	Os05g34830	Os05g0421600	ONAC088	323	3	2	
XXXVI	Os07g12340	Os07g0225300	ONAC067	277	1	0	

Clade	TIGR Locus Name[1]	RAP-DB Locus Name[2]	ONAC Name[3]	Predicted Length (amino acids)	No. Exons	No. Introns	Predicted Gene Structure[4]
XXXVI	Os11g08210	Os11g0184900	ONAC009, ONAC020, ONAC071	330	3	2	
XXXVII	Os07g37920	Os07g0566500	ONAC010	426	3	2	
XXXVIII	Os01g01430	Os01g0104200	ONAC089	340	2	1	
XXXVIII	Os03g21060	Os03g0327800	ONAC098	393	3	2	
XXXVIII	Os07g48450	Os07g0683200		347	3	2	
			ONAC086	642	2	1	
				200	1	0	
XXXVIII	Os11g03300	Os11g0126900	ONAC136	392	3	2	
				181	1	0	
XXXVIII	Os12g03040	Os12g0123700	ONAC143	394	3	2	
XL	Os01g66490	Os01g0888300	ONAC075	453	3	2	
XL	Os02g51120	Not Found	ONAC095	387	3	2	
XL	Os03g04070	Os03g0133000	ONAC022	317	3	2	
XL	Os03g56580	Os03g0777000	ONAC066	363	3	2	
XL	Os06g51070	Os06g0726300	ONAC116	293	1	0	
XL	Os07g04560	Os07g0138200	ONAC117	344	3	2	
XL	Os08g02160	Os08g0113500	ONAC124	335	3	2	
XL	Os08g33910	Os08g0436700	ONAC063	386	3	2	
XL	Os12g43530	Os12g0630800	ONAC152	376	3	2	
XLI	Os01g64310	Os01g0862800	ONAC059	257	3	2	
XLI	Os05g37080	Os05g0442700	ONAC012	275	3	2	
XLI	Os11g05610	Not Found	ONAC139	N/A	N/A	N/A	N/A
XLI	Os12g05990	Os12g0156100	ONAC146	308	1	0	
XLII	Os02g12310	Os02g0214500	ONAC023	253	2	1	
XLII	Os05G34310	Os05g0415400	ONAC024, ONAC078	311	3	2	
XLIII	Os01g70110	Os01g0925400	ONAC041, ONAC050	229	2	1	
XLIII	Os08g42400	Os08g0535800	ONAC005, ONAC035, ONAC046, ONAC052	233	3	2	
				168	2	1	
				165	3	2	

Clade	TIGR Locus Name[1]	RAP-DB Locus Name[2]	ONAC Name[3]	Predicted Length (amino acids)	No. Exons	No. Introns	Predicted Gene Structure[4]
XLIII	Os09g33490	Os09g0509100	ONAC001	248	3	2	
XLIV	Os11g31330	Os11g0512000	ONAC025	301	3	2	
XLIV	Os11g31360	Os11g0512200	ONAC141	291	3	2	
XLIV	Os11g31380	Os11g0512600	ONAC142	241	3	2	
XLV	Os10g09820	Os10g0177000	ONAC128	497	4	3	
XLV	Os11g04960	Os11g0146900	ONAC138	452	3	2	
XLVI	Os10g25620	Os10g0395700	ONAC129	533	3	2	
XLVI	Os10g25640	Not Found	ONAC130	196	3	2	
XLVI	Os10g26240	Os10g0401800	ONAC131	173	2	1	
XLVI	Os10g27360	Os10g0413700	ONAC133	422	3	2	
XLVI	Os10g27390	Os10g0414000	ONAC134	544	3	2	
XLVII	Os03g39050	Os03g0587700	ONAC099	443	3	2	
XLVII	Os03g39100	Os03g0588000	ONAC100	501	1	0	
XLVII	Os05g25960	Os05g0325300	ONAC111	568	1	0	
XLVII	Os07g09740	Os07g0195600	ONAC118	665	1	0	
XLVII	Os07g09830	Os07g0196500	ONAC119	707	1	0	
XLVII	Os07g09860	Os07g0196800	ONAC120	679	1	0	
XLVII	Os07g13920	Os07g0242800	ONAC121	375	1	0	
XLVII	Os07g17180	Os07g0272700	ONAC122	606	1	0	
XLVII	Os07g27330	Os07g0456900	ONAC065	777	1	0	
XLVII	Os07g27340	Os07g0457100	ONAC123	638	1	0	
XLVII	Os08g23880	Os08g0327800	ONAC125	448	1	0	
XLVIII	Os01g48460	Not Found	ONAC091	291	1	0	
XLVIII	Os03g61250	Not Found	ONAC102	317	1	0	
XLVIII	Os03g61320	Not Found	ONAC103	N/A	N/A	N/A	N/A
XLVIII	Os03g62470	Os03g0841500	ONAC104	132	1	0	
XLVIII	Os12g22630	Os12g0415000	ONAC148	471	1	0	
XLVIII	Os12g22940	Os12g0418300	ONAC149	418	1	0	
XLVIII	Os12g23090	Os12g0419600	ONAC150	405	2	1	
LI	Os10g21560	Os10g0359500	ONAC061	394	5	4	
				334	4	3	
LII	Os02g38130	Os02g0594800	ONAC082	394	6	5	
LII	Os04g40130	Os04g0477200	ONAC106	N/A	N/A	N/A	N/A

Clade	TIGR Locus Name[1]	RAP-DB Locus Name[2]	ONAC Name[3]	Predicted Length (amino acids)	No. Exons	No. Introns	Predicted Gene Structure[4]
No Clade	Os01g09550	Os01g0191300	ONAC003, ONAC016, ONAC085, ONAC090	434	6	5	
No Clade	Os03g42630	Os03g0624600	ONAC006	324	2	1	
No Clade	Os03g59730	Not Found	ONAC101	497	3	2	
No Clade	Os05g10620	Os05g0194500	ONAC110	358	5	4	
No Clade	Os05g43960	Os05g0515800	ONAC112	495	3	2	
No Clade	Os06g36480	Os06g0560300	ONAC115	568	9	8	
No Clade	Os09g32040	Os09g0493700	ONAC042, ONAC051	703	4	3	
No Clade	Os10g26270	Os10g0402100	ONAC132	415	3	2	
No Clade	Os11g31340	Os11g0512100	ONAC140	278	3	2	
No Clade	Os12g04150	Os12g0135800	ONAC145	N/A	N/A	N/A	N/A
No Clade	Os12g07790	Os12g0177600	ONAC147	572	1	0	
No Clade	Os12g29330	Os12g0477400	ONAC151	261	2	1	

1 TIGR uses a similar convention for loci names as the one used for the *Arabidopsis* genome, with minor modification for the larger size of the rice genome. Each nuclear gene is labeled LOC_OsXXg##### with LOC_Os referring to Oryza sativa locus, XX referring to chromosome 01-12, g referring to gene, and a 5-digit number referring to the gene order on the chromosome (http://www.tigr.org/tdb/e2k1/osa1/tigr_gene_nomenclature.shtml).

2 Rice Annotation Project Database (RAP-DB) assigned systematic locus identifiers to the RAP loci on the International Rice Genome Sequencing Project genome assembly using an ID (OsXXg#######) consists of the species name (Os for Oryza sativa), a two-digit number for chromosomes, and a seven-digit number that indicates a sequential order of loci in a chromosome. These Os identifier are different from the ones used by TIGR see the RAP-DB website for more information (http://rapdb.dna.affrc.go.jp/note.html).

3 ONAC naming scheme coined from Ooka et al 2003 paper. Red ONAC names represent NACs who were assigned ONAC names in this study.

4 Predicted gene structures came from (insert website) and are not to scale. The red boxes indicate exons, the grey lines indicate introns, and the black boxes indicate UTR sequences.

Table S2.4. *Arabidopsis* and poplar NAC domain orthologs.

Class	*Arabidopsis*			Poplar			
1:1	At2G27300			PtNAC084			
	At1G69490			PtNAC048			
1:n	At5G50820			PtNAC062	PtNAC071		
	At5G13180			PtNAC052	PtNAC053		
	At5G14000			PtNAC114	PtNAC115		
	At3G17730			PtNAC017	PtNAC020		
	At1G54330			PtNAC001	PtNAC007		
	At5G62380			PtNAC037	PtNAC050		
	At1G71930			PtNAC055	PtNAC060		
	At1G32770			PtNAC063	PtNAC068		
	At1G56010			PtNAC070	PtNAC081		
	At4G28530			PtNAC109	PtNAC111		
	At1G76420			PtNAC024	PtNAC031		
	At5G53950			PtNAC012	PtNAC016		
	At3G18400			PtNAC032	PtNAC042		
	At2G24430			PtNAC011	PtNAC013		
	At1G77450			PtNAC030	PtNAC034		
	At1G61110			PtNAC002	PtNAC005		
	At2G43000			PtNAC073	PtNAC075	PtNAC076	PtNAC077
				PtNAC079	PtNAC078		
	At2G17040			PtNAC072	PtNAC074		
	At2G02450			PtNAC093	PtNAC100		
n:1	none			none			
n:n	At5G09330	At5G64060		PtNAC067	PtNAC069		
	At1G33060	At4G35580		PtNAC018	PtNAC019		
	At3G49530	At5G24590		PtNAC041	PtNAC057		
	At1G12260	At1G62700		PtNAC038	PtNAC046		
	At1G33280	At4G10350		PtNAC064	PtNAC066		
	At3G04060	At5G18270		PtNAC014	PtNAC086		
	At5G07680	At5G61430		PtNAC015	PtNAC021		
	At5G08790	At5G63790		PtNAC043	PtNAC045		
	At1G52890	At3G15500		PtNAC036	PtNAC044		
	At1G52880	At3G15510		PtNAC003	PtNAC006		
	At1G26870	At5G39820		PtNAC051	PtNAC059		
	At1G32510	At1G64108	At4G17980	PtNAC008	PtNAC010		
	At5G46590						
	At3G03200	At5G17260	At1G65910	PtNAC004	PtNAC009		
	At2G18060	At4G36160	At5G66300	PtNAC025	PtNAC039		
	At3G44350	At5G22380		PtNAC106	PtNAC108	PtNAC113	

Table S2.5. *Arabidopsis* and rice NAC domain orthologs

Class	*Arabidopsis*			Rice			
1:1	At1g65910			Os03g02800			
	At1g76420			Os08g40030			
	At2g24430			Os09g32260			
1:n	At5g64530			Os02g34970	Os04g35660		
	At1g71930			Os08g01330	Os04g59470		
	At1g56010			Os06g46270	Os02g06750	Os12g41680	
	At4g28530			Os04g43560	Os10g33760	Os03g01870	Os02g56600
				Os03g21060	Os07g48450	Os01g01430	
	At1g69490			Os02G51120	Os08G33910		
	At1G26870			Os03G56580	Os07G04560	Os12G43530	
	At2G43000			Os03G04070	Os06G51070		
	At2G17040			Os01G66490	Os08G02160		
	At2G02450			Os03G39050	Os03G39100	Os05G28960	Os07G13920
				Os07G09830	Os07G09860	Os07G09740	Os07G17180
	At5G18300			Os07G27330	Os07G27340	Os08G23880	
n:1	At3G44290	At5G22290	At2G27300	Os01G15640			
	At1G32510	At1G64105	At4G17980	Os10G42130			
	At5G46590						
	At1G52880	At3G15510	At1G61110	Os07G37920			
n:n	At1G33060	At4G35580	At3G49530	Os08G06140	Os06G01230		
	At5G24590						
	At1G12260	At1G62700	At5G62380	Os02G42970	Os04G45340	Os06G01480	
	At2G18060	At4G36160	At5G66300	Os03G03540	Os10G38830		
	At2G46770	At3G61910	At1G32770	Os06G04090	Os08G02300		
	At1G79580	At1G33280	At4G10350	Os02G15340	Os06G33940		
	At3G04060	At5G18270		Os01G01470	Os01G29840	Os03G21030	Os07G48550
				Os02G36880	Os04G38720	Os12G03050	Os11G03370
				Os11G03310			
	At1G01720	At1G77450	At5G63790	Os01G66120	Os05G34830		
	At3G44350	At5G22380		Os01G64310	Os05G37080	Os11G05610	Os12G05990

Table S2.6. Poplar and rice NAC domain orthologs.

Class	Poplar			Rice			
1:1	PtNAC084			Os01G15640			
	PtNAC139			Os10G21560			
1:n	PtNAC048			Os03G21060	Os07G48450	Os01G01430	
n:1	PtNAC130	PtNAC131	PtNAC149	Os01G59640			
	PtNAC125						
	PtNAC008	PtNAC010		Os10G42130			
	PtNAC004	PtNAC009		Os03G02800			
	PtNAC024	PtNAC031		Os08G40030			
	PtNAC011	PtNAC013		Os09G32260			
	PtNAC003	PtNAC006	PtNAC002	Os07G37920			
	PtNAC005						
n:n	PtNAC025	PtNAC039		Os03G03540	Os10G38830		
	PtNAC055	PtNAC060		Os08G01330	Os04G59470		
	PtNAC051	PtNAC059		Os02G51120	Os08G33910		
	PtNAC093	PtNAC100		Os01G66490	Os08G02160		
	PtNAC118	PtNAC128	PtNAC122	Os02G34970	OS04G35660		
	PtNAC129						
	PtNAC041	PtNAC057	PtNAC018	Os08G06140	Os06G01230		
	PtNAC019						
	PtNAC037	PtNAC037	PtNAC038	Os02G42970	Os04G45340	Os06G01480	
	PtNAC046	PtNAC050					
	PtNAC063	PtNAC063	PtNAC068	Os06G04090	Os08G02300		
	PtNAC061	PtNAC065					
	PtNAC054	PtNAC058	PtNAC064	Os02G15340	Os06G33940		
	PtNAC066						
	PtNAC070	PtNAC081		Os06G46270	Os02G06950	Os08G10080	Os12G41680
	PtNAC109	PtNAC111		Os04G43560	Os10G33760	Os03G01870	Os02G56600
	PtNAC014	PtNAC086		Os01G01470	Os01G29840	Os03G21030	Os07G48550
				Os02G36880	Os04G38720	Os12G03050	Os11G03370
				Os11G03310			
	PtNAC043	PtNAC030	PtNAC045	Os01G66120	Os05G34830		
	PtNAC034						
	PtNAC073	PtNAC075	PtNAC076	Os03G56580	Os07G04560	Os12G43530	
	PtNAC077	PtNAC079	PtNAC078				
	PtNAC072	PtNAC074	PtNAC116	Os03G04070	Os06G51070		
	PtNAC106	PtNAC108	PtNAC113	Os01G64310	Os05G37080	Os11G05610	Os12G05990
	PtNAC099	PtNAC104	PtNAC110				

Chapter 3

Functional Analysis of PtNAC068[1]

[1]Kimberly Diane Hunt, Joe Nairn, Steve Strauss, and Sarah Covert, to be submitted to Plant Physiology.

Abstract

Secondary wall thickening and lignification are hallmarks of plant vascular cells involved in structural support or water transport. Identifying the molecular switches that control secondary wall deposition, therefore, is important for understanding the development of vascular tissues in plants. In this study, we elucidated the role of a *Populus* NAC domain transcription factor, PtNAC068, in vascular development by analyzing PtNAC068 over-expression, knockdown, and dominant repression lines. PtNAC068 over-expression mutants displayed ectopic lignin deposition in leaf epidermal cells and primary vein pith cells, as well as an increase in lignin deposition in phloem fiber bundles and associated cells in leaves and stems. PtNAC068 dominant repressor mutants had a reduction in the size and number of phloem fiber bundles in primary leaf veins and stems, as well as delayed development of the sclereids associated with phloem fibers in stems. These findings indicate that PtNAC068 is a positive regulator of secondary wall development in phloem fibers and their associated sclereid cells in poplar.

Introduction

The secondary, or lateral, growth of woody plant stems is produced by cellular division and differentiation at the vascular cambium. This process creates the secondary xylem and phloem tissues, which function in support, as well as water and sugar transport [1]. How secondary vascular tissues develop in woody plants is not very well understood, but it is an important topic not only because of its relevance to basic plant biology, but also because of its influence on industrial fiber production, solid wood quality and bioenergy feedstocks.

Recently, the NAC domain transcription factor family has been implicated in vascular development in *Arabidopsis*. VASCULAR-RELATED NAC DOMAIN PROTEIN 6 (VND6) and 7 (VND7) were positive switches for metaxylem and protoxylem vessel formation [2]. NAC SECONDARY WALL THICKENING PROMOTING FACTOR 1 (NST1), 2 (NST2), and 3 (NST3/SND1/ANAC012) function as regulators of secondary wall thickening [3-6]. NST1 and NST2 influence secondary wall thickening during anther dehiscence and tracheary element formation [4], while the secondary wall thickening of interfascicular and xylary fibers is influenced by NST3/SND1/ANAC012[3, 5, 6].

Transcription factors in the NAC domain family are known for overlapping or redundant functions that mask mutant phenotypes when single gene loss-of-function mutants are used. This was the case for the VND and NST genes because loss-of-function mutations in individual genes did not produce any detectable phenotypes [2, 4, 6]. To get around the potential functional redundancy of the NAC domain transcription factor family, dominant repressor technology was employed. In this approach, transcription factors fused to the EAR repression domain act as repressors of their target genes and this activity is dominant over the function of any redundant, unmodified NAC domain proteins [2, 4-7].

In a previous phylogenetic analysis (see Chapter 2), the NAC domains from *Arabidopsis* NST1, NST2, and NST3 cluster with four NAC domains from *Populus trichocarpa* (poplar) and two NAC domains from *Oryza sativa* (rice) (Figure 3.1). This subgroup of the NAC family will be referred to herein as the NST subfamily. Based on this analysis, two poplar proteins, PtNAC063 and PtNAC068, were potential orthologs of NST3. Therefore, PtNAC068 was predicted to regulate secondary wall thickening in vascular cells. In this study PtNAC068 over-

expression, *knockdown*, and dominant repression lines indicate that PtNAC068 is a positive regulator of secondary wall thickening in phloem fiber cells and their associated sclereids.

Methods and Materials

Plant Transformation and Growth Conditions

Appropriate plasmids were transformed into hybrid aspen clone INRA 717-IB4 (*Populus tremula* × *Populus alba*) via an *Agrobacterium tumefaciens*-mediated transformation procedure [8] at the Strauss lab at Oregon State University. Transformed plantlets were transferred to greenhouse and grown with 16 hours of light and 8 hours of darkness every 24 hours. Temperatures in the greenhouse were maintained no lower than 60^{o}. Poplar plants were fertilized once a week with 25% solution of 16-17-18 Peatlite fertilizer and watered as necessary to maintain plant vigor.

Plasmids

The protein coding region of PtNAC068 and 3'UTR of PtNAC063 were amplified from cDNAs using appropriate primers. The PtNAC068 coding region was directionally cloned into the AvrII and SacII sites of p35N23 to produce p35S:PtNAC068:NOST. For the dominant repression of PtNAC068 the full-length coding region of PtNAC068 was fused in-frame with the dominant EAR repression sequence [5], and cloned into the AvrII and SacII of the p35S23N plasmid to produce p35S:PtNAC068:EAR:NOST. For the RNAi construct, the 3'UTR of PtNAC063 in sense and antisense orientation (separated by a piece of GUS) was cloned into the AvrII and SacII sites of p35N23 to produce 35S:PtNAC063 3'UTR Sense:GUS Linker:PtNAC063 3'UTR Antisense: NOST. The corresponding 35S to NOST regions of the

over-expression, RNAi, and dominant repression p35N23 plasmids were cloned into the PZPNPT plant transformation vector and selected for using kanamycin.

Microscopy

Fixed Tissues

Leaf primary veins and stems section from transgenic and wild-type poplar plants were fixed in a solution containing 2% glutaraldehyde and 3% paraformaldehyde in 0.05M sodium phosphate buffer and dehydrated in an ethanol series (2 hours per step at 4C). A vacuum was pulled at each step to insure optimal infiltration of the fixatives and ethanol series. After dehydration tissues were embedded in JB-4 medium (JB-4 Plus Embedding Kit, Polysciences). Four to five micrometer-thick sections from embedded tissues were cut using a microtome, covered with mounting media, and observed under UV.

Fresh Mounts

Whole disc-sized sections of fresh transgenic and wild-type poplar primary and secondary leaf veins were cut and cleared in 70% lactic acid. These sections were then observed under bright-field and UV light. Fresh epidermal peals of over-expression and wild-type poplar plants were stained with 1% evans blue or 1% sudan black. The stained epidermal peals were washed in water for a few seconds and viewed under bright-field microscopy.

Results

Whole Plant Phenotypes of PtNAC068 Mutants

To elucidate the role of PtNAC068 in poplar, we produced three different types of transgenic plants. One type expressed PtNAC068 ectopically under the control of the CaMV

35S promoter. Of thirteen 35S:PtNAC068 lines, three showed consistent phenotypes. The second type of transgenic plant was transformed with an RNAi construct made from the 3'UTR of PtNAC063. This construct was expected to target both PtNAC068 and PtNAC063 because their 3'UTR sequences were similar. Of 17 RNAi transformed lines, one had a detectable phenotype. The third type of transgenic plant made use of dominant repression technology by fusing the PtNAC068 coding region to the EAR repression domain (a.k.a. the strong repression domain of SRDX) [4, 7] under the control of the CaMV 35S promoter. Of seventeen 35S:PtNAC068:SRDX lines, seven had similar, novel phenotypes.

All three types of transgenic poplar lines had abnormal leaves. Wild-type leaves were all held in the same plane (Figure 3.2A). The PtNAC068 over-expression leaves were more rigid and cup-shaped (Figure 3.2B) than wild-type. These leaves were cupped through all stages of development and some folded over on themselves. In contrast, the RNAi and PtNAC068 SRDX mutant leaves drooped downwards (Figure 3.2C and 3.2D). The leaves in the RNAi and SRDX lines displayed slightly different leaf phenotypes. The blades of the RNAi leaves folded downwards along a nearly erect mid-vein (Figure 3.2C), while the SRDX leaves and mid-veins drooped downwards from the petiole-leaf junction (Figure 3.2D). Unlike the over-expression lines whose leaves were cup-shaped from the time they were formed, the drooping associated with the loss-of-function mutants was not apparent until after the leaves had matured and reached their maximum size (data not shown).

PtNAC068 Affects Lignin Deposition in Leaf Cells.

In order to determine if changes in leaf cellular structure were contributing to the whole-leaf phenotypes in the PtNAC068 over-expression and loss-of-function lines, we examined the

epidermis and veins of the leaves under UV light. UV light makes lignin (a marker of secondary wall development) autofluoresce. Mutations in *Arabidopsis* NST1, NST2, and NST3 affect lignin deposition [4, 5, 9], therefore, mutations in PtNAC068 were predicted to also affect the deposition of lignin.

PtNAC068 over-expression plants deposited lignin in epidermal cell walls, whereas wild-type plants did not (Figure 3.3A and 3.3B). In addition, the epidermal cells of the over-expression plants were more rectangular and elongated than wild-type epidermal cells (Figure 3.3). The epidermal cells in plants over-expressing PtNAC068 lacked a spherical, sub-cellular structure seen in wild-type epidermal cells (Figure 3.3A-B). Vascular cells that undergo secondary wall development usually undergo programmed cell death and are dead at maturity. In order to determine if the epidermal cells in the PtNAC068 over-expression lines were alive or dead, they were stained with Evan's Blue, which only stains dead cells. No blue staining was seen in the PtNAC068 over-expression epidermal cells (Figure 3.3C and 3.3D). Therefore, these cells were still alive and had not undergone programmed cell death. Sudan black stained the structures in the wild-type epidermal cells black (Figure 3.3E and 3.3F), suggesting that the structures missing from the over-expression cells were lipid bodies.

To determine the affect of PtNAC068 over-expression on wall thickening in vascular tissue, cross-sections of primary leaf veins were examined. One of the most striking features of the primary veins in these leaves was the increase in lignin deposition in and around the phloem fibers (Figure 3.3L). In wild-type, the phloem cells are not lignified (Figure 3.3K, 3.3O). However, in the over-expression plants ectopic lignin deposition occurred in cells associated with the phloem layer; in some cases, cells with ectopic lignin deposition completely span the phloem layer from the outer fibers to the xylem (Figure 3.3L, 3.3P). Wild-type pith cells in the

veins (Figure 3.3K) are non-lignified, parenchyma cells [10]. In contrast, the pith of the PtNAC068 over-expression plants was lignified, indicative of secondary wall development in this tissue (Figure 3.3L, 3.3P). Cross-sections of PtNAC068 over-expression veins showed ectopic lignin deposition in the first 1-2 cell layers under the abaxial and adaxial epidermis (data not shown).

Under UV light, intact wild-type leaf veins autofluoresced brightly (Figure 3.3G and 3.3H). In contrast, the leaf veins of the PtNAC068:SRDX plants were not visible under UV light (Figure 3I and 3J), and lignin deposition was noticeably reduced in the phloem fibers of these veins (Figure 3.3N). In the PtNAC063:RNAi plants, the leaf veins could be seen under UV light, but their autofluorescence was not as bright as that of wild-type (data not shown). In cross-section, a reduction in phloem fiber lignification in the RNAi veins was difficult to detect by eye (Figure 3.3M), but a preliminary analysis with ImageJ software confirmed that the area of lignified tissue in the RNAi veins was slightly less than that of wild-type (data not shown).

PtNAC068 Affects Lignin Deposition in Stem Cells

To determine if PtNAC068 played a role in stem development, we examined cross-sections from three sampling points on wild-type, PtNAC068 over-expression, and PtNAC068:SRDX stems. The first position, referred to as zone 1, was approximately 10 centimeters from the shoot apical meristem while the second position, zone 2, was harvested from the middle of each tree trunk (Figure 3.4A). Samples from the third position, zone 3, were harvested 10 centimeters from the base of each tree trunk (Figure 3.4A). In wild-type, zone 1 contained a single row of phloem fiber bundles and a large area of non-lignified pith inside the xylem layer (data not shown). Wild-type zone 2 contained a row of mature phloem fiber bundles

and a row of developing phloem fiber bundles (Figure 3.4B), while wild-type zone 3 contained multiple rows of phloem fibers. In zone 3, the oldest row of phloem fiber bundles (those laid down originally in zone 1) were interleaved with lignified parenchymatous elements known as sclereids [11], thus a continuous band of lignified tissue encircled the outer stem in zone 3 of wild-type plants (Figure 3.4E). Of all the mutant lines, only PtNAC068 over-expression plants had a novel phenotype in zone 1. They exhibited ectopic lignin deposition in the parenchyma cells between the phloem fiber bundles and the epidermis and in normally non-lignified pith cells (data not shown). PtNAC068 over-expression stem sections from zone 2 deviated from wild-type zone 2 stem sections by having three disorganized rows of phloem fibers and an overall increase in the area of lignification (Figure 3.4B and 3.4C). They also differed from wild-type zone 2 because they contained sclereids in association with the phloem fiber bundles (Figure 3.4B and 3.4C). The expansion of lignin deposition associated with PtNAC068 over-expression was also present in zone 3 stem sections as there was an increase in the number and size of phloem fiber bundles (Figure 3.4F). However, the continuous band of lignified tissue in wild-type zone 3 was missing in the PtNAC068 over-expression plants, even though sclereids were present in association with the fiber bundles in these samples (Figure 3.4F). In contrast, the size and the number of phloem fiber bundles in zone 2 and zone 3 of the PtNAC068:SRDX plants were reduced compared to wild-type (Figure 3.4D and 3.4G). In addition, the PtNAC068:SRDX zone 3 stem sections almost completely lacked phloem fiber-associated sclereids (Figure 3.4G) which normally function in protection and support of phloem cells.

Discussion

Secondary wall thickening is a hallmark of cells involved in support or water transfer and is usually seen in plant vascular cells. Understanding the molecular switches that control secondary wall deposition in plants is important for understanding the development of vascular tissues. Three lines of evidence support the conclusion that PtNAC068 is a positive regulator of secondary wall development, particularly in poplar phloem fibers and their associated sclereids. First, 35S:PtNAC068 mutants displayed ectopic lignin deposition in leaf epidermal cells and primary vein pith cells. Second, 35S:PtNAC068 mutants had increased lignin deposition in phloem fibers in primary leaf veins and stems, and accelerated development of sclereids in stem phloem. Third, PtNAC068:SRDX mutants had decreased lignin deposition in phloem fibers in primary leaf veins and stems, and delayed development of sclereids in stem phloem.

When compared phylogenetically, PtNAC068 clusters with three *Arabidopsis* NAC domain proteins (NST1, NST2, NST3) involved in secondary wall development, but is most closely related to NST3 (Figure 3.1). NST3 is involved in regulating secondary wall synthesis in *Arabidopsis* xylary and interfascicular fibers (Zhong, Demura et al. 2006; Ko, Yang et al. 2007; Mitsuda, Iwase et al. 2007; Zhong, Demura et al. 2007). NST3 over-expression mutants produce plants with severely curled rosette leaves [5, 6]. The severe curling of *Arabidopsis* leaves in NST3 over-expression plants is due to ectopic lignin deposition in epidermal and mesophyll cells, and increased secondary wall development in interfascicular and xylary fiber cells [5, 6]. This is similar to the patterns of lignin deposition seen in the PtNAC068 over-expression mutants (Figure 3.4), although PtNAC068 affected only phloem fibers and not xylary fibers or leaf mesophyll. Some functional redundancy is shared between NST1 and NST3 [5, 9], as well as

between NST1 and NST2 [4]. The over-expression of NST1 results in leaves that are upwardly curled with ectopic lignin expression in leaf epidermal cells and mesophyll cells [5].

NST1, NST2, and NST3 activate genes involved in secondary wall thickening [4-6, 9], a process that is usually followed by programmed cell death to produce mature vascular cells. However, NST1 and NST2 over-expression mutants were able to activate secondary cell wall associated genes (e.g., cellulose and lignin biosynthetic enzymes) without activating programmed cell death associated genes (e.g., xylanases, peptidases and nucleases), indicating that NST1 and NST2 do not have the ability to regulate the entire process of tracheary element differentiation [4]. The function of PtNAC068 may be similar because the epidermal cells in 35S:PtNAC068 plants displayed patterns of lignin deposition and cell shape consistent with secondary wall thickening, and they lacked lipid bodies normally seen in wild-type cells, but they appeared to still be alive because they did not stain with Evan's blue (Figure 3.3A-F). Therefore, it is possible that PtNAC068 can stimulate secondary wall development and thus initiate aspects of fiber differentiation in developmentally plastic cells, but it can not complete fiber maturation because programmed cell death is not initiated.

The drooping leaf phenotypes seen in the PtNAC063 RNAi and PtNAC068:SRDX mutants are consistent with a loss in phloem fibers in leaf veins. The more severe drooping in the SRDX lines could be due to a dominant negative effect on functionally redundant NAC protein(s) that are not affected in the RNAi lines. Alternatively, it could be that the RNAi construct, which was designed to target transcription of PtNAC068 and PtNAC063, was not highly effective at knocking down the expression of either gene in the single transformed line that was recovered after transformation with this construct. The whole plant phenotypes in the

dominant repression and RNAi mutants in *Arabidopsis* NST proteins [4-6, 9] are somewhat analogous to those seen in the PtNAC068 dominant repression mutants.

Dominant repression SND1 (NST3) plants had leaning inflorescence stems caused by a reduction in lignin deposition in xylary and interfascicular fibers [6]. This is similar to the drooping leaf phenotype resulting from a loss of phloem fiber cells in PtNAC068 SRDX mutants. PtNAC068:SRDX mutant leaves begin to droop near the petiole and lack the support needed to hold the leaf vein and blade erect (Figure 3.2D). The inflorescence stems of double RNAi NST1 and NST3 (SND1) mutants completely lack the support needed to hold the stem upright and the inflorescence stems lay on the ground [5, 9]. The PtNAC068 dominant repression mutants, however, do not have stems that lay on the ground, most likely due to the difference in the amount of xylem in *Arabidopsis* and poplar. Poplar has a lot more secondary xylem than *Arabidopsis,* and its stature was not noticeably affected by changes in PtNAC068 expression. This tissue was able to hold the poplar stems in an upright manner despite the loss of lignification seen in the phloem fiber bundles (Figure 3.4D and 3.4G).

Vascular cells in plants are produced from cambial cells that differentiate into fusiform initials [12]. *Arabidopsis* HOMEOBOX GENE-8 (ATHB8), a member of the homeodomain leucine zipper family, is a positive regulator of proliferation and differentiation of vascular cells from procambial and cambial cells [13]. WOODEN LEG (WOL) is another regulator of proliferation of procambial cells. Mutations in WOL result in loss of phloem cells and altered xylem organization in *Arabidopsis* root vascular tissue [14, 15]. Several genes are known to affect xylem tissues produced from fusiform inititals. VND6 and VND7 are transcription factors affecting the formation of protoxylem and metaxylem [2]. Xylary fibers in *Arabidopsis* are affected by the NST NAC domain transcription factor subfamily [5, 6, 9]. Phloem tissues

resulting from fusiform initials are not as well studied as xylem tissues. At least one transcription factor [13] and various phytahormones (auxin, cytokinin, and gibberellic acids) are suspected players in phloem tissue development (phytahormones reviewed in [16]). The ALTERED PHLOEM DEVELOPMENT (APL), MYB coiled-coil transcription factor, is required for phloem development because mutants in APL eliminate phloem development throughout *Arabidopsis* plants [17]. PtNAC068 is the first woody plant transcription factor shown to be involved in phloem fiber development.

References

1. Berleth T, Mattsson J, Hardtke CS: **Vascular continuity and auxin signals**. *Trends in Plant Sciences* 2000a, **5**:387-393.
2. Kubo M, Udagawa M, Nishikubo N, Horiguchi G, Yamaguchi M, Ito J, Mimura T, Fukuda H, Demura T: **Transcription switches for protoxylem and metaxylem vessel formation**. *Genes & Development* 2005, **19**(16):1855-1860.
3. Ko JH, Yang SH, Park AH, Lerouxel O, Han KH: **ANAC012, a member of the plant-specific NAC transcription factor family, negatively regulates xylary fiber development in Arabidopsis thaliana**. *Plant Journal* 2007, **50**(6):1035-1048.
4. Mitsuda N, Seki M, Shinozaki K, Ohme-Takagi M: **The NAC transcription factors NST1 and NST2 of Arabidopsis regulate secondary wall thickenings and are required for anther dehiscence**. *Plant Cell* 2005, **17**(11):2993-3006.
5. Mitsuda N, Iwase A, Yamamoto H, Yoshida M, Seki M, Shinozaki K, Ohme-Takagi M: **NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of Arabidopsis**. *Plant Cell* 2007, **19**(1):270-280.
6. Zhong RQ, Demura T, Ye ZH: **SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of Arabidopsis**. *Plant Cell* 2006, **18**(11):3158-3170.
7. Hiratsu K, Matsui K, Koyama T, Ohme-Takagi M: **Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in Arabidopsis**. *Plant Journal* 2003, **34**(5):733-739.

8. Han KH, Meilan R, Ma C, Strauss SH: **An Agrobacterium tumefaciens transformation protocol effective on a variety of cottonwood hybrids (genus Populus)**. *Plant Cell Reports* 2000, **19**(3):315-320.

9. Zhong RQ, Richardson EA, Ye ZH: **Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of Arabidopsis**. *Planta* 2007, **225**(6):1603-1611.

10. Pandey SN: **A textbook of Botany Angiosperms**, 4th edn. Ltd ,India: S. Chand & Co 2001.

11. Tucker CM, Evert RF: **SEASONAL DEVELOPMENT OF SECONDARY PHLOEM IN ACER NEGUNDO**. *American Journal of Botany* 1969, **56**(3):275-&.

12. Mauseth JD: **Botany an Introduction to Plant Biology**, 2nd edn. Philadelphia: Saunders College Publishing; 1995.

13. Baima S, Possenti M, Matteucci A, Wisman E, Altamura MM, Ruberti I, Morelli G: **The Arabidopsis ATHB-8 HD-zip protein acts as a differentiation-promoting transcription factor of the vascular meristems**. *Plant Physiology* 2001, **126**(2):643-655.

14. Mahonen AP, Bonke M, Kauppinen L, Riikonen M, Benfey PN, Helariutta Y: **A novel two-component hybrid molecule regulates vascular morphogenesis of the Arabidopsis root**. *Genes & Development* 2000, **14**(23):2938-2943.

15. Scheres B, Dilaurenzio L, Willemsen V, Hauser MT, Janmaat K, Weisbeek P, Benfey PN: **MUTATIONS AFFECTING THE RADIAL ORGANIZATION OF THE ARABIDOPSIS ROOT DISPLAY SPECIFIC DEFECTS THROUGHOUT THE EMBRYONIC AXIS**. *Development* 1995, **121**(1):53-62.

16. Aloni R: **DIFFERENTIATION OF VASCULAR TISSUES**. *Annual Review of Plant Physiology and Plant Molecular Biology* 1987, **38**:179-204.

17. Bonke M, Thitamadee S, Mahonen AP, Hauser MT, Helariutta Y: **APL regulates vascular tissue identity in Arabidopsis**. *Nature* 2003, **426**(6963):181-186.

Figure 3.1. Phylogenetic tree of the NST NAC domain subfamily in Arabidopsis, poplar and rice. This diagram is an excerpt from the MrBayes tree presented in Figure 2.2.

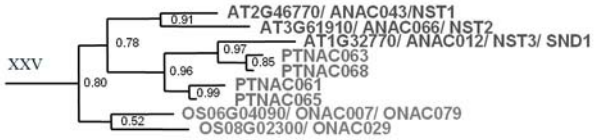


Figure 3.2. Whole leaf phenotypes of wild type and PtNAC068 mutants. (A) Wild type. (B) PtNAC068 over-expressor. (C) PtNAC063 RNAi plant (D) PtNAC068-SRDX plant.

Figure 3.3. Leaf cell phenotypes. (A) Wild type leaf epidermal cells under UV light. (B) PtNAC068 over-expression leaf epidermal cells under UV light. (C) Wild type leaf epidermal cells stained with Evan's blue. (D) PtNAC068 over-expression leaf epidermal cells stained with Evan's blue. (E) Wild type leaf epidermal cells stained with Sudan black. (F) PtNAC068 over-expression leaf epidermal cells stained with Sudan black. (G) Bright field image of wild type leaf vein. (H) Same image as in G under UV light. (I) Bright field image of PtNAC068-SRDX leaf vein. (J) Same image as in I under UV light. (K) Cross section of wild type primary leaf vein under UV light. (L) Cross section of PtNAC068 over-expressor primary leaf vein under UV light. (M) Cross section of PtNAC063 RNAi leaf vein under UV light. (N) Cross section of PtNAC068-SRDX leaf vein under UV light. (O) Longitudinal section of wild type leaf vein under UV light. P = Phloem. Pi = Pith. (P) Longitudinal section of PtNAC068 over-expression leaf vein under UV light. P = Phloem. Pi = Pith.

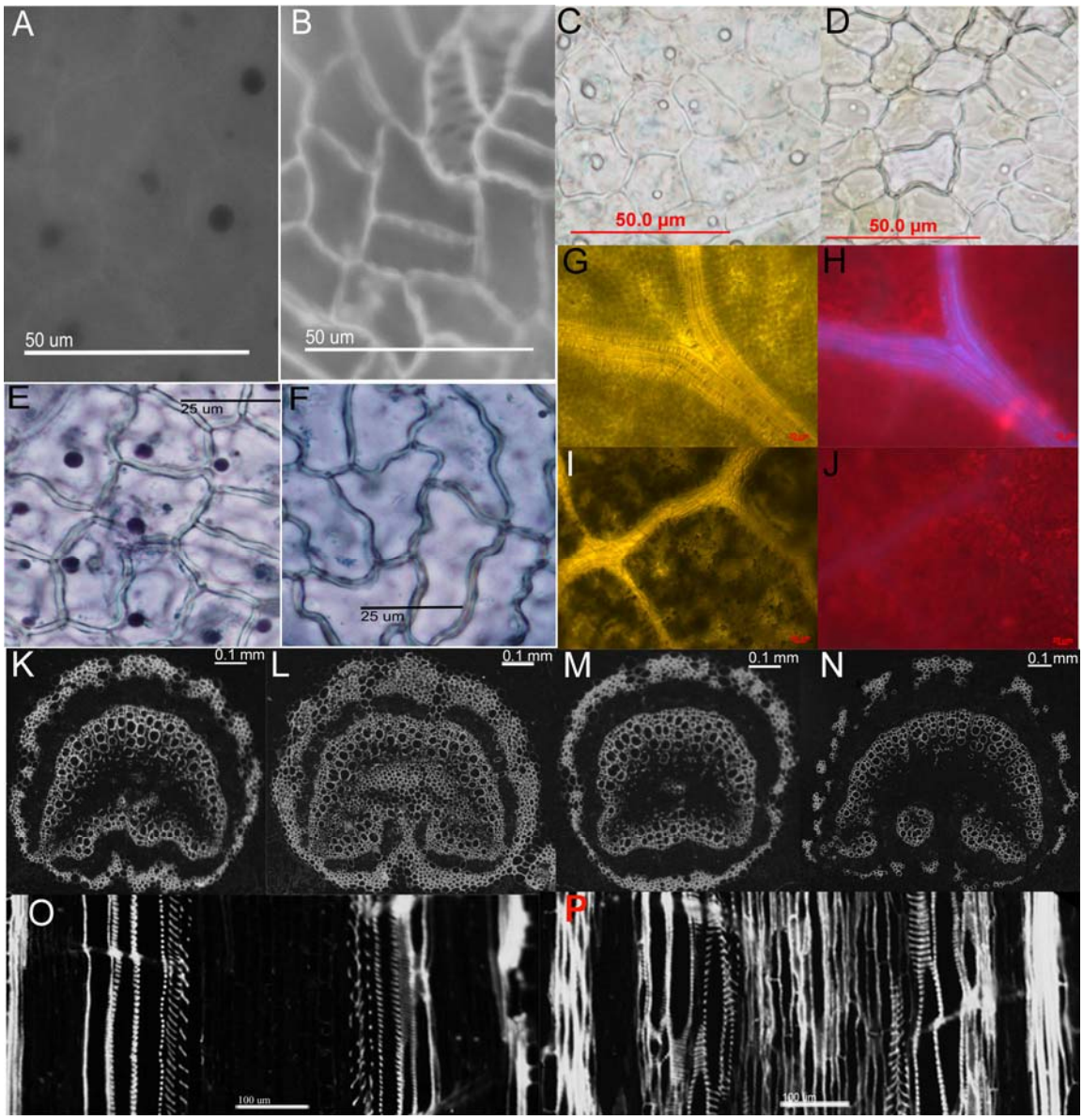
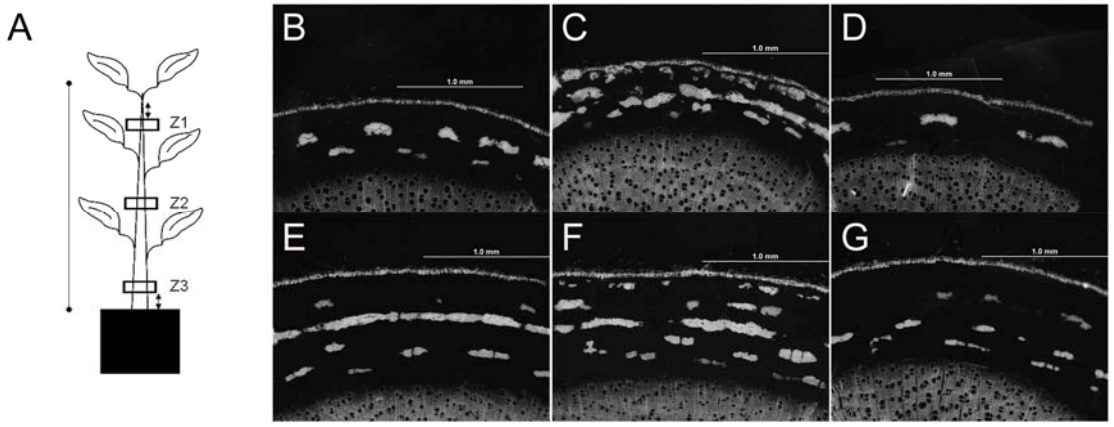

A
50 um
B
50 um
C
50.0 µm
D
50.0 µm
E
25 um
F
25 um
G
H
I
J
K
0.1 mm
L
0.1 mm
M
0.1 mm
N
0.1 mm
O
100 um
P
100 um

Figure 3.4. Stem Phenotypes. (A) Positions of sampling zones 1-3. Double-headed arrows represent 10 cm. The solid line on the left represents 1.6 meters. All pictures in B-G are of stem cross-sections under UV light. (B) Wild type zone 2. (C) PtNAC068 over-expressor zone 2. (D) PtNAC068-SRDX zone 2. (E) Wild type zone 3. (F) PtNAC068 over-expressor zone 3. (G) PtNAC068-SRDX zone 3.



Chapter 4

Discussion

In each higher plant genome, the NAC domain family encodes a hundred or more plant-specific transcription factors that are involved in many important developmental, defense, and stress responses. In order to expand our understanding of the NAC domain family and to elucidate the function of one of its poplar members, PtNAC068, phylogenetic, motif and functional analyses of the NAC domain family were done. The phylogenetic and motif analyses of the Arabidopsis, poplar, and rice NAC domain family provided a solid foundation for future functional and evolutionary studies, while the functional analysis of PtNAC068 provided a proof of concept for the phylogenetic analysis and determined that PtNAC068 regulates phloem fiber development.

The phylogenetic analysis divided the 387 member Arabidopsis, poplar and rice NAC domain family into 52 closely related clades ranging from 2 to 26 sequences per clade. Dividing large gene families into smaller, closely related groups allows researchers to more efficiently assess functional and evolutionary relationships among the genes of greatest interest to them. For example, some researchers might want to focus on potential dicot- or rice-specific genes. There are 33 potential dicot-specific genes (i.e. those in Arabidopsis and poplar, but not rice) and 45 potential rice-specific genes in the Arabidopsis, poplar and rice NAC domain family. Determining if the rice-specific genes encode functions that might be shared broadly among the monocots awaits the sequencing of additional monocot genomes. Other researchers might be more interested in studying tree-specific transcription factors. Our analysis has identified 50 potential tree-specific NAC domain proteins (i.e. those found only in poplar). It is possible, however, that some of these "tree-specific" NACs are unique to poplar and not found in other tree species. The sequencing of additional tree genomes would help to address this issue.

The motif analysis of the variable regions in the NAC domain proteins identified conserved protein sequences shared within and between the phylogenetic clades. Therefore, researchers interested in studying NAC domain transcription factors involved in specific processes, like vascular development, can target clades with known vascular function, as well as clades that share motifs with those involved in vascular development. The VND (clade XXIV) and NST (clade XXV) clades contain transcription factors involved in vascular development [1-6]. Two motifs (XXIV-M1, XXIV-M2) are present in most members of these two vascular-specific clades. These two motifs are also in the majority of sequences in clade XXVI, and a third motif (XXV-M5/XXVI-M8) is shared by several sequences in clades XXV and XXVI. To date, none of the NAC domain proteins from clade XXVI have been studied, but our motif analysis suggests that members of this clade are likely to be regulators of vascular development.

Due to the absence of gene replacement techniques in plants, it will be difficult to assess the functional significance of the motifs present in the variable regions of NAC domain proteins with experiments done in plants. However, their potential roles in transcriptional activation and protein-protein interactions may be assessed via assays in yeast [7]. Previous research indicates that the ability to activate transcription is encoded in the variable region of multiple NAC domain proteins [7-10]. However, specific sequences involved in transcriptional activation have not been identified. Motifs shared between all or a majority of clade members provide good targets for these types of experiments. In addition, it may be possible to determine if specific variable region motifs influence the functional specificity of individual NAC domain proteins by fusing deletion constructs lacking individual motifs to the EAR dominant repressor domain [11]. . If the variable region motifs specify which gene targets are regulated by a given NAC domain

protein, then the phenotypes of transgenic plants expressing a motif-deletion+EAR construct should differ from those expressing an intact protein+EAR construct.

The functional study of PtNAC068 (Chapter 3) provides evidence that the phylogenetic relatedness of NAC domain proteins from different species (Chapter 2) can be predictive of functional similarity. In the phylogenetic analysis, PtNAC068 was identified as a member of clade XXV (the NST subfamily). Arabidopsis members (NST1, NST2, NST3) of this clade are involved in secondary wall thickening of fiber cells [1, 3-6]. Of the three Arabidopsis proteins in this subfamily, NST3 appears to be the ortholog of PtNAC068 (Figure 2.2). NST3 regulates secondary wall thickening of interfascicular and xylary fiber cells and shares functional redundancy with NST1 [3, 6]. Three other poplar NAC domain proteins are found in clade XXV and it is possible that some of them might share functional redundancy with PtNAC068. Previous experiments involving NAC domain proteins in clades XXIV and XXV have demonstrated the difficulty of detecting phenotypes using single gene mutants (RNAi or T-DNA insertional lines) in the NAC domain family [2, 4, 5]. The issue of functional redundancy masking phenotypes was over-come by fusing the gene of interest to the EAR repressor domain (strong repressor domain) for Arabidopsis NST1 and NST3 NACs [4, 5]. Over-expression and strong repressor domain PtNAC068 mutant plants showed that PtNAC068, like NST3, functions in fiber development. However, PtNAC068 appears to function specifically as a positive regulator of secondary wall development in phloem fibers and associated cells because it was not observed to have an effect on any xylem cells. PtNAC068 might be an example of a gene which has undergone subfunctionalization to make it specialized for phloem fibers, while PtNAC063 or another poplar NAC domain protein in that clade might affect secondary wall development of xylary fibers.

Over all, the phylogenetic analysis of Arabidopsis, poplar, and rice NAC domain transcription factor proteins and the functional analysis of PtNAC068 add to our current knowledge of the NAC domain family in several ways. The phylogenetic analysis is the first study of the NAC domain family to include a complete set of poplar and rice NAC domain proteins, and it is the first to determine if variable region motifs in NAC domain proteins are shared between phylogenetic clades, as well as within them. The functional characterization of PtNAC068 represents the first mutational analysis of a poplar NAC domain protein, and it is the first use of the strong repression technology (EAR domain) in poplar. In addition, PtNAC068 is the first transcription factor reported to be involved in phloem fiber development in trees.

References

1. Ko JH, Yang SH, Park AH, Lerouxel O, Han KH: **ANAC012, a member of the plant-specific NAC transcription factor family, negatively regulates xylary fiber development in Arabidopsis thaliana**. *Plant Journal* 2007, **50**(6):1035-1048.
2. Kubo M, Udagawa M, Nishikubo N, Horiguchi G, Yamaguchi M, Ito J, Mimura T, Fukuda H, Demura T: **Transcription switches for protoxylem and metaxylem vessel formation**. *Genes & Development* 2005, **19**(16):1855-1860.
3. Mitsuda N, Iwase A, Yamamoto H, Yoshida M, Seki M, Shinozaki K, Ohme-Takagi M: **NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of Arabidopsis**. *Plant Cell* 2007, **19**(1):270-280.
4. Mitsuda N, Seki M, Shinozaki K, Ohme-Takagi M: **The NAC transcription factors NST1 and NST2 of Arabidopsis regulate secondary wall thickenings and are required for anther dehiscence**. *Plant Cell* 2005, **17**(11):2993-3006.
5. Zhong RQ, Demura T, Ye ZH: **SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of Arabidopsis**. *Plant Cell* 2006, **18**(11):3158-3170.
6. Zhong RQ, Richardson EA, Ye ZH: **Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of Arabidopsis**. *Planta* 2007, **225**(6):1603-1611.
7. Hegedus D, Yu M, Baldwin D, Gruber M, Sharpe A, Parkin I, Whitwill S, Lydiate D: **Molecular characterization of Brassica napus NAC domain transcriptional activators induced in response to biotic and abiotic stress**. *Plant Molecular Biology* 2003, **53**(3):383-397.

8. Duval M, Hsieh TF, Kim SY, Thomas TL: **Molecular characterization of AtNAM: a member of the Arabidopsis NAC domain superfamily**. *Plant Molecular Biology* 2002, **50**(2):237-248.

9. Taoka K, Yanagimoto Y, Daimon Y, Hibara K, Aida M, Tasaka M: **The NAC domain mediates functional specificity of CUP-SHAPED COTYLEDON proteins**. *Plant Journal* 2004, **40**(4):462-473.

10. Xie Q: **Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development**. *Genes & Development* 2000, **14**(23):3024-3036.

11. Hiratsu K, Matsui K, Koyama T, Ohme-Takagi M: **Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in Arabidopsis**. *Plant Journal* 2003, **34**(5):733-739.