

DIVERSIFICATION OF CLASS III OF THE HD-ZIP GENE FAMILY:
IMPLICATIONS FOR LAMINAR EVOLUTION IN ANGIOSPERMS

by

WENLI LI

(Under Direction the of Robert K. Kuzoff)

ABSTRACT

This thesis addresses natural history of the HD-Zip gene family, especially class III (HDZ-III). I explore patterns of sequence variation in HDZ-III across the land plants and provide estimates of the phylogenetic history of the gene family using a variety of inference methods. I also calibrate a molecular clock of the gene family and use it to provide estimates of minimum divergence times for key clades and gene duplication events in the family. I investigate patterns of gene expression for members of HDZ-III and selected members of the YABBY family. I provide evidence of clade-specific divergences in expression patterns, suggesting that promoter function changed following some ancient gene duplication events. Finally, I attempt to draw upon the phylogenetic, temporal, and gene expression data to reconstruct a tentative model of the relationships between events in the natural history of HDZ-III and the diversification of land plants.

INDEX WORDS: Class III HD-ZIP, patterns of sequence evolution, angiosperms
Molecular clock, divergence and duplication time, expression

DIVERSIFICATION OF CLASS III OF THE HD-ZIP GENE FAMILY:
IMPLICATIONS FOR LAMINAR EVOLUTION IN ANGIOSPERMS

by

WENLI LI

B.S., Yunnan University, P. R. China, 2000

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

MASTER OF SCIENCES

ATHENS, GEORGIA

2006

© 2006

Wenli Li

All Rights Reserved

DIVERSIFICATION OF CLASS III OF THE HD-ZIP GENE FAMILY:
IMPLICATIONS FOR LAMINAR EVOLUTION IN ANGIOSPERMS

by

WENLI LI

Major Professor: Robert K. Kuzoff

Committee: Chris Peterson
Joseph Nairn

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2006

DEDICATION

To Xiangsheng Li and Yuhua Xu

ACKNOWLEDGEMENTS

Words could not possibly explain how I am grateful to my major advisor, Dr. Robert Kuzoff. He is the one guiding me into this intriguing field of developmental evolution. I took several high quality courses from him and gained a tremendous amount by working with him as a graduate student in the area of molecular developmental evolution and computational phylogenetic analysis. Countless technical help and stimulating discussions from him make this research possible. Whenever I need help, he is always there for me. It is hard to imagine what my life in graduate school would be without his constant support and encouragement.

I thank Dr. Chris Peterson and Dr. Joeseeph Nairn for serving as my committee members. I encounter a lot of difficulties along the way of my graduate study. But my committee members are very encouraging to me. Special thank to Dr. Nairn for providing me with the *Pinus* cDNA library.

I feel fortunate to have Bing Ma working with me in the same lab. She helped me with many of the lab works.

I appreciate all the help from the faculty, staff and fellow graduate students in the Plant Biology Department. Especially thank Drs. Lee Pratt for helping with growing seedlings, Dr. David Porter for support in my graduate study here and writing wonderful recommendations for me, Michael Boyd and Andy Tull for assistance in the greenhouse, Brian Perkins for solving the problems I had with computer and network, main office staff for assistance in various things I need help with, friendship and pleasant experience with fellow graduate students.

The wonderful teaching experience and encouraging comments I got from the students I taught are always something that motivates me to do better in my teaching and research.

This work is supported by Start-up funds of my advisor and the graduate student small grant from the Plant Biology Department.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	V
LIST OF TABLES.....	IX
LIST OF FIGURES	X
CHAPTER	
1 INTRODUCTION: AN OVERVIEW OF PLANT DEVELOPMENTAL	
EVOLUTION	1
PLANT DIVERSITY	2
PLANT DEVELOPMENT	3
HD-ZIP GENE FAMILY	6
GENE FAMILY EVOLUTION	7
THESIS OBJECTIVES	8
REFERENCES	10
2 EXPLORE THE NATURAL HISTORY OF HD-ZIP GENE FAMILY,	
CLASS III.....	17
INTRODUCTION.....	18
RESULTS	27
DISCUSSION	31
CONCLUSION.....	34
REFERENCES	36

3	COMPARATIVE STUDY OF GENES REGULATING LAMINAR DEVELOPMENT IN <i>LYCOPERSICON ESCULENTUM</i> AND <i>MEDICAGO TRUNCULATA</i>	67
	INTRODUCTION.....	68
	MATERIAL AND METHODS	71
	RESULTS	75
	DISCUSSION	78
	REFERENCES	82
4	CONCLUSION.....	108
	REFERENCES	111

LIST OF TABLES

	Page
Table 2.1: Dgenerate primers used to amplify initial fragments of HD-ZIP, class III genes	40
Table 2.2: Taxa sampled, gene names and GenBank numbers for HD-ZIP genes analyzed in this study	41
Table 2.3: L _g Likelihood calculated by JTT amino acid substitution matrix in TREEPUZZLE for trees saved with branch length and generated through a series of weighting matrix and character-weighting method	43
Table 2.4: L _g Likelihood calculated by Dayhoff amino acid substitution matrix in TREEPUZZLE for trees saved with branch length and generated through a series of weighting matrix and character-weighting method	44
Table 2.5: Calibration of molecular clock using a pool of trees from weighted parsimony and Bayesian analysis	45
Table 2.6: Molecular clock calibration and dating of ancient duplication events within HD-ZIP, class III gene family	47
Table 2.7: Molecular clock calibration and dating of recent duplication events within HD- ZIP, class III gene family	48
Table 3.1: In-situ hybridization conditions and results in this study	86

LIST OF FIGURES

	Page
Figure 2.1 Relative locations of forward and reverse primers used in amplification of HDZ-III homologs from diverse land plant species	49
Figure 2.2 Relative abundance of each amino acid in the full alignment of all the sequences analyzed in this study	51
Figure 2.3 Relative sequence variation across full alignment.....	53
Figure 2.4 Frequency of unambiguous changes among amino acids	55
Figure 2.5 One of two optimal weighted parsimony trees	57
Figure 2.6 Plots indicating the “burn-in” period for McMC analyses using three amino acid transition matrices	59
Figure 2.7 Majority-rule consensus tree based on 1,000 trees generated from a Bayesian analysis using the BLOSSUM AATM.....	61
Figure 2.8 Minimum age estimates of events in the ancient history of the HDZ-III gene lineage based on BLOSSUM45 weighted-parsimony tree	63
Figure 2.9 Minimum age estimates of eight recent gene duplication events in the HDZ-III gene lineage based on BLOSSUM weighted-parsimony tree	65
Figure 3.1 Expression for <i>TomFIL</i> in longitudinal and cross section of leaves.....	88
Figure 3.2 Expression for <i>TomFIL</i> and <i>TomCOR</i> in flower meristem, young petal and petal primordium.....	90

Figure 3.3 Expression for <i>TomREV</i> and <i>TomCOR</i> in FM, SAM, petal primordium and cross section of older leaf	92
Figure 3.4 Expression of <i>TomPHB</i> in SAM compared to the expression of sense probe	94
Figure 3.5 Expression of <i>TomCOR</i> , <i>TomPHB</i> in Tomato immature root compared to the expression of sense probe in immature root	96
Figure 3.6 Expression of <i>TomCOR</i> , <i>TomREV</i> and <i>TomPHB</i> in immature shoot compared to the expression of sense probe.....	98
Figure 3.7 Expression for <i>MedFIL</i> in young leaf primordial and anlagen	100
Figure 3.8 Expression of <i>MedREV</i> and <i>MedPHB</i> in mature shoot apical meristem (SAM) compared to the expression of sense probe.....	102
Figure 3.9 Expression of <i>MedPHB</i> and <i>MedREV</i> in immature root in contrast with the sense probe expression	104
Figure 3.10 Expression of <i>MedPHB</i> and <i>MedCOR</i> in cross section of immature shoot	106

CHAPTER 1

INTRODUCTION: AN OVERVIEW OF PLANT DEVELOPMENTAL EVOLUTION

PLANT DIVERSITY

At least since the days of Goethe (1790), humans have puzzled over the tremendous morphological variety seen in terrestrial plant species and the historical forces that have given rise to it. From the time they first emerged from the oceans, land plants have evolved an astonishing range of sizes and shapes in every part of their body (Stern, 2006). Beneath the ground they may have little more than thinly filamentous rhizoid hairs, as seen in many mosses and seedless vascular plants, or enormous woody root systems sturdy enough to lift the foundations of a house or highway, as in many arboreal species of gymnosperms and angiosperms. Above the ground, shoots range from frail and ephemeral, as with many short-lived herbaceous species, like the glacier lily, to elongate and serpentine, as with the kudzu vine, to massive and enduring, as with eucalyptus or the coastal redwoods.

Perhaps the most visually stunning variation in plant form is seen in the sizes and shapes of structures borne laterally on the shoot. For example leaves, if present, may be one to many cell layers thick and have blades that are entire with smooth margins, as in *Gnetum*, or dissected with jagged edges, as in mahonia. Reproductive organs may consist of a cluster of cells on the undersurface of an apparently amorphous disc, as in the archegonia and antheridia of free-living gametophytes, to regal and expansive structures, such as in the massive solitary flowers of *Rafflesia*. Though undeniably complex and daunting, resolving the details of the natural processes that produced this seemingly endless variety of plant forms is an enthralling problem.

A necessary prerequisite for untangling the evolutionary processes that gave rise to the countless forms that plant structures have taken is an understanding of their phyletic relationships. Fortunately, progress in recent years toward reconstructing the lineal relationships within the land plants has been substantial (e.g., Pryer *et al.*, 2001, 2004; reviewed in Palmer *et*

al., 2004). As has long been suspected, the earliest diverging lineages of embryophytes are the mosses, hornworts, and liverworts. Though the specific branching order among these groups remains unclear, most recent studies agree that the bryophytes are the sister to all vascular plants. Within the vascular plant lineage, the first-diverging branch is the clade comprising all lycopods. The remainder of the vascular plants split into two groups, which are the seed plants and the monilophytes, a clade that includes all the ferns and fern allies (other than the lycopods). Within the seed plants, angiosperms appear to be the sister group to a monophyletic gymnosperm lineage in recent molecular systematic studies (Burleigh and Mathews, 2004). Relationships among the major lineages of flowering plants have been independently and corroboratively determined by a suite of recent studies that show *Amborella*, the Nymphaeales, and the Austrobaileyales as the first three successive branches of angiosperms (reviewed in Kuzoff and Gasser, 2000). Thereafter, monocots split from the core magnoliid and eudicot clades, the latter comprising approximately three fourths of all flowering plants. Finally, a recent study by Schneider *et al.* (2004) provides estimates for minimum divergence times between these clades. Encouraged by this clarified picture of land plant phylogeny, contemporary evolutionary biologists have sought to incorporate recent findings from the field of developmental biology into their analysis of plant diversification.

PLANT DEVELOPMENT

Plant development is a complex process that can be studied through a variety of means. Historically, light microscopy-based investigation of sectioned plant organs and experimental micro-dissection helped to define the range of tissue types and morphogenic fields present in the plant body (e.g., Johansen, 1940; Esau, 1977; Steeves and Sussex, 1989; Fahn, 1990). Studies

that use scanning electron microscopy provided further intimate insights into the patterns of organogenesis in vegetative and floral apices (e.g., Endress, 1996; Kaplan, 2001). Although these lines of evidence form a necessary foundation to comprehending plant developmental processes, the major emphasis of more recent study has been the elucidation of the molecular determinants of plant morphogenesis (reviewed in Leyser and Day, 2003). Sustained genetic and molecular-genetic analysis of mutant phenotypes in a variety of model research organisms has yielded an elaborate picture of the regulation of developmental pathways governing development in roots, shoots, and many kinds of appendages.

Of special interest to the present study is the establishment of dorsoventral organ polarity in shoot lateral organs, especially leaves and floral appendages (reviewed in Bowman *et al.*, 2002). Mutants of snapdragon (Waites and Hudson, 1995; Luo *et al.*, 1996; Golz *et al.*, 2004), maize (Timmermans *et al.*, 1999; Juarez *et al.*, 2004a, b), tomato (Kim *et al.*, 2003a, b), *Petunia* (Nakagawa *et al.*, 2004), and *Arabidopsis* (Sawa *et al.*, 1999a; McConnell and Barton, 1998; Eshed *et al.*, 1999; McConnell *et al.*, 2001; Emery *et al.*, 2003; Eshed *et al.*, 2004; Kidner and Martienssen, 2004) have all provided conceptual inroads into the genetic mechanisms underlying dorsoventral differentiation in lateral shoot organs. In *Arabidopsis*, the *filamentous flower* (*fil*) mutation lacks a leaf phenotype, but produces flowers which are either reduced to adaxialized filaments or have adaxialized appendages. Consequently, it was proposed that *FIL* promotes abaxial identity in flowers and floral appendages (Sawa *et al.*, 1999b). *CRABS CLAW* (*CRC*) was identified in a screen for carpel mutants and was later shown to be involved in abaxial differentiation (Bowman and Smyth, 1999). *KANADI* (*KAN*), an additional genetic factor regulating dorsoventrality, was identified through a screen for mutations that enhance the *crc* mutant phenotype (Eshed *et al.*, 1999). *KAN* was shown to promote abaxial identity and to

negatively regulate adaxial factors (Eshed *et al.*, 2001; Kersetter *et al.*, 2001). Other genetic factors regulating adaxial identity in lateral shoot organs identified through mutant analyses include *PHABULOSA* (*PHB*), *PHAVOLUTA* (*PHV*), and *REVOLUTA* (*REV*). It has been proposed that *PHB* and *PHV* may interact directly with a signaling-ligand traveling from the shoot apex to nascent leaves (McConnell *et al.*, 2001). However, more recent studies reveal that *PHB*, *PHV*, and *REV* are more likely regulated by miRNAs (Emery *et al.*, 2003). Based on these findings, a model has been proposed in which adaxial- and abaxial-determining factors are expressed in their respective portions of emerging lateral organ primordia and act as negative regulators of one another.

Published sequences of genes that govern abaxial-adaxial polarity determination reveal that they belong primarily to four gene families. *PHANTASTICA* belongs to the MYB family (Waites *et al.*, 1998); *CRC*, *FIL* and *INO* belong to the YABBY family (Bowman and Smyth, 1999; Siegfried *et al.*, 1999); and *KAN* belongs to the GARP family (Eshed *et al.*, 2001; Kersetter *et al.*, 2001). *PHB*, *PHV* and *REV* belong to the narrowly circumscribed and well-supported class III lineage within the HD-ZIP family (McConnell *et al.*, 2001; Sakakibara *et al.*, 2001; Emery *et al.*, 2003). Phylogenetic relationships and the molecular evolutionary history of each of these gene families have been studied previously, at least preliminarily, providing an additional set of tools with which to explore processes of plant developmental evolution (Martin and Paz-ares, 1997; Keiko *et al.*, 1999; Eshed *et al.*, 2001; Lee and Shiefelbein, 2001; Sakakibara *et al.*, 2001).

HD-ZIP GENE FAMILY

Among the gene families that play an integral role in both the abaxial-adaxial differentiation of shoot lateral organs, the HD-ZIP gene family, particularly class III, will be the focus of attention throughout much of this thesis. Developmentally regulated members of the HD-ZIP family of genes that play crucial roles in plant ontogeny are required for vascular differentiation in roots, stems and for pattern formation in leaves, diverse shoot lateral organs, and ovule integuments. Accordingly, this family has been the recent focus of intense molecular-genetic research in a set of model research organisms, especially *Arabidopsis*, maize, and tomato (e.g., Emery *et al.*, 2003; Kim *et al.*, 2003a; Kim *et al.*, 2003b; Prigge *et al.*, 2004). Investigated protein products of these gene family members, including *REVOLUTA* (*REV*), *PHABULOSA* (*PHB*), and *PHAVOLUTA* (*PHV*), are transcription factors, with two conserved DNA binding domains, a homeodomain (HD) and beta-leucine zipper (ZIP), and are the targets of both hormonal and miRNA regulation through a StAR-related lipid-transfer (START) domain (Ponting and Arvind, 1999; Schrick *et al.*, 2004).

Genes with either a leucine zipper motif or a homeodomain have been reported in metazoa, fungi and green plants. Schena and Davis (1997) speculated that the HD-ZIP genes originated in green plant lineages by exon shuffling between a gene encoding a homeodomain and another gene encoding a leucine zipper motif. HD-ZIP genes have been isolated from moss, suggesting that HD-ZIP genes originated before the divergence of the mosses and vascular plants (Sakakibara, *et al.*, 2001).

Previous investigation reveals that the HD-ZIP family is conserved across land plant phylogeny and contains four distinct constituent lineages, or classes (I-IV). These four distinct lineages within the HD-ZIP gene family have been recognized based on prior phylogenetic

analysis : HD-ZIP I, HD-ZIP II, HD-ZIP III and HD-ZIP IV (Meijer *et al.*, 1997; Aso *et al.*, 1999). Investigated members of the four sub-clades appear to have distinct functional roles: HD-ZIP I and II genes play roles in light, auxin and ABA signal transduction networks (Carabelli *et al.*, 1993, 1996; Steindler *et al.*, 1997, 1999; Kawahara *et al.*, 1995), which are involved in general plant growth regulation. HD-ZIP III genes play important roles in plant vascular and leaf development, establishment of apical meristem, organ polarity and vascular tissue patterning (McConnell and Barton, 1998; McConnell *et al.*, 2001; Otsuga *et al.*, 2001; Emery *et al.*, 2003; Ohashi-Ito and Fukuda, 2003; Eshed *et al.*, 2004). HD-ZIP IV genes appear to play a developmental role in the differentiation of the outmost cell layer of diverse shoot organs (Cristina *et al.*, 1996; Ingram *et al.*, 1999; Kubo *et al.*, 1999).

GENE FAMILY EVOLUTION

A vast majority of genes identified in sequenced genomes belong to gene families. Gene families are evolutionary units that share a common ancestor and a unique natural history. A gene family grows as paralogous genes are formed by gene duplication events. Gene duplication is prevalent across the tree of life and gene duplication events arise at an approximate rate of one per million years in eukaryotes (Lynch and Conery, 2000). By far, the most common fate of duplicate gene pairs is for one of them to become a pseudogene, which is either unexpressed or functionless (Walsh, 1995; Lynch *et al.*, 2001). If the gene pairs are retained, they have potential three fates. The first, functional redundancy, is the least common outcome and is most likely with highly expressed genes, such as housekeeping genes. The second is neofunctionalization, which is the most common fate of retained duplicate gene pairs in single celled organisms. Neofunctionalization of duplicated genes involves an elevated rate of amino acid substitutions in

generally conserved regions that were previously under stabilizing selection (Zhang, 2003). The third possible fate of duplicate gene pairs is subfunctionalization, which is most common in multicellular organisms. There are at least two modes of subfunctionalization: coding region subfunctionalization and promoter subfunctionalization. Coding region subfunctionalization occurs at the protein function level and entails complementary loss of function in more than one domain between duplicate gene pairs. Subfunctionalization leads to functional specialization of the daughter genes (Hughes, 1999). Promoter subfunctionalization refers to a similar process in which duplicate gene pairs exhibit complementary loss of regulatory elements in their promoters, leading to differential expression patterns (Force, *et al.*, 1999).

THESIS OBJECTIVES

The goals of this thesis are to describe and interpret the results of a series of experiments into the natural history of the HD-Zip gene family, especially class III (HDZ-III). In Chapter 2, I explore patterns of sequence variation in HDZ-III across the land plants and provide estimates of the phylogenetic history of the gene family using a variety of inference methods. Although Aso *et al.* (1999) provide a preliminary phylogenetic study for the gene family, they sampled only a small number of genes and only one angiosperm from class III. In Chapter 2, I attempt to provide a more detailed phylogenetic history of this crucial lineage within the family. I also calibrate a molecular clock of the gene family and use it to provide estimates of minimum divergence times for key clades in the family, and to estimate minimum ages of ancient and recent gene duplication events. In Chapter 3, I investigate patterns of gene expression for members of HDZ-III and selected members of the YABBY family, which play an antagonistic role to at least some HDZ-III genes during development of shoot lateral organs. I provide evidence of clade-specific

divergences in expression patterns, suggesting that promoter function changed following some ancient gene duplication events. Finally, I attempt to draw upon the phylogenetic, temporal, and gene expression data to reconstruct a tentative model of the relationships between events in the natural history of HDZ-III and the diversification of land plants.

REFERENCES

- Aso, K., Kato, M., J. A. Banks, M. Hasebe. 1999. Characterization of homeodomain leucine zipper genes in the fern, *Ceratopteris richardii* and the evolution of the homeodomain-leucine zipper gene family in vascular plants. *Mol. Biol. Evol.* **16**: 544-551.
- Bharathan, G., B. Janssen, E. A. Kellogg, N. Sinha. 1997. Did homeodomain proteins Duplicate before the origin of angiosperms, fungi, and metazoa? *Proc. Natl. Acad. Sci.* **94**: 13749-13753.
- Bowman, J. L., Eshed, Y., and Baum, S. 2002. Establishment of polarity in angiosperm lateral organs. *Trends Genet.* **18**: 134-141.
- Bowman, J. L., and Smyth, D. R. 1999. *CRABS CLAW*, a gene that regulates carpel and nectary development in *Arabidopsis*, encodes a novel protein with zinc finger and helix-loop-helix domains. *Development* **126**: 2387-2396.
- Burglin, T. R. Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. 1997. *Nucleic Acids Res.* **25**: 4173-80.
- Burleigh, J. G., and Mathews, S. Phylogenetic signal in nucleotide data from seed plants: implications for resolving the seed plant tree of life. *Amer. J. Bot.* **91**: 1599-1613.
- Carabelli, M., G. Morelli, G. Whitelam, and I. Ruberti. 1996. Twilight-zone and canopy shade induction of the *Athb-2* homeobox gene in green plants. *Proc. Natl. Acad. Sci. USA.* **93**: 3530-3535.
- Carabelli, M., G. Sessa, S. Baima, G. Morelli, and I. Ruberti. 1993. The *Arabidopsis* *Athb-2* and *-4* genes are strongly induced by far-red-rich light. *Plant Journal.* **4**: 469

479.

- Chan, R. L., G. M. Gago, C. M. Palena, and D. H. Gonzalez. 1998. Homeoboxes in plant development. *Biochem. Biophys. Acta*. **1442**: 1-19.
- Cristina, M. D., G. Sessa, L. Dolan, P. Linstead, S. Baima, I. Ruberti, and G. Morelli. 1996. The Arabidopsis Athb-10 (*GLABRA2*) is an HD-Zip protein required for regulation of root hair development. *Plant J.* **10**: 393-402.
- Emery, J. F, Floyd, S. K., Alvarez, J., Eshed, Y., Hawker, N. P., Izhaki, A., Baum, S. F., Bowman, J. L. 2003. Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes. *Curr Biol.* **13**: 1768–1774
- Endress, P. K. 1996. *Diversity and evolutionary biology of tropical flowers*. New York: Cambridge University Press.
- Esau, K. 1977. *Anatomy of Seed Plants*, 2nd ed. Somerset, NJ: John Wiley & Sons.
- Eshed, Y., Baum, S. F., and Bowman, J. L. 1999. Distinct mechanisms promote polarity establishment in carpels of Arabidopsis. *Cell.* **99**: 199-209.
- Eshed, Y., Baum, S.F., Perea, J.V., and Bowman, J. L. 2001. Establishment of polarity in lateral organs of plants. *Curr. Biol.* **11**: 1251-1260.
- Eshed, Y. Izhaki, A., Baum, S. F., Floyd, S. K., and Bowman, J. L. 2004. Asymmetric leaf development and blade expansion in *Arabidopsis* are mediated by KANADI and YABBY activities. *Development* **131**:2997-3006.
- Fahn, A. 1990. *Plant Anatomy*, 4th ed. New York: Elsevier.
- Floyd, S. and Bowman, J. 2004. Ancient microRNA target sequences in plants. *Nature.* **428**: 485-485.
- Force, A., Lynch, M., Bryan Pickett, F., Amores, A., Yan, Y. and Postlethwait, J. 1999.

- Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531-1545.
- Goethe, J. W. von. 1790. Versuch die Metamorphose der Pflanzen zu erklären. Ettinger, Gotha, Germany.
- Golz, J. F., M. Roccaro, R. K. Kuzoff, and A. Hudson. 2004. *GRAMINIFOLIA* promotes growth and polarity of *Antirrhinum* leaves. *Development* **131**: 3661-3670.
- Hughes, A. L. 1999. Adaptive evolution of genes and genomes. Oxford University Press.
- Ingram, G. C., J. L. Magnard, P. Vergne, C. Dumas, and P. M. Rogowsky. 1999. *ZmOCL1*, an *HDGL2* family homeobox gene, is expressed in the outer cell layer throughout maize development. *Plant Mol. Biol.* **40**: 343-354.
- Johansen, D. A. 1940. Plant microtechnique. McGraw-Hill, New York, New York, USA.
- Juarez, M. T., Kui, J. S., Thomas, J., Heller, B. A., and Timmermans, M. C. P. 2004a. microRNA-mediated repression of *rolled leaf1* specifies maize leaf polarity. *Nature* **428**: 84-88.
- Juarez, M. T., Twigg, R. W., and Timmermans, M. C. P. 2004b. Specification of adaxial cell fate during maize leaf development. *Development* **131**: 4533-4544.
- Kaplan, D. R. 2001. Fundamental concepts of leaf morphology and morphogenesis: a contribution to the interpretation of molecular genetics. *Int. J. Pl. Sci.* **162**: 465-474.
- Kersetter, R. A., Bollman, K., Taylor, R.A., Bomblies, K., and Peothig, R.S. 2001. *KANADI* regulates organ polarity in Arabidopsis. *Nature* **411**: 706-709.
- Kidner, C. A., and Martienssen, R. A. 2004. Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. *Nature* **428**, 81-84.
- Kim, M., McCormick, S., Timmermans, M., and Sinha, N. R. 2003. The

- expression domain of *PHANTASTICA* determines leaflet placement in compound leaves. *Nature* **424**: 438-443.
- Kim, M., Pham, T., Hamidi, A., McCormick, S., Kuzoff, R. and Sinha, N. R. 2003. Reduced leaf complexity in tomato wiry mutants suggests a role for *PHAN* and *KNOX* genes in generating compound leaves. *Development* **130**: 4405-4415.
- Kuzoff, R. K., and Gasser, C. S. Recent progress in reconstructing angiosperm phylogeny. *Trends Pl. Sci*, **5**: 330-336.
- Luo, D. *et al.* 1999. Control of organ asymmetry in flowers of *Antirrhinum*. *Cell* **99**: 367-376.
- Lynch, M. and Conery, J. S. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151-1155.
- Lynch, M., O'Hely, M., Walsh, B. and Force, A. 2001. The probability of preservation of a newly arisen gene duplicate. *Genetics* **159**: 1789-1804.
- McConnell, J. R, Barton, M. K. 1998. Leaf polarity and meristem formation in *Arabidopsis*. *Development* **125**: 2935-2942
- McConnell, J. R., Emery, J. F., Eshed, Y., Bao, N., Bowman, J., Barton, M. K. 2001. Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. *Nature* **411**: 709-713
- Meijer, A. H., E. Scarpella, E. L. van Dijk, L. Qin, A. J. Taal, S. Rueb, S. E. Harrinton, McCouch, S. R., Schilperoort, R. A. and Hoge, J. H. 1997. Transcriptional repression by Oshox1, a novel homeodomain leucine zipper protein from rice. *Plant J.* **11**: 263-76.

- Nakagawa, H. Ferrario, S., Angenent, G.C., Kobayashi, A., and Takatsuli, H. 2004. The *Petunia* ortholog of *Arabidopsis SUPERMAN* plays a distinct role in floral organ morphogenesis. *The Plant Cell* **16**:920-932.
- Ohashi-Ito, K. and Fukuda, H. 2003. HD-Zip III homeobox genes that include a novel member, ZeHB-13 (*Zinnia*)/ATHB-15 (*Arabidopsis*), are involved in procambium and xylem cell differentiation. *Plant Cell Physiol* **44**: 1350–1358
- Otsuga, D., Deguzman, B., Prigge, M., Drews, G., and Clark, S. 2001. REVOLUTA regulates meristem initiation at lateral positions. *Plant J.* **25**: 223–236.
- Palmer, J. D., Soltis, D. E., and Chase, M. W. 2004. The plant tree of life: an overview and some points of view. *Amer. J. Bot.* **91**: 1437-1445.
- Ponting, C.P., and Aravind, L. 1999. START: A lipid-binding domain in StAR, HD-ZIP and signalling proteins. *Trends Biochem. Sci.* **24**: 130–132.
- Prigge, M. J., Otsuga, D., Alonso, J. M., Ecker J. R., Drews, G. N., Clark, S. E. 2005. Class III homeodomain-leucine zipper gene family members have overlapping, Antagonistic, and distinct roles in *Arabidopsis* development. *Plant Cell.* **17**: 61-76.
- Pryer, K. M., Schneider, H., Smith, A. R., Cranfill, R., Wolf, P. G., Hunt, J. S., and Snipes, S. D. 2001. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* **409**: 618-622.
- Pryer, K. M. Schuettpelz, E., Wolf, P. G., Schneider, E., Smith, A. R., and Cranfill, R. 2004. Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate

- divergences. *Amer. J. Bot.* **91**: 1582-1598.
- Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B., Bartel, B., Bartel, D.P. 2002. Prediction of plant microRNA targets. *Cell.* **110**: 513-20.
- Reiser, L., Sánchez-Baracaldo, P. and Hake, S. 2000. Knots in the family tree: Evolutionary relationships and functions of *knox* homeobox genes. *Plant Molecular Biology* **42**: 151-166.
- Sakakibara, K., Nishiyama, T., Sumikawa, N., Kofuji, R., Murata, T., and Hasebe, M. 2003. Involvement of auxin and a homeodomain-leucine zipper I gene in rhizoid development of the moss *Physcomitrella patens*. *Development* **130**: 4835–4846.
- Sakakibara, K., Nishiyama, T., Kato, M. and Hasebe, M. 2001. Isolation of homeodomain-leucine zipper genes from the moss *Physcomitrella patens* and the evolution of homeodomain-leucine zipper genes in land plants. *Mol. Biol. Evol.* **18**: 491-502.
- Sawa, S., Ito, T., Shimura, Y., and Okada, K. 1999a. *FILAMENTOUS FLOWER* controls the formation and development of Arabidopsis inflorescences and floral meristems. *Plant Cell* **11**: 69-86.
- Sawa, S., Watanabe, K., Goto, K., Kanaya, E., Morita, E. H., and Okada, K. 1999b. *FILAMENTOUS FLOWER*, a meristem and organ identity gene of Arabidopsis, encodes a protein with a zinc finger and HMG-related domains. *Genes Develop.* **13**: 1079-88.
- Schena, M. and Davis, R. W. 1994. Structure of homeobox-leucine zipper genes suggests a model for the evolution of gene families. *Proc Natl. Acad. Sci. U S A.* **91**: 8393-8397.
- Schneider, H. Schuettpelz, E., Pryer, K. M., Cranfill, R., Magallon, S., and Lupia, R. 2004. Ferns

- diversified in the shadow of angiosperms. *Nature* **428**: 553-557.
- Schrick, K., Nguyen, D., Karlowski, W. M., Mayer, K. F. 2004. START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors. *Genome Biology*. **5**: R41
- Steeves, T. A. and Sussex, I. M. 1989. *Patterns in Plant Development*, 2nd ed. Cambridge University Press, Cambridge.
- Steindler, C., M. Carabelli, U. Borello, G. Morelli, and I. Ruberti. 1997. Phytochrome A, phytochrome B and other phytochrome(s) regulate ATBB-2 gene expression in etiolated and green Arabidopsis plants. *Plant Cell Environ.* **20**: 759-763.
- Steindler, C., A. Matteucci, G. Sessa, T. Weimar, M. Ohgishi, T. Aoyama, G. Morelli, and I. Ruberti. 1999. Shade avoidance responses are mediated by the ATHB-2 HD-Zip protein, a negative regulator of tene expression. *Development* **126**: 4235-4245,
- Timmermans, M.C.P. *et al.* 1999. ROUGH SHEATH2: a myb protein that represses *knox* *homeobox* genes in maize lateral organ primordial. *Science* **284**: 151-153.
- Waites, R. and Hudson, A. 1995. *phantastica*: a gene required for dorsiventrality in leaves of *Antirrhinum majus*. *Development* **121**, 2143-2154.
- Walsh, J. B. 1995. How often do duplicated genes evolve new functions? *Genetics* **139**: 421-428.
- Zhang, J. 2003. Evolution by gene duplication: an update. *Trends in Ecology and Evolution*. **18**: 292-298.

CHAPTER 2

EXPLORE THE NATURAL HISTORY OF HD-ZIP GENE FAMILY, CLASS III

INTRODUCTION

Developmentally regulated members of the HD-ZIP family of genes play crucial roles in plant ontogeny and are required for vascular differentiation in roots and stems and for pattern formation in leaves, diverse shoot lateral organs, and ovule integuments (Bharathan, *et al.*, 1997; Bürglin, 1997, 1998; Chan, *et al.*, 1998; Bowman, 2003). Accordingly, this family has been the recent focus of intense molecular-genetic research in a set of model organisms, especially *Arabidopsis*, maize, and tomato (e.g., Emery, *et al.* 2003; Kim, *et al.* 2003a; Kim, *et al.* 2003b; Prigge, *et al.* 2004). Investigated protein products of gene family members, including *REVOLUTA* (*REV*), *PHABULOSA* (*PHB*), and *PHAVOLUTA* (*PHV*), are transcription factors, with two conserved DNA binding domains, a homeodomain (HD) and beta-leucine zipper (ZIP), and are the targets of both hormonal and miRNA regulation through a StAR-related lipid-transfer (START) domain (Ponting and Arvind, 1999; Schrick, *et al.* 2004).

The homeodomain (or homeobox) is a conserved DNA-binding region made up of approximately 60 amino acids that is found in transcription factors from distantly related eukaryotic species, including animals, fungi, and plants (Bharathan, *et al.*, 1997). Genes that contain a homeodomain are divided into two groups: typical and TALE homeobox genes. TALE homeobox genes have three more residues between helix 1 and helix 2 than do typical homeobox genes. Homeodomains have a wide range of functions including control of cell fate determination, mating-type recognition in fungi, and various forms of morphogenesis in plants (Bürglin, 1997, 1998; Chan, *et al.*, 1998). In angiosperms, KNOTTED -like homeobox genes (KNOX genes) are known to be crucial regulators of meristem function and they likely contributed to the evolution of compound leaves in tomato and nectar spurs in *Antirrhinum* (Reiser, *et al.*, 2000; Golz, *et al.*, 2002; Kim, *et al.*, 2003).

At the carboxy-terminal end of the homeobox in all HD-ZIP genes is the beta-leucine zipper, a region that is rich in leucine residues. Hydrophobic amino acids frequently occupy regions in which leucine residues are also present. In a beta-leucine zipper, four to five highly conserved leucine residues are spaced exactly seven residues apart within an alpha-helical hydrophobic portion of a protein. This hydrophobic region serves to stabilize protein:protein interactions between protein monomers (Aso *et al.*, 1999). Accordingly, this region is believed to be involved in both dimerization of proteins and stabilizing DNA-protein interactions.

Interestingly, the combination of beta-leucine zipper adjacent to a homeodomain is believed to be unique to plants. In fact, no known homeobox-containing proteins in animals or fungi also have a beta-leucine zipper region. This suggests that the HD-ZIP super-domain arose after the divergence of plants from animals and fungi (Schrack, *et al.*, 2004; Schena and Davis, 1994). Schena and Davis (1997) speculated that HD-ZIP genes originated in the green plant lineage by an event of exon shuffling between a gene encoding a homeodomain and another gene encoding a leucine zipper motif. HD-ZIP genes have recently been isolated from moss, indicating that the origin of HD-ZIP genes antedates the most recent common ancestor of all vascular plants (Sakakibara, *et al.*, 2001).

The StAR-related lipid-transfer (START) domain consists of approximately 200 amino acids and functions in binding lipid-sterol molecules. This ancient domain is present across plants, animals, bacteria and protists. But only in multicellular eukaryotes (plants, animals and multicellular protists) are START domains found in relation to domains that have established functions in signal transduction or transcriptional control. START domains are more common in plants than in animals and are not closely related to those of animals, implying collateral evolution to accommodate organism-specific lipid-sterols. In humans, about half of the START

domain-containing proteins are multi-domain proteins, whereas in Arabidopsis and rice approximately more than three-quarters of all START proteins contain an additional domain. START domain containing genes in plants fall into two categories: HD-ZLZ (zipper-loop-zipper)-START proteins and HD-ZIP-START proteins, of which class III of the HD-Zip family is an example. These two categories have differentiated from each other substantially in both their leucine-zipper and START domains (Schrack, *et al.*, 2004).

Dominant mutations in three Arabidopsis HD-ZIP III transcription factors, *REVOLUTA*, *PHABULOSA* and *PHAVOLUTA*, occur within a narrow region of the START domain. This led McConnell *et al.* (2001) to speculate that disruption of the sterol-binding pocket was the cause of the disrupted function of these mutated genes. However, this domain also contains a negative-regulatory target site to which microRNA (miRNA) derived from miR166 and miR165 adheres (Rhoades, *et al.*, 2002). Similar miRNA-target sequences are found in START domains of all surveyed HD-ZIP III genes in land plants, suggesting that they are generally regulated through a conserved microRNA-mediated mechanism that had existed since the common ancestor of the embryophytes, more than 400 Myr ago (Floyd and Bowman, 2004).

In Arabidopsis five HD-Zip III genes are known: *REVOLUTA* (REV), *PHABULOSA* (PHB/AtHB14), *PHAVOLUTA* (PHV/AtHB9), *CORONA* (COR/AtHB8) and AtHB15 (Emery, *et al.*, 2003). These family members have somewhat overlapping and somewhat divergent roles in Arabidopsis development. *REV* plays a crucial role in apical patterning in the embryo as well as embryonic and postembryonic initiation of both the shoot apical and floral meristems (SAM and FM, respectively). *REV* is the only member of HD-ZIP III that has an apparent (single) mutant phenotype. *PHB* and *PHV* are thought to be largely functionally redundant with one another. They play vital roles in postembryonic SAM and FM initiation, lateral organ and apical embryo

patterning, and meristem size regulation. *PHB*, *PHV* and *REV* are each expressed in the adaxial domain of shoot lateral organs, in the apical meristem and in vascular tissues (Zhong, *et al.*, 1997; Zhong and Ye, 1999; Otsuga, *et al.*, 2001). Triple loss-of-function *rev phb phv* plants exhibit complete loss of function in the shoot apical meristem and have abaxialized, radial cotyledons which envelope the persisting apical pole resulting in a wine glass shape embryo (Emery, *et al.*, 2003; Prigge, *et al.*, 2005). Based on comparison of this with mutant phenotypes of the *pin1 stm* double mutant and the *pin1 cuc1* and *cuc2* triple mutant, which also has radially-symmetrical embryos (though not cup shaped) it was suggested that HD-ZIP III genes may contribute to the differentiation of central and peripheral cells in the shoot apex. The expression of *AtHB8* is restricted to vascular development while *COR* is expressed at high levels in vascular tissue, but is detectable in developing lateral shoot meristem and flower meristem (Kang and Dengler, 2002; Prigge, *et al.*, 2005).

It has been hypothesized that a mutual antagonism between HD-ZIP III genes, on the one hand, and KANADI and YABBY genes, on the other, leads to the establishment of both adaxial and abaxial cell identities during morphogenesis of shoot lateral organs (McConnell, *et al.*, 2001; Eshed, *et al.*, 2001; Kerstetter, *et al.*, 2001; Bowman, *et al.* 2002; Bowman, 2003). Additionally, because the HD-ZIP III gene lineage diverged early in the embryophyte lineage and is integral to the development of several morphologically diverse plant structures, it was suggested that the initial function of this gene lineage was in auxin signaling and that it evolved as part of the auxin-signaling pathway (Bowman, 2003). It was also suggested that the ancestral function of HD-ZIP III gene lineage was involved in specifying central against peripheral polarity (Prigge, *et al.*, 2005). Further testing of these hypotheses awaits more detailed analysis of HD-ZIP III gene function in mosses and ferns, an addition to angiosperms.

Previous phylogenetic investigation reveals that the HD-ZIP family contains four distinct constituent lineages: HD-ZIP I, HD-ZIP II, HD-ZIP III and HD-ZIP IV (Meijer, *et al.*, 1997; Aso *et al.*, 1999). Investigated members of the four sub-clades appear to have distinct functional roles. HD-ZIP I and II genes are integral to light, auxin and ABA signal transduction networks (Carabelli, *et al.*, 1993, 1996; Steindler, *et al.*, 1997, 1999; Kawahara, *et al.*, 1995), which are involved in general plant growth regulation. HD-ZIP III genes play important roles in plant vascular and leaf development, establishment of the apical meristem, organ polarity, and vascular tissue patterning (McConnell and Barton, 1998; McConnell, *et al.*, 2001; Otsuga, *et al.*, 2001; Emery *et al.*, 2003; Ohashi-Ito and Fukuda, 2003; Eshed, *et al.*, 2004). HD-ZIP IV genes are integral to the differentiation of the outmost cell layer of diverse plant organs (Cristina *et al.*, 1996; Ingram *et al.*, 1999; Kubo *et al.*, 1999).

Despite their prominence in recent studies of plant development, class III of the HD-ZIP gene family has been poorly sampled in previous phylogenetic analyses. The lack of a rigorous phylogenetic context for HD-ZIP III hinders our understanding of patterns of gene family diversification and the potential role that it plays in the evolution of land plant form. Here I report the results of a broad-scaled phylogenetic analysis of HD-ZIP III across land plant phylogeny, with dense sampling across the angiosperms. I reconstruct key events in the diversification of the gene family and calibrate a molecular clock for HD-ZIP III. I use my phylogenetic topology and the molecular clock to estimate minimum ages of gene duplication and radiation events in the history of the family and evaluate their implications for embryophyte diversification.

MATERIAL AND METHODS

In silico data collection

I searched for members of the HD-ZIP III sub-family in the GenBank (<http://www.ncbi.nlm.nih.gov/>), KOME (<http://cdna01.dna.affrc.go.jp/Cdna/>), and TIGR (<http://www.tigr.org/>) databases using Arabidopsis *PHB* (NP_181018) as query sequences. 22 full length sequences were detected and subsequently aligned to facilitate degenerate primer design (Table 2.2). Retrieved protein sequences were aligned using ClustalX 1.8 (Thompson, Gibson *et al.* 1997) in the slow-accurate mode with a pair-wise alignment gap-opening penalty of 35 and extension-penalty of 0.75 and a multiple alignment gap-opening penalty of 15 and extension-penalty of 0.3. Based on this alignment, eighteen degenerate primers were designed in conserved regions to amplify internal fragments of HD-ZIP III genes from diverse land plants (Table 2.1; Fig. 2.1).

Plant tissue collection, RNA extraction and cDNA synthesis

Shoot apical meristems were collected from a wide range of plants including ferns, gymnosperms and a broad range of flowering plants (Table 2.1). Freshly collected tissue was immediately frozen in liquid nitrogen and stored at -80°C for subsequent RNA extraction. Total RNA was extracted using QIAGEN RNeasy Plant Mini Kit, following manufacturer's instructions. First strand cDNA was synthesized using Superscript III reverse transcriptase, using an 18-mer poly-T primer and following manufacturer's instructions (Invitrogen).

Degenerate PCR

I used combinations of the above mentioned 18 degenerate primers to amplify fragments

of HD-ZIP III genes from all target species. Reverse transcribed cDNA served as template for the primary PCR reactions using the combination of one of the forward and reverse primer pairs: HDZIII5'A and HDZIII3'A or HDZ_MF and HDZ_OR. The PCR products derived from primary PCR reactions using HDZIII5'A and HDZIII3'A were used to perform nested PCR reaction using HDZ_MF and HDZ_OR. The products amplified from primary PCR using HDZ_MF and HDZ_OR were used to perform nested PCR reaction using HDZ_MF or HDZ_IF (subsequently additional clade-specific primers were designed and used in a similar manner in nested PCR to increase sampling of HD-ZIP III in each species; Table 2.1 and Fig. 2.1). All PCR amplifications were performed in 30 μ l of PCR buffer (200 Mm Tris-HCl, Ph 8.4; 500 mM KCl; 50 Mm MgCl₂) containing 3 μ l of each 10 μ M 5' and 3' primers, 3 μ l dNTPs at 2 μ M, 16.8 μ l sterilized QH₂O. The PCR reaction mixtures were covered with mineral oil. PCR amplifications were carried out on an MJ Research PCT-200 Peltier Thermal-Cycler as follows: (1) a hot start at 94°C for 3 min; (2) 30 amplification cycles of 94°C for 0.5 min, 0.5 min annealing at 51°C, 3min extension 72°C; (3) a terminal extension phase at 72°C for 3 min. Amplified PCR products that were around 750 ~1150 bp were cloned using a TOPO-TA Cloning kit (Invitrogen, Carlsbad, Calif.). At least two clones were sequenced using m13-20 and m13-reverse universal primers. DNA sequences were translated into protein sequence using Sequencher based on the presence of special indels.

To increase the representation of full-length sequences in our analysis, one *Lycopodium*, *Selaginella*, *Gnetum*, *Pinus* and *Marchantia* partial sequence and all Cucumber, Medicago, Snapdragon, and Tomato partial sequences were completed by 5' and 3' RACE. For each partial sequences, two gene-specific 5' primers and two gene specific 3' primers were designed (data not shown) and used for 5' and 3' RACE using the GeneRacer kit (Invitrogen), following

manufacturer's instructions.

Alignment and phylogenetic analysis

I used ClustalX 1.8 (Thompson, Gibson *et al.*, 1997) in slow-accurate mode with the above mentioned parameters to perform a global alignment of all HD-ZIP protein sequences retrieved from GenBank and those amplified through degenerate PCR and RACE. This alignment was adjusted manually to correct obvious errors generated by ClustalX and saved as a Nexus file for phylogenetic analysis. Additionally, this alignment was used to calculate the relative abundance of amino acids in the data matrix using MacClade 4.06 (Maddison and Maddison 2003; Fig. 2.2).

Aligned sequences were analyzed phylogenetically through both substitution-weighted and character-weighted parsimony analyses. Substitution weighting depended on a series of amino acid transition matrices (AATMs): PAM250, WAG, WAG*, Minimum-Evolution, BLOSUM45, BLOSUM 62, BLOSUM80 and BLOSUM100 using PAUP* (Swofford, 2001). Following a preliminary parsimony analysis in which all characters were equally weighted, I then implemented an analysis in which characters were differentially weighted based on the mean value of their rescaled consistency index also using PAUP* (Swofford, 2001). Maximum parsimony trees were generated through heuristic searches with 100 random taxon additions, TBR branch swapping and saving multiple parsimonious trees (MULTREES on). Gaps were encoded as missing data. Bootstrap support was estimated by performing 500 nonparametric bootstrap replicates with the above settings, except with five random taxa additions per replicate.

Following all equally weighted and differentially-weighted parsimony analyses, trees were saved with branch lengths. The likelihood of all trees resulting from parsimony analyses was calculated and a KH-test conducted using TreePuzzle 5.0. Shape parameters for the gamma distribution were estimated with eight rate categories and either the JTT or Dayhoff amino acid

substitution model, separately (Table 2.3 and 2.4, respectively). The AATM with the appropriate character-weighting or non character weighting method that produced the tree that conferred the highest likelihood on the data was used subsequently for calculation of per site sequence variation and amino acid transition frequencies (using MacClade 4.06, Maddison and Maddison, 2003; Figs. 2.3 and 2.4) and for molecular clock calibration (see below).

A series of Bayesian analyses were performed using MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) running 200,000 generations and allowing for among site rate heterogeneity and weighting with one of the following AATMs: blossom, dayhoff, or jones. Two runs, each with four Markov chains (one “hot” and three “cold”) were implemented for each AATM. MCMC results (trees) were sampled every 100 generations. The first 100,000 generations were discarded to account for a “burn in” period for the Markov chain. Trees sampled after 100,000 generations were used to compute a majority rule consensus tree and to assess posterior probabilities of resolved clades.

Molecular clock calibration

Based on the composition of the clades resolved in the above analyses, we determine three internal nodes for which minimum age estimates are already available: the most recent common ancestor (MRCA) of an all angiosperm clade, the MRCA of an all spermatophyte clade, and the MRCA of an all bryophyte clade (minimum ages provided in Schneider, *et al.*, 2004). Ten weighted-parsimony trees and ten Bayesian trees were selected for clock calibration. I used the Min-Evol AATM to calculate branch lengths in each of these trees. Branches connecting each of ten full-length sequence drawn from each resolved clade and spanning the breadth of HD-ZIP III were used to calculate the average branch length between them and each of the three

calibration points. By dividing the age of the node selected by the average branch length to the ten calibration terminals, we estimated a global average for the number of steps that occur per million years (Tables 2.5). This rate was subsequently used to estimate the minimum ages of clades and key events in the diversification of HD-ZIP III (Tables 2.6 and 2.7).

RESULTS

Final data matrix

Our final data matrix included 969 aligned characters (AA positions) from 72 sequences (65 “ingroup” and 7 “outgroup”) sampled from across land plant phylogeny (including a liverwort, two lycopods, four gymnosperms, and 22 angiosperms; Table 2.2). In total, forty new HD-ZIP class III genes were isolated and sequenced from twenty-four embryophytes. Of the ingroup sequences, 40 out of 65 (61%) were full length. Although all outgroup sequences (obtained from GenBank) were full length only the homeodomain and beta-leucine zipper regions were included in the alignment, as the remaining portions of the outgroup sequences were either not homologous or unalignable. Unequal frequencies of amino acid residues (Fig. 2.2), apparent variation in rates of evolution among sites in the alignment (Fig. 2.3), and unequal frequencies of inter-conversions among amino acids (Fig. 2.4) all suggest that suitable correction methods may be needed for accurate phylogeny reconstruction based this data matrix.

Parsimony analyses

Results based on parsimony analyses using any of the AATMs and equally weighted characters are generally concordant. HD-ZIP III is monophyletic with high bootstrap support in all optimal trees (Fig. 2.5). Sequences from all spermatophytes form a clade as do all sequences

from angiosperms, though with only modest bootstrap support. Non-seed plants form a clade in the most parsimonious trees recovered through analyses weighted with most of the AATMS, but are split into two successive branches at the base of the tree in results based on the PAM250 and BLOSSUM100 AATMs (not shown). Three clades are resolved within the angiosperms with moderate to very high bootstrap support. These are the COR, PHB, and REV clades. Of these, the PHB clade is the smallest with 16 sequences. The COR and REV clades each have 21 sequences. A sister relationship between the PHB and REV clades receives moderate support and the COR clade is sister to the PHB/REV composite clade.

Results based on analyses in which characters in the data matrix were re-weighted based on their rescaled-consistency indices were nearly identical to their counterparts in analyses in which the characters were equally weighted, but conferred slightly lower likelihoods on the data in KH tests based on either the JTT or Dayhoff weighting matrices (Tables 2.3 and 2.4). Though character-weighted trees were not significantly worse than those based on equally weighted characters, this complication of the analytic methods did not improve the fit of the results to the data. Accordingly, character-weighted trees were not used in subsequent analyses.

Bayesian analyses

The Markov chains in each of the Bayesian analyses appear to have reached stationarity within 10,000 to 20,000 generations, which is well within the allotted burn-in period of 100,000 generations (Fig. 2.6). A majority-rule consensus tree based on results from the Bayesian analysis using the BLOSSUM transition matrix appears to be largely concordant to results based on weighted-parsimony analyses, with two notable exceptions. As with the parsimony analyses, HD-ZIP III is monophyletic and the COR, PHB, and REV subclades are resolved, though clade

support metrics are generally low. However, the angiosperm taxa do not form a monophyletic group in a majority of Bayesian trees. Instead, the non-angiosperm sequences appear as sister to the COR clade, though with a low inferred posterior probability. Additionally, PHB and REV do not resolve as sister clades. Results based on Bayesian analyses using the JONES transition matrix are largely similar to those based on the BLOSSUM matrix, except that PHB and REV do resolve as sister clades (as in the weighted-parsimony analyses), but in a majority of these trees, the non-angiosperm branch is sister to the PHB/REV-composite clade, again with a low inferred posterior probability (not shown). Results based on the Dayhoff matrix were similar to those based on the BLOSSUM matrix, except that the REV clade was split into two groups (not shown).

Dating events in HD-ZIP III

Estimates for the minimum ages of the most recent common ancestors (MRCAs) for the embryophyte, spermatophyte-only, and angiosperm-only clades were taken from Schneider *et al.* (2004). Estimated rates for an HD-ZIP III molecular clock based on these three calibration points are variable, suggesting that the clock rate varies across the phylogeny (Table 2.5). The clock rates increase, progressively, moving from the oldest to the most recent calibration point. Accordingly, the rate based on the angiosperm-MRCA calibration point was taken as the most appropriate for event dating within the angiosperm-only clade.

Within the angiosperm-only clade three ancient gene duplication events are evident (Table 2.6; Fig. 2.8). The first is the duplication that produced the COR clade, on the one hand, and the PHB/REV-composite clade, on the other. This gene duplication event has a minimum estimated age of 252 MYA, indicating that it likely occurred after the divergence of the

angiosperms from other seed plants, but before the MRCA of the angiosperms. This duplication event accounts for seven detected duplicate-gene pairs in HD-ZIP III. The second gene duplication event produced the separate PHB and REV clades and has a minimum estimated age of 222 ± 4 MYA. This duplication likely occurred after the origin of flowering plants, but before the split of the monocots from the eudicots and core magnoliids (taxon names sensu Soltis *et al.* 2000; APG II 2003). Ten detected duplicate-gene pairs can be traced back to this duplication event. The third ancient gene duplication event occurred within the COR clade and has a minimum estimated age of 94 ± 2 MYA, but must have occurred before the MRCA of the core eudicots (sensu APG II 2003). This event accounts for four detected duplicate-gene pairs.

Eleven more recent gene duplication events each contributed duplicate gene pairs to only one surveyed species (Fig. 2.9). Two of these occurred in a monocot species (rice, REV clade) and have minimum estimated ages of ~ 37 and ~ 2 MYA. A third occurred in a core magnoliid species (*Peperomia*, REV clade) and has a minimum estimated age of ~ 4 MYA. Two duplications occurred in Asterid species (*Antirrhinum*, COR clade, and *Zinnia*, REV clade; ~ 5.4 and 27.4 MYA, respectively). Six duplicate gene pairs were detected in Rosid species (one in *Arabidopsis*, PHB clade, and five in *Populus*, all angiosperm clades), but the number of duplication events that gave rise to them is unclear. A maximum of six gene duplication events may have been involved in the production of these duplicate gene pairs, but as few as three could account for the data. The first duplication event occurred in the early evolutionary history of the eurosid II clade, to which *Arabidopsis* belongs, and has a minimum estimated age of ~ 80 MYA. The second occurred in the Eurosids I clade, to which *Populus* belongs, and has a minimum estimated age of ~ 20 MYA. Three additional duplicate gene pairs in the REV and COR clades were produced in the recent evolutionary history of *Populus* and date back to at least ~ 8 to ~ 11

MYA. These duplicate gene pairs may have arisen from a single chromosomal duplication event affecting multiple loci.

DISCUSSION

Choosing the most accurate topology for HD-ZIP III

I regard the results of the parsimony analyses that used any of six AATMs to be more accurate than those based on Bayesian analyses for several reasons. The amount of disagreement among relationships close to the terminals in each tree is highly concordant (Figs. 2.5 and 2.7). Differences between parsimony and Bayesian results are more evident closer to the root of the tree. Here parsimony results are consistent with the organismal phylogeny for the taxa surveyed, but Bayesian results are at odds with the organismal phylogeny. If the Bayesian results were correct, it would be necessary to infer six independent events of non-random gene silencing along the branches leading from the root of the organismal tree for embryophytes to the terminals corresponding to the non-angiosperm species surveyed. For example, if a non-angiosperm clade really were the sister group to the COR clade, as is suggested in the Bayesian results (Fig. 2.7), then the molecular common ancestor of this composite clade would have to be present at the root of the organismal phylogeny for the embryophytes. Additionally, the origin of the REV and PHB clades would have to antedate the MRCA of the COR/non-angiosperm. Along three distinct branches in the organismal phylogeny, those leading to the liverworts, lycopods, and gymnosperms, independent gene loss (silencing) events would have to have occurred for both REV and PHB orthologs, while REV and PHB orthologs were each retained along the succession of branches leading to the common ancestor of the angiosperms. It is much simpler, and more likely, account for the Bayesian results as a case of long-branch attraction (LBA).

Weighted-parsimony analyses of the HD-ZIP III data matrix appear to be less sensitive to the apparent artifact described above, presumably because parsimony minimizes inferred branch lengths. Although model-based methods are generally thought to be more accurate in phylogeny reconstruction than those based on parsimony, recent analyses have shown that parsimony-based methods sometimes recover an accurate phylogeny for a gene family in cases where likelihood-based methods do not (Kolaczkowski and Thornton 2004; Thornton and Kolaczkowski; but see Steel 2005 for an opposing view). Parsimony outperforms likelihood-based methods in cases where heterotachy (unequal rates of molecular evolution in duplicate gene pairs) is present in the evolutionary history of a gene family. In a theoretical study, Kolaczkowski and Thornton (2004) demonstrated that likelihood-based methods will absolutely converge on the incorrect topology when confronted with some forms of heterotachy. In a recent empirical study, Fares *et al.* (2005) discovered multiple cases of LBA in likelihood-based phylogenetic analyses of gene families that contain duplicated gene pairs derived from a whole-genome duplication event in the evolutionary history of yeast. They argued that these cases of LBA were generally associated with positive selection acting on or both members of the gene pair after duplication. As the HD-ZIP III gene family has many duplicate gene pairs, it is possible that cases of heterotachy, and perhaps also positive selection, are partially confounding phylogenetic results based on Bayesian methods.

Coupled with the implausibility of six independent and non-random cases of gene loss in the early history of the gene family, I am inclined to regard the phylogenetic results based on parsimony as more accurate. As the KH-tests based on two different AATMs (Tables 2.3 and 2.4) each indicate that the most parsimonious trees recovered through the BLOSSUM45-

weighted analyses had the highest likelihood, I consider this the best available estimate of the phylogeny for HD-ZIP III and have used this topology in molecular dating analyses.

Duplication events

The HD-ZIP III gene family has numerous gene pairs that are derived from either ancient or recent gene duplication events. In total 31 duplicate gene pairs (30 in angiosperms, and one in *Pinus*) were detected through our analyses, involving nearly all of the genes isolated from surveyed species (95%). Had no gene duplication events occurred, the size of the gene family would be less than half its current size (a maximum of 29 genes). Our results reinforce a long-standing view that gene duplication is the major force driving the expansion of gene families (Ohno, 1970). Additionally, gene duplication events provide the raw material for functional diversification of genes, as purifying selection may be relaxed on one or both daughter genes (Zhang, 2003).

Some of the gene duplication events detected here may be attributable to chromosomal duplication events in the evolutionary history of angiosperms. Bowers *et al.* (2003) suggested that three distinct chromosomal duplication events in the evolutionary history of flowering plants contributed to many of the duplicate gene pairs found in the Arabidopsis genome. The events are α , β , and γ . The α chromosomal event occurred in the early evolutionary history of the eurosid II lineage, after its split from the eurosid I lineage, between 14.5 and 86 MYA. The β event occurred after the divergence of the monocots from the core magnoliids and eudicots, but prior to the MRCA of the core eudicots, between 112 and 235 MYA. The γ event, which is the least well supported by their data, occurred between 170 and 334 MYA, somewhere between the MRCA of the seed plants and MRCA of the monocots and eudicots.

In total, four gene duplication events detected in the present study may have been the result of the chromosomal duplication events identified by Bowers *et al.* (2003). The timing of two of the ancient gene duplication events detected in this study is consistent with the timing of the γ event. These are the gene duplication events that produced the MRCAs of the COR and PHB/REV composite clade and the duplication event that produced the MRCAs of the separate PHB and REV clades (Fig. 2.8). However, since these events occurred at distinct points in time, at most only one could be attributed to the γ event. The apparent timing of the third ancient duplication event, in the COR clade, is consistent with the β event (Fig. 2.8). Although the inferred minimum age of this event falls after of the range of dates for the β event, the membership of the clades derived from this duplication event (including asterid, rosid, and caryophyllid species) indicates that it occurred prior to the radiation of the core eudicots, which Schneider *et al.* (2004) date at approximately 179 MYA. This places the ancient COR duplication at the early end of the range of potential dates for the β event. Finally, one of the more recent duplication events that occurred in Arabidopsis PHB clade lead to the PHABULOSA and PHAVOLTA genes. The minimum estimated age for this duplication event is approximately 80 MYA (Fig. 2.9), which falls at the early end of the α event, somewhere in the early history of the eurosid II clade. All other recent duplication events can be excluded from the α event, as they occur in lineages outside the eurosid II clade.

CONCLUSION

Resolving the interplay between events in the natural history of gene families and the morphological diversification of organisms is an exciting area of research in contemporary biology. To be successful in this endeavor, biologists will need to gather and understand results

based on several disparate lines of evidence. Better understanding is needed of the molecular-genetic pathways that govern the development of model species, the phylogenetic history of the evolutionary relatives of model research organisms, and the evolutionary histories of the genes that mutated to create morphological variation for selection to act upon. The present study provides insights into the last category, but is aimed at building bridges between the study of molecular ancestry and the study of organismal diversification. The timing of events in the natural history of the HD-ZIP III lineage provide a foundation for investigation of gene function in model research organisms. For example, numerous gene duplicate pairs are identified here, providing potential insights for gene knock-out/knock-down studies in species surveyed here. Additionally, this study provides a potential basis for comparative studies of gene function among model research species. The potential roles of neofunctionalization or subfunctionalization of promoters and encoded proteins in the HD-ZIP III family requires an accurate understanding of its ancestry. Finally, efforts at recovering organismal phylogenies are sometimes hampered by the lack of suitable molecular tools to use for molecular systematic studies. The phylogenetic context provided here shows that some lineages within HD-ZIP III are likely poor sources of markers for molecular systematic studies. For example, the REV clade is replete with gene duplications and paralogy-related problems would likely hamper efforts to reconstruct organismal phylogenies with REV homologs. In contrast, the PHB clade has relatively few detected duplication events and is much more likely to serve as an accurate marker for reconstructing organismal phylogenies. Accordingly, reconstructing the natural history of gene families facilitates progress in several sib-disciplines of biology centered around the study of developmental evolution.

REFERENCES

- Angiosperm Phylogeny Group. 2003. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.* **141**: 399-436.
- Aso, K., Kato, M., Banks, J. A., Hasebe, M. 1999. Characterization of homeodomain leucine zipper genes in the fern, *Ceratopteris richardii* and the evolution of the homeodomain-leucine zipper gene family in vascular plants. *Mol. Biol. Evol.* **16**: 544-551.
- Bharathan, G., Janssen, B., Kellogg, E. A. and Sinha, N. 1997. Did homeodomain Proteins duplicate before the origin of angiosperms, fungi, and metazoa? *Proc. Natl. Acad. Sci. USA.* **94**: 13749-13753.
- Bowers, J. E., Chapman, B. A., Rong, J., and Paterson, A. H. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature.* **422**: 433-438.
- Bowman, J. L. 2003. Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes. *Curr Biol.* **13**: 1768–1774
- Chan, R. L., Gago, G. M., Palena, C. M., and Gonzalez, D. H. 1998. Homeoboxes in plant development. *Biochem. Biophys. Acta.* **1442**: 1-19.
- Bowman, J. L., Eshed, Y., and Baum, S. 2002. Establishment of polarity in angiosperm lateral organs. *Trends Genet.* **18**: 134-141.
- Bowman, J. L. 2003. Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes. *Curr Biol.* **13**: 1768–1774
- Eshed, Y., Izhaki, A., Baum, S. F., Floyd, S. K., Emery, J. F. and Bowman, J. L. 2004.

- Asymmetric leaf development and blade expansion in *Arabidopsis* are mediated by KANADI and YABBY activities. *Development*. **131**:2997–3006
- Fares, M. A., Bryne, K. P., and Wolfe, K. H. 2005. Rate asymmetry after genome duplication causes substantial long-branch attraction artifact in the phylogeny of *Saccharomyces* species. *Mol. Biol. Evol.* **23**: 245-253.
- Floyd, S. K. and Bowman, J. L. 2004. Gene regulation: ancient microRNA target sequences in plants. *Nature*. **428**: 485-486.
- Golz, J. F., Keck, E. J., and Hudson, A. 2002. Spontaneous mutations in KNOX genes give rise to a novel floral structure in *Antirrhinum*. *Curr. Biol.* **12**: 515-522.
- Kim, M., Pham, T., Hamidi, A., McCormick, S., Kuzoff, R. and Sinha, N. R. 2003. Reduced leaf complexity in tomato wiry mutants suggests a role for *PHAN* and *KNOX* genes in generating compound leaves. *Development*. **130**: 4405-4415.
- Kang, J. and Dengler, N. 2002. Cell cycling frequency and expression of the homeobox gene *AtHB-8* during leaf vein development in *Arabidopsis*. *Planta* **216**: 212-219.
- Kolaczowski, B. and Thornton, J. W. 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature*. **431**: 980-984.
- Maddison, W. P. and Madson, D. R. 2003. "MacClade. Version 4.06."
- McConnell, J. R., Barton, M. K. 1998. Leaf polarity and meristem formation in *Arabidopsis*. *Development*. **125**: 2935–2942
- McConnell, J. R., Emery, J. F., Eshed, Y., Bao, N., Bowman, J. and Barton, M. K. 2001. Role of *PHABULOSA* and *PHAVOLUTA* in determining radial patterning in shoots. *Nature*. **411**: 709–713
- Ohashi-Ito, K. and Fukuda, H. 2003. HD-Zip III homeobox genes that include a novel member, *ZeHB-13* (*Zinnia*)/*ATHB-15* (*Arabidopsis*), are involved in procambium

- and xylem cell differentiation. *Plant Cell Physiol.* **44**:1350–1358
- Ohno, S. 1970. *Evolution by gene duplication*. New York: Springer-Verlag
- Otsuga, D., Deguzman, B., Prigge, M., Drews, G., and Clark, S. 2001. REVOLUTA regulates meristem initiation at lateral positions. *Plant J.* **25**: 223–236.
- Ponting, C.P., and Aravind, L. 1999. START: A lipid-binding domain in StAR, HD-ZIP and signalling proteins. *Trends Biochem. Sci.* **24**:130–132.
- Prigge, M. J., Otsuga, D., Alonso, J. M., Ecker J. R., Drews, G. N., Clark, S. E. 2005. Class III homeodomain-leucine zipper gene family members have overlapping, Antagonistic, and distinct roles in Arabidopsis development. *Plant Cell.* **17**: 61-76.
- Pryer, K. M., Schneider, H., Smith, A. R., Cranfill, R., Wolf, P. G., Hunt, J. S., Sipes, S. D. 2001. Horsetail and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* **409**: 618-622.
- Reiser, L., Sánchez-Baracaldo, P. and Hake, S. 2000. Knots in the family tree: Evolutionary relationships and functions of *knox* homeobox genes. *Plant Mol. Biol.* **42**: 151-166.
- Ronquist, F. and Huelsenbeck, J. P. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* **19**:1572-1574.
- Sakakibara, K., Nishiyama, T., Kato, M. and Hasebe, M. 2001. Isolation of Homeodomain Leucine Zipper Genes from the Moss *Physcomitrella patens* and the Evolution of Homeodomain-Leucine Zipper Genes in Land Plants. *Mol. Biol. and Evol.* **18**: 491-502.
- Schneider, H. Schuettpelz, E., Pryer, K. M., Cranfill, R., Magallon, S., and Lupia, R. 2004. Ferns diversified in the shadow of angiosperms. *Nature* **428**: 553-557.
- Schrick, K., Nguyen, D., Karlowski, W. M. and Mayer, K. F. 2004. START lipid/sterol

- binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors. *Genome Biology* **5**: R41.
- Soltis, D. E., *et al.* (15 co-authors). 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Bot J. Linn. Soc.* **133**: 381-461.
- Steel, M. 2005. Should phylogenetic models be trying to 'fit an elephant'? *Trends in Genetics* **21**: 307-309.
- Swofford, D. L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sunderland, MA, Sinauer.
- Thompson, J. D., Gibson, T. J., *et al.* 1997. "The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." *Nucleic Acids Research*. **25**: 4876-82.
- Thornton, J. W. and Kolaczowski, B. 2005. No magic pill for phylogenetic error. *Trends in Genetics* **21**: 310-311.
- Zhong, R. and Ye, Z.H. 2001. Alteration of auxin polar transport in the *Arabidopsis* ifl1 mutants. *Plant Physiol.* **126**: 549–563.
- Zhong, R. and Ye, Z.H. 1999. IFL1, a gene regulating interfascicular fiber differentiation in *Arabidopsis*, encodes a homeodomain-leucine zipper protein. *Plant Cell* **11**: 2139–2152.
- Zhang, J. 2003. Evolution by gene duplication: an update. *Trends in Ecology and Evolution*. **18**: 292

Table 2.1 Degenerate primers used to amplify initial fragments of HD-ZIP, class III genes.

Primer	Direction	5' to 3' Sequencing
HDZIII 5'A	Forward	ARAAYMGMGUTGYMGIGARAARCA
HDZIII_MF	Forward	GARAAYGAIMGIYTICARAARCAR
HDZIII_IF	Forward	ACIGCNGTIGANTGGGTNCARATG
PHB290_5F	Forward	GGIATHGTNGCIRTISIMGIAAY
PHB430_5F_N	Forward	YATIGTIGAYCAYRTIGAYYTIGAYG
REV220_5F	Forward	ACICCARCAIWSIYTNMGIGAY
REV280_5F_N	Forward	CNGAYWSIRTIGGIATITTYGC
ATHB8_280_5F	Forward	GGIATIRTNGCIATHWSICAYGG
ATHB8_430_5F_N	Forward	RTIGTIGAYCAYWTIRAYYTIGAR
HDZIII3'A	Reverse	ACNARIGTIGTYTCIARCATRTC
HDZIII_OR	Reverse	RTCIGCCCAATCISWICKRTGYTC
HDZIII_MR	Reverse	ARIARCATISWIGCYTTNGCRCA
PHB680_3R_N	Reverse	CCISWRCAIARYTGIARIARRTACAT
PHB910_3R	Reverse	CRAAIGTRAAIACIGSISWIGCRTT
REV650_3R	Reverse	GYTCDATIGTITGNCCIARIGGCAT
REV630_3R_N	Reverse	AYYTGISWICCNNGTRAAICKIGT
ATHB8_960_3R	Reverse	RAAICCYTGYTGCATDATYTGNGG
ATHB8_700_3R_N	Reverse	SWIGCRTCIATIGGNGCRAADAT

Table 2.2. Gene names, source taxa, and GenBank accession numbers for HD-Zip genes analyzed in this study.

Gene Name	Taxon	GenBank Accession No.
NicoPHB	Nicotiana tabacum	AAS66760
PopuPHB1	Populus trichocarpa	AAX19052
RicePHB	Oryza sativa	AK103284
PopuPHB2	Populus trichocarpa	AAX19053
PHV_Arab	Arabidopsis thaliana	NP_174337
PHB_Arab	Arabidopsis thaliana	NP_181018
SnapPHB	Antirrhinum major	This study
CucPHB	Cucumis sativas	This study
MedPHB	Medicago sativa	This study
LiriPHB	Liriodendron tulipifera	This study
SorgPHB	<i>Sorghum</i> sp.	This study
TritPHB	Triticum sp.	This study
OcauPHB	<i>Ornithogalum</i> sp.	This study
AfimPHB	Aristolochia fimbriata	This study
TomPHB	Lycopersicon esculentum	This study
CucREV	Cucumis sativas	This study
PopuREV1	Populus trichodarpa	AAX19051
PopuREV2	Populus trichodarpa	AAX19050
PopuREV3	Populus trichodarpa	AAS77254
REV_Arab	Arabidopsis thaliana	AAF42938
ZinniaREV1	Zinnia elegans	BAC22514
ZinniaREV2	Zinnia elegans	BAC22513
RiceREV1	Oryza sativa	AK102830
RiceREV2	Oryza sativa	AK102603
RiceREV3	Oryza sativa	AK102378
ZeaREV	Zea may	AAR97952
PvelREV1	Peperomia velvatina	This study
PvelREV2	Peperomia velvatina	This study
PhisREV	Piper hispida	This study
TritREV	Triticum sp.	This study
SorgREV	Sorghum sp.	This study
SnapREV	Antirrhinum major	This study
MedREV	Medicago stiva	This study
TomREV	Lycopersicon esculentum	This study
CalyREV	Calycanthus floridus	This study
MedCOR	Medicago sativa	This study
CucCOR	Cucumis sativas	This study
PopuCOR1	Populus trichocarpa	AAX19057
PopuCOR2	Populus trichocarpa	AAX19056
PopuCOR3	Populus trichocarpa	AAX19054
PopuCOR4	Populus trichocarpa	AAX19055
ArabCOR-Like	Arabidopsis thaliana	NP_195014
ZinniaCOR1	Zinnia elegans	CAC84276

Gene Name	Taxon	GenBank Accession No.
ArabCOR	<i>Arabidopsis thaliana</i>	AAW88440
ZinniaCOR2	<i>Zinnia elegans</i>	BAC22512
TomCOR	<i>Lycopersicon esculentum</i>	This study
SnapCOR1	<i>Antirrhinum major</i>	This study
SnapCOR2	<i>Antirrhinum major</i>	This study
PhysCOR	<i>Physotegia</i> sp.	This study
RivinaCOR1	<i>Rivina humilis</i>	This study
RivinaCOR2	<i>Rivina humilis</i>	This study
ChloCOR	<i>Chloranthus spicatus</i>	This study
LiriCOR	<i>Liriodendron tulipifera</i>	This study
CalyCOR	<i>Calycanthus fliridus</i>	This study
AfimCOR	<i>Aristolochia fimbriata</i>	This study
AgigCOR	<i>Aristolochia gigantea</i>	This study
EpheHDZ	<i>Ephedra distachya</i>	This study
PinusHDZ1	<i>Pinus taeda</i>	This study
GnetHDZ	<i>Gnetum gnemon</i>	This study
PinusHDZ2	<i>Pinus teada</i>	DR384908
SelaHDZ	<i>Selaginella erythopus</i>	This study
LycOHDZ	<i>Lycopodium esculentum</i>	This study
MarcHDZ	<i>Marchantia</i> sp.	This study
PhahoxIV	<i>Phalaenopsis</i> sp.	S71477
AHDIVArab	<i>Arabidopsis thaliana</i>	AAB41901
GmhIIGly	<i>Glycine max</i>	T06438
PpHB4II	<i>Physcomitrella patens</i>	AB028075
CrateroII	<i>Craterostigma plantagineum</i>	AJ005833
T2PIArab	<i>Arabidopsis thaliana</i>	AAP88361
HAHBI_He	<i>Helianthus annus</i>	T12634

Note. Sequences are organized by the clade they belong to.

Table 2.3 Ln-Likelihoods (LogL) for most parsimonious trees calculated using the JTT amino acid substitution matrix in TREEPUZZLE 5.0. Topologies tested include branch lengths.

Matrix/Method	Tree#	LogL	Diff- erence	Sign. Worse
EqualBL45	1	-26681.36	0	Best
	2	-26681.36	0	Best
WeightedBL45	3	-26700.5	19.14	no
EqualBL62	4	-26690.05	8.7	no
WeighredBL62		-26685.83	4.47	no
EqualBL80	6	-26685.75	4.4	no
WeightedBL80	7	-26689.33	7.97	no
EqualBL100	8	-26682.1	0.74	no
	9	-26682.1	0.74	no
WeightedBL100	10	-26681.96	0.6	no
EqualPAM250	11	-26689.23	7.87	no
Weighed PAM250	12	-26690.49	9.13	no
EqualWAG	13	-26708.66	27.3	no
WeightedWAG	14	-26700.76	19.4	no
EqualWAG*	15	-26700.35	18.99	no
Weighted WAG*	16	-26694.73	13.37	no
Equal MiniEvol	17	-26696.72	15.37	no
	18	-26687.54	6.18	no
	19	-26692.72	11.36	no
Weighted MiniEvol	20	-26705.03	23.67	no

Table 2.4 Ln-Likelihoods (LogL) for most parsimonious trees calculated using the Dayhoff amino acid substitution matrix in TREEPUZZLE 5.0. Topologies tested include branch lengths

Matrix/Method	Tree #	LogL	Diff erence	Sign. worse
EqualBL45	1	-27029.46	0	Best
	2	-27029.46	0	Best
WeightedBL45	3	-27048.38	18.92	no
EqualBL62	4	-27038.52	9.06	no
WeighredBL62	5	-27034.88	5.41	no
EqualBL80	6	-27032.74	3.28	no
WeightedBL80	7	-27038.07	8.61	no
EqualBL100	8	-27030.83	1.37	no
	9	-27030.83	1.37	no
WeightedBL100	10	-27031.32	1.85	no
EqualPAM250	11	-27039.74	10.28	no
Weighed PAM250	12	-27038.02	8.55	no
EqualWAG	13	-27062.64	33.18	no
WeightedWAG	14	-27057.31	27.84	no
EqualWAG*	15	-27053.82	24.36	no
Weighted WAG*	16	-27050	20.54	no
Equal MiniEvol	17	-27048.99	19.52	no
	18	-27039.73	10.27	no
	19	-27045.13	15.67	no
Weighted MiniEvol	20	-27057.83	28.36	no

Table **2.5**. Calibration of the molecular clock using a pool of trees drawn from both weighted-parsimony and Bayesian analyses.

	Clade1	Clade2	Clade3
*Tree # (Angiosperm) (Spermatophyte) (Embryophyte)			
1	421.6	423.8	389.4
2	422	424	389.5
3	423.3	424.4	390.2
4	422.7	423.8	389.15
5	429.8	427.4	390.9
6	412.8	421.4	304.9
7	416.5	421.8	388.9
8	421	426	306
9	422.5	426.3	306.1
10	420.7	426.6	306.2
11	408.7	417.8	386.1
12	423.3	432.2	390.35
13	434.8	430.9	392.7
14	439.3	436.7	393.3
15	433	433.5	369.25
16	416.3	422.7	387.6
17	425.2	427.1	390.05
18	423.5	426.3	392.2
19	417.5	428	374.5
20	434.3	432.9	392
Total	8468.8	8533.6	7429.3
Mean	423.44	426.68	371.465
Error	7.67349505	4.65907151	34.1906847
Max	431.113495	431.339072	405.655685
Min	415.766505	422.020928	337.274315
*Calibration Date (Relaxed) 252(MYA)	334(MYA)	432(MYA)	
MYA/Step (MinEvol Rate) Mean	0.59512564	0.78278804	1.16296286
MYA/Step (MinEvol) Error	0.01078829	0.00854856	0.10795698
MYA/Step (MinEvol Rate) Min	0.58453285	0.77433282	1.06494255
MYA/Step (MinEvol Rate) Max	0.60610943	0.79142994	1.2808565

*Tree1-Tree10 are from a pool of trees generated by non Character weighted parsimony analysis using BLOSSUM45 (2 trees), BLOSSUM62 (1 tree), BLOSSUM80 (1 tree), BLOSSUM100 (2 trees), MinimumEvolution (3 trees) and PAM250 (1 tree) amino acid transition matrices.

*Tree11-Tree21 are from a pool of trees generated by Bayesian analysis using BLOSSUM (1 tree), VT (4 trees) and Dayhoff (5 trees).

* Based on the clade composition within the phylogeny tree of HD-ZIP, class III in this study, we use the estimation dates from Pryer group (Pryer, *et al.*, 2001).

Table 2.6. Molecular clock calibration and dating of ancient duplication events within the angiosperm-only clade of the HD-ZIP, class III gene family.

	Trees PHB (A1)	PB/RV (A2)	REV (A3)	Cor_Dp (A4)	Cor_All (A5)
WtPar_Tree1	288	377	214	164.75	182.75
WtPar_Tree2	289.25	365.62	193	160	178
WtPar_Tree3	347	374.5	192	167.75	192.75
WtPar_Tree4	347	374.5	192	158.5	185.5
WtPar_Tree5	292.25	375	206.75	159.5	183.5
Bayes_Tree6	346	376.5	196	149	173
Bayes_Tree7	283.75	371.38	208	155.25	183.25
Bayes_Tree8	292.25	366.62	197	140.5	161.5
Bayes_Tree9	290	369.5	196	165.75	191.75
Bayes_Tree10	342.25	381.5	212.75	154.5	187.5
Average	311.775	373.212	200.75	157.55	181.95
MYA/Step Mean	0.595126	0.595126	0.59513	0.595126	0.595126
MYA/Step Error	0.010788	0.010788	0.01079	0.010788	0.010788
Date	185.5454	222.1082	119.472	93.7621013	108.283176
Error	3.363429	4.026211	2.16569	1.6996494	1.9628766

* PB/RV (A2) represents the most recent common ancestor (MRCA) of PHB and REV clade;

PHB (A1) represents the MRCA of PHB clade;

REV (A3) represents the MRCA of REV clade;

Cor_Dp (A4) represents the MRCA that leads to the duplication event within CORONA clade.

Cor_All (A5) represents the MRCA of CORONA clade.

Table 2.7. Molecular clock calibration and dating of recent duplication events within the angiosperm-only clade of the HD-ZIP, class III gene family.

	R1	R2	R3	R4	R5	R6	R7	R8
WtPar_Tree1	132	22	31.67	13	46	3	13	18
WtPar_Tree2	146.5	22	47.5	13	46	3	13	18
WtPar_Tree3	135	22	31.67	13	46	3	13	18
WtPar_Tree4	135	22	31.67	13	46	3	13	18
WtPar_Tree5	133	22	31.67	13	46	3	13	18
Bayes_Tree6	132	22	31.67	13	46	3	13	18
Bayes_Tree7	133.5	22	31.67	13	46	3	13	18
Bayes_Tree8	133.5	22	31.67	13	46	3	13	18
Bayes_Tree9	133	22	31.67	13	46	3	13	18
Bayes_Tree10	132	22	31.67	13	46	3	13	18
Average	134.6	22	33.25	13	46	3	13	18
MYR/Step Mean	0.595	0.6	0.595	0.6	0.6	0.6	0.6	0.6
MYR/Step Err	0.011	0.01	0.011	0.01	0.01	0.01	0.01	0.01
Date	80.07	13.1	19.79	7.74	27.4	1.79	7.74	10.7
Error	1.452	0.24	0.359	0.14	0.5	0.03	0.14	0.19

R1: Recent duplication event leading to *PHB* and *PHV* in *Arabidopsis*.

R2: Recent duplication event leading to *PHB* duplication in *Populus*.

R3 and R4: Recent duplication events leading to *REV1*, *REV2* and *REV3* in *Populus*;

R5: Recent duplication event leading to *REV1* and *REV2* in *Zinnia*;

R6: Recent duplication event leading to *REV1* and *REV2* in *Oryza*;

R7 and R8: Recent duplication events leading to *CORNA1*, *CORNA2*, *CORNA3* and *CORNA4*.

Figure **2.1**. Relative locations of forward and reverse degenerate primers used in amplification of HDZ-III homologs from diverse land plant species. The approximate locations of three conserved functional domains, including the homeodomain (red), the leucine-zipper domain (blue), and the START domain (green), are also indicated relative to their locations in the full alignment.

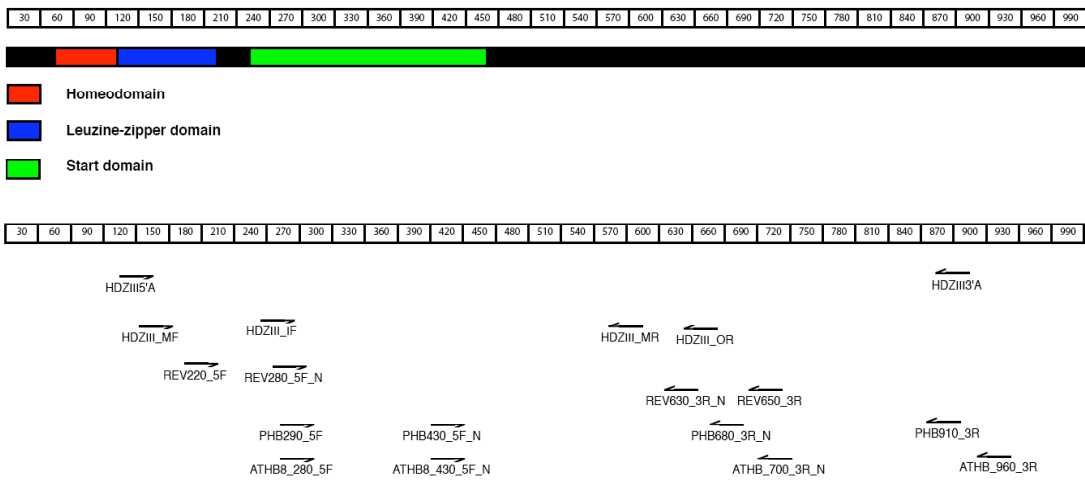


Figure **2.2**. Relative abundance of each amino acid in the full alignment of all the sequences analyzed in this study. Especially abundant amino acids include Ala, Glu, Leu, Ser, and Val. Relatively uncommon amino acids include Cys, His, Trp, and Tyr. Leu is more than seven times more abundant than Trp. This indicates the potential need for a model of sequence evolution that does not assume equal prior probabilities for residues.

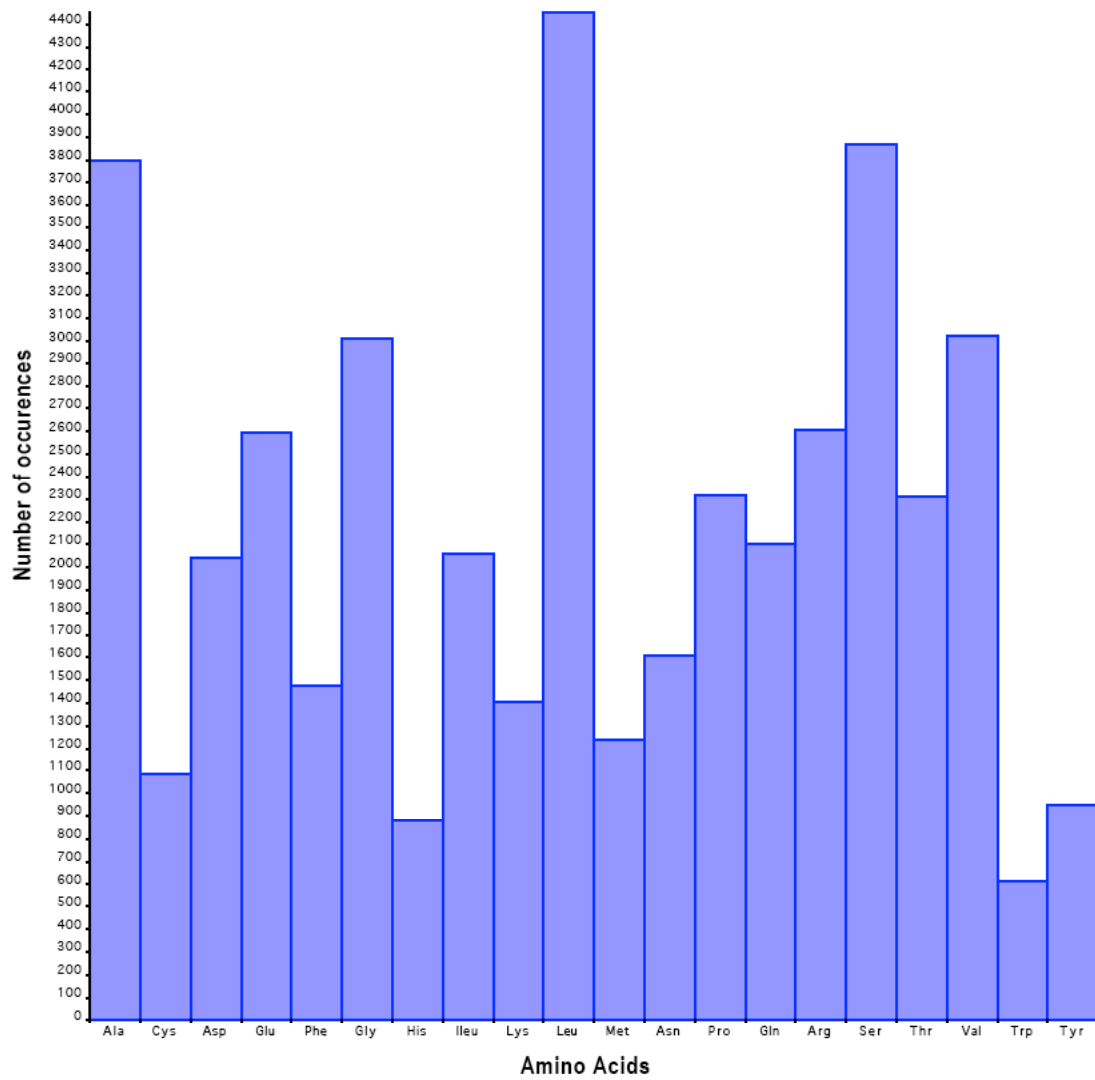


Figure **2.3**. Relative sequence variation across full alignment. Variation in residue substitution rates among 72 protein sequences are analyzed using a window size of five consecutive bases and graphed using MacClade 4.08 (Madison and Madison, 2003). Substitution rates are highly variable among sites, indicating the potential need for a model of sequence evolution that accounts for among site rate variation.

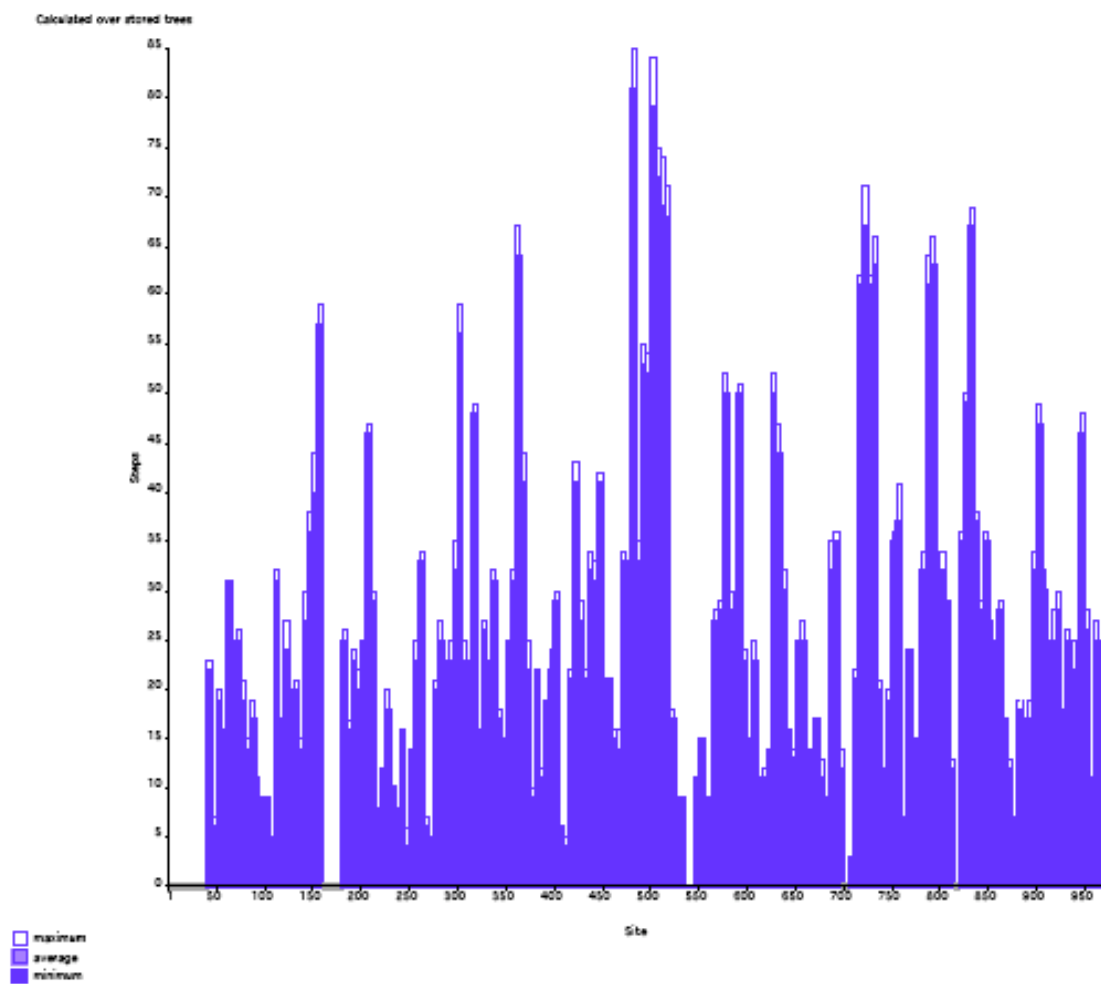


Figure 2.4. Frequency of unambiguous changes among amino acids. Amino acids at a site prior to change are listed on the left. Amino acids after change are listed across the top. The relative abundance of individual amino acid substitutions are proportional to the radius of each circle. Some substitutions are quite common (e.g., A→S, E→D, I→V, S→T, V→I, etc.), while others are quite rare (e.g., C→D, C→E, C→M, C→P, C→Q, etc.). This indicates the potential need for a model of sequence evolution that does not assume that all transitions among amino acids are equally likely.

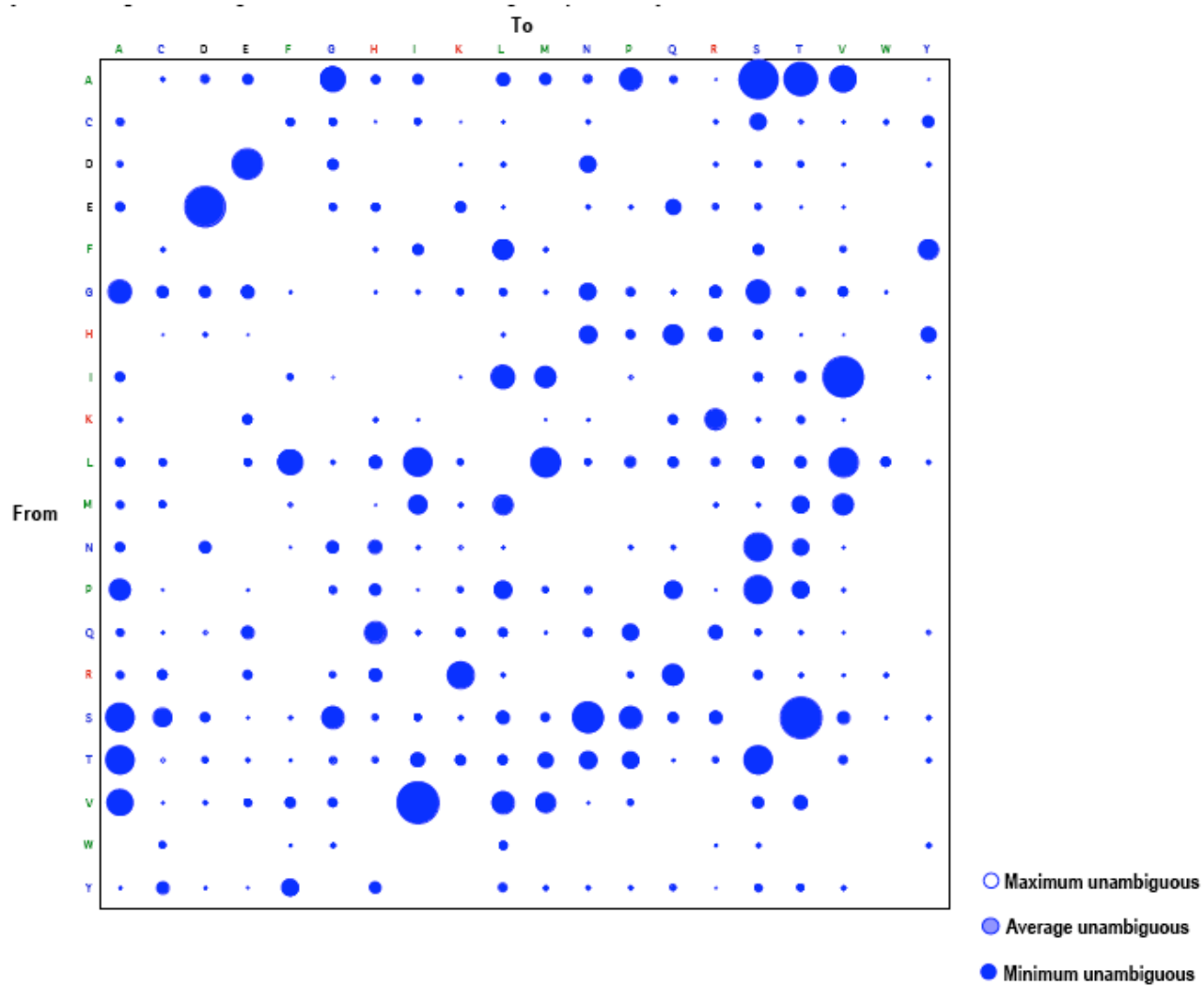


Figure **2.5**. One of two optimal weighted parsimony trees ($-\text{LnL}=-26681.36$). This tree resulted from a parsimony analysis that was weighted with the BLOSSUM 45 amino acid transition matrix (AATM). This tree indicates that class III of the HD-Zip gene family is monophyletic. Non-angiosperm sequences for two successive early branching lineages are near the base of the tree. Angiosperm sequences for a large, but poorly supported clade comprising three sub-clades; COR, REV, and PHB. These three subclades are moderately to well supported. Within the angiosperms, PHB and REV clades are sister to one another. Result based on weighted parsimony analyses using the BLOSSUM 62, BLOSSUM 80, BLOSSUM 100, PAM250, and Minimum Evolution AATMs are nearly identical, except that *Maracantia* is the first diverging lineage at the base of the tree in results based on BLOSSUM 100 and PAM250.

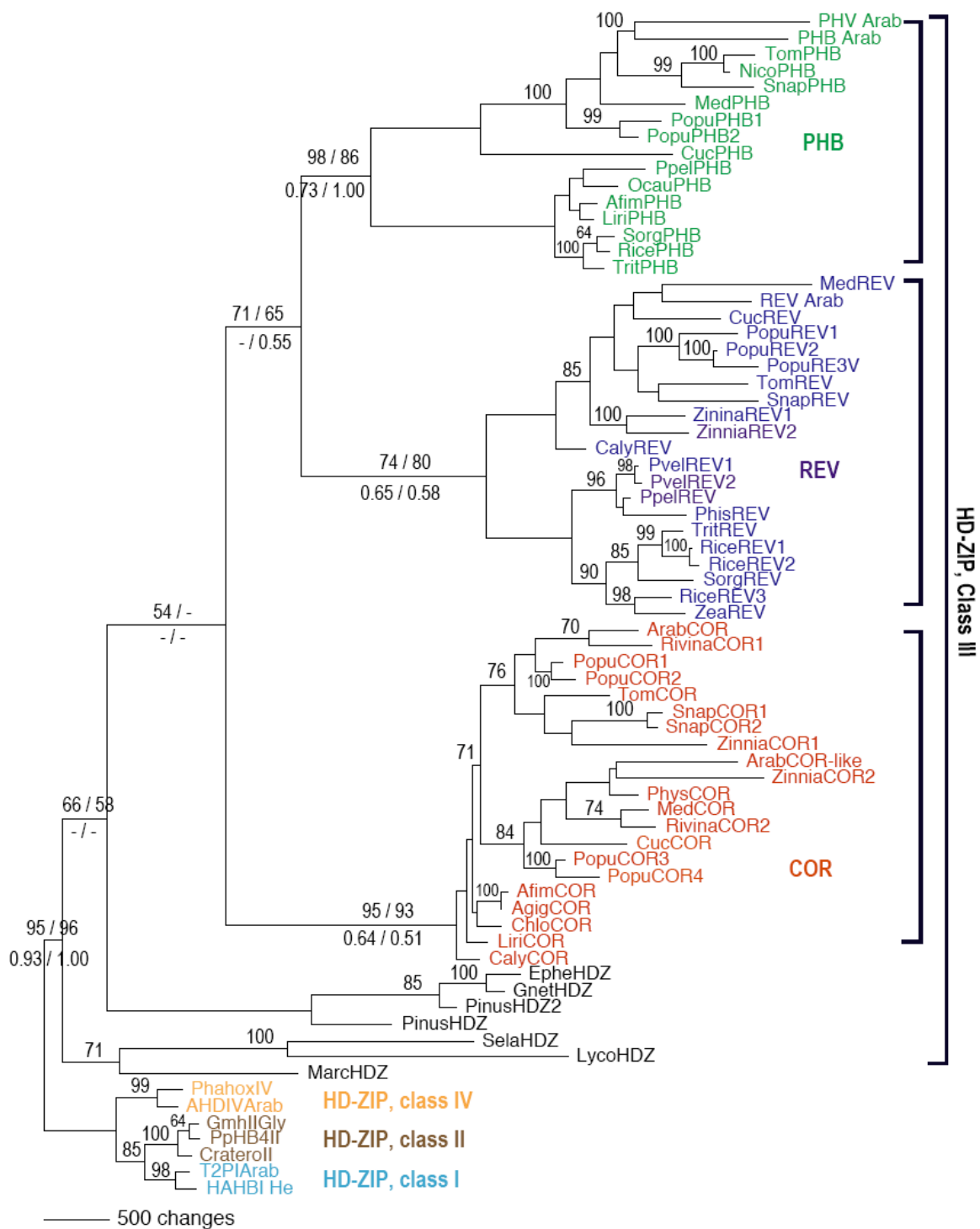


Figure **2.6**. Plots indicating the “burn-in” period for McMC analyses using three amino acid transition matrices: a) averaged BLOSSUM AATM; b) JONES AATM, and c) Dayhoff AATM. For each plot the generation number is plotted against the Ln-Likelihood for each tree. After approximately 25K generations, plots stabilize and little change is visible in subsequent generations of the random walk, indicating that the Markov chain has been burned in and parameter values in “posterior probability space” are being adequately sampled.

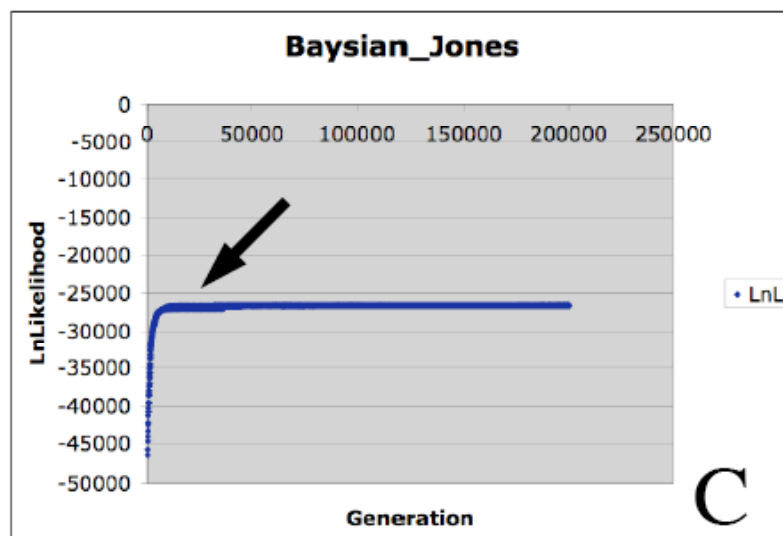
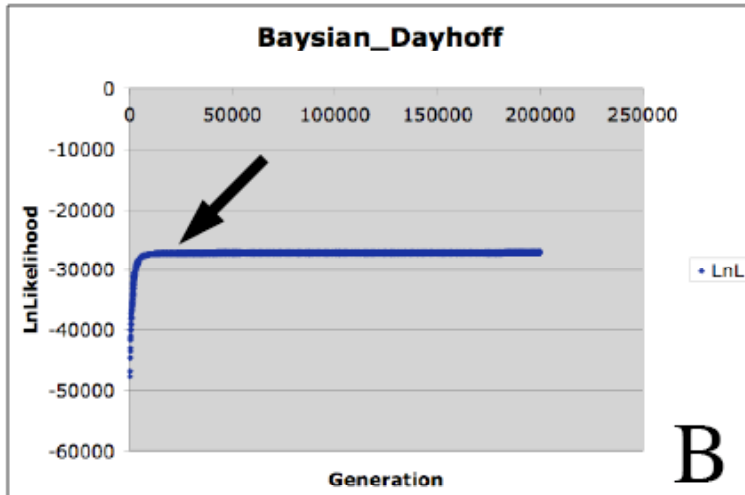
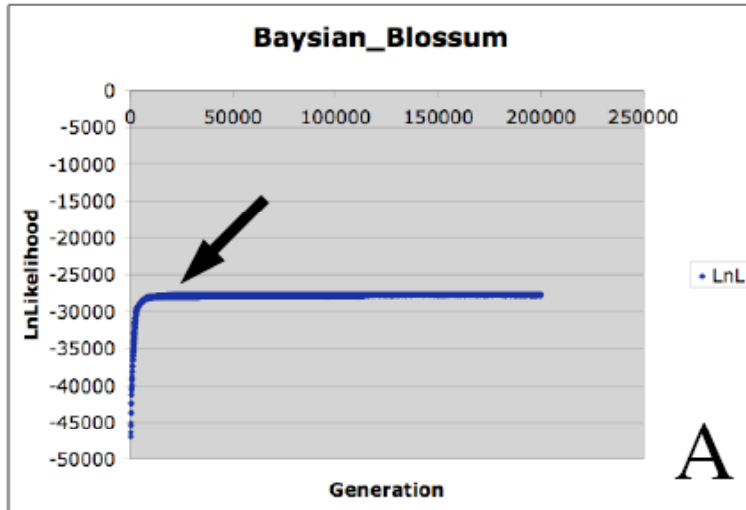


Figure 2.7. Majority-rule consensus tree based on 1,000 trees generated from a Bayesian analysis using the BLOSSUM AATM. As with the weighted parsimony analyses, class III of the HD-Zip gene family is monophyletic and PHB, REV, and COR clades are all monophyletic. In contrast to weighted parsimony results, non-angiosperm sequences appear as the sister clade to the COR clade, though this result is poorly supported. Other Bayesian trees are generally concordant with the results shown here, except that the non-angiosperm clade is sister to the REV/PHB clade in results based on the JONES AATM, and the REV clade is split into two clades in results based on the Dayhoff AATM.

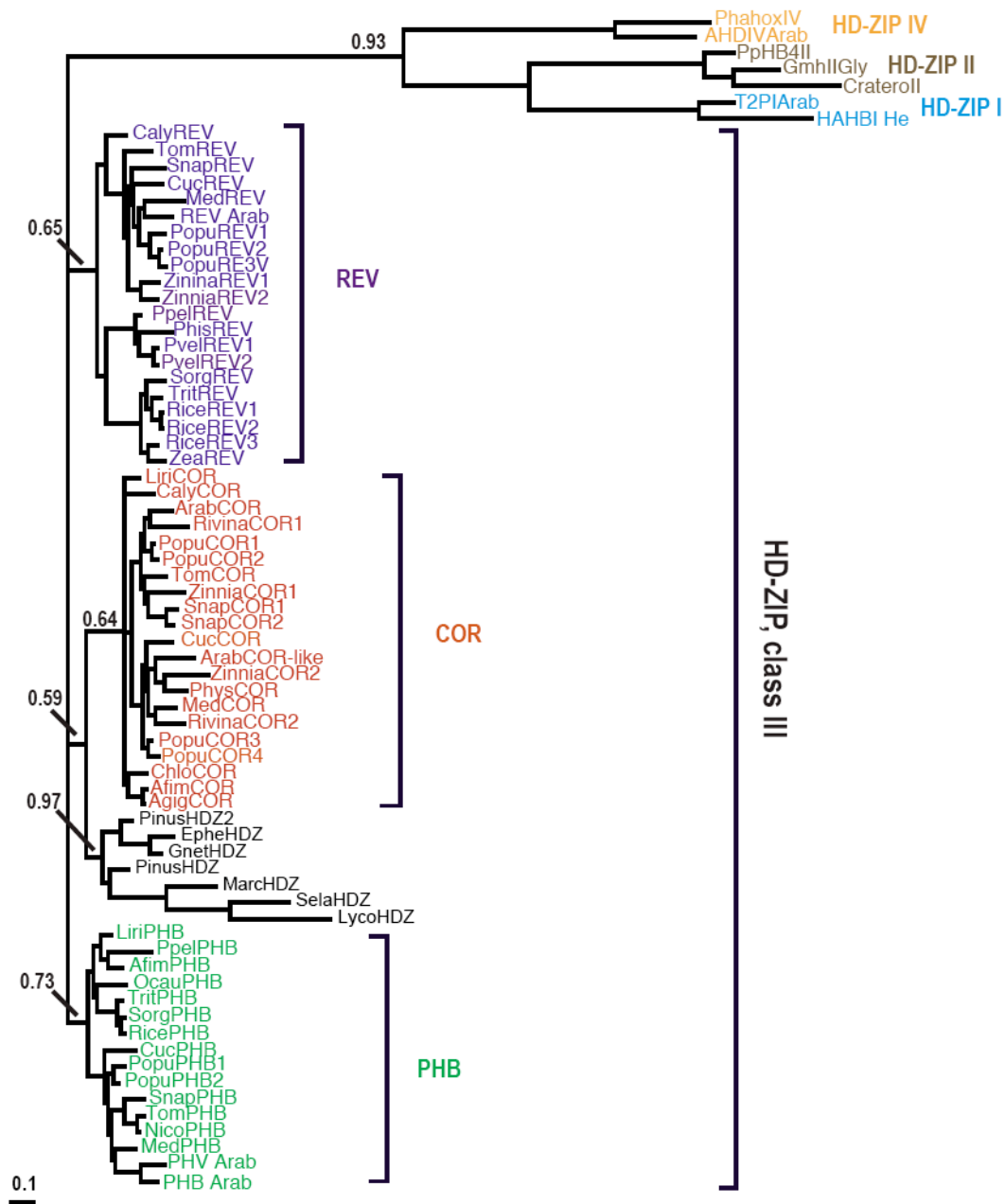


Figure 2.8. Minimum age estimates of events in the ancient history of the HDZ-III gene lineage based on BLOSSUM45 weighted-parsimony tree. Minimum age estimates for three nodes near the base of the tree, indicated by thick black arrows, are based on the composition of the clades derived from these nodes. They are the most-recent-common-ancestors (MRCAs) of an all-embryophyte, a spermatophyte only, and an angiosperm-only clade, aged 432 MYA, 334 MYA, and 252 MYA, respectively (based on minimum clade ages estimated in Schneider *et al.* 2004). The node at the base of the angiosperm-only clade corresponds to the first of three ancient gene duplication events in HDZ-III. Following calibration of a molecular clock for HDZ-III, ages for the COR, PHB, REV, and PHB/REV clades were estimated. The node at the base of the PHB/REV clade indicates a second ancient gene duplication event in HDZ-III (with a minimum age of ~222 MYA). Within the COR clade there is a third ancient gene duplication event somewhere in the ancestry of the eudicots, prior to the divergence of the Caryophyllids, Rosids, and Asterids (with a minimum age of ~ 94 MYA). These results indicate that ancient gene duplication events played a role in shaping the composition of HDZ-III.

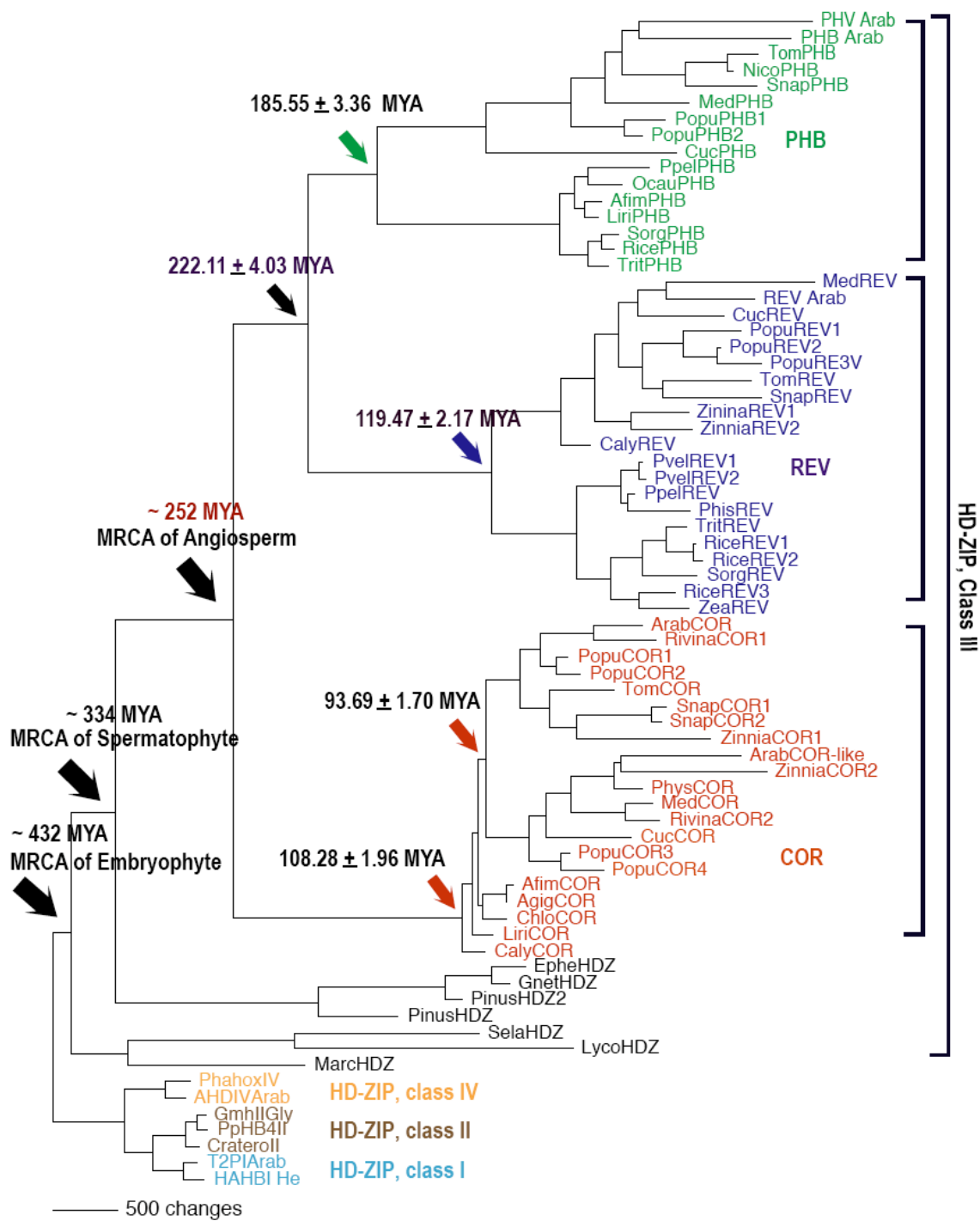
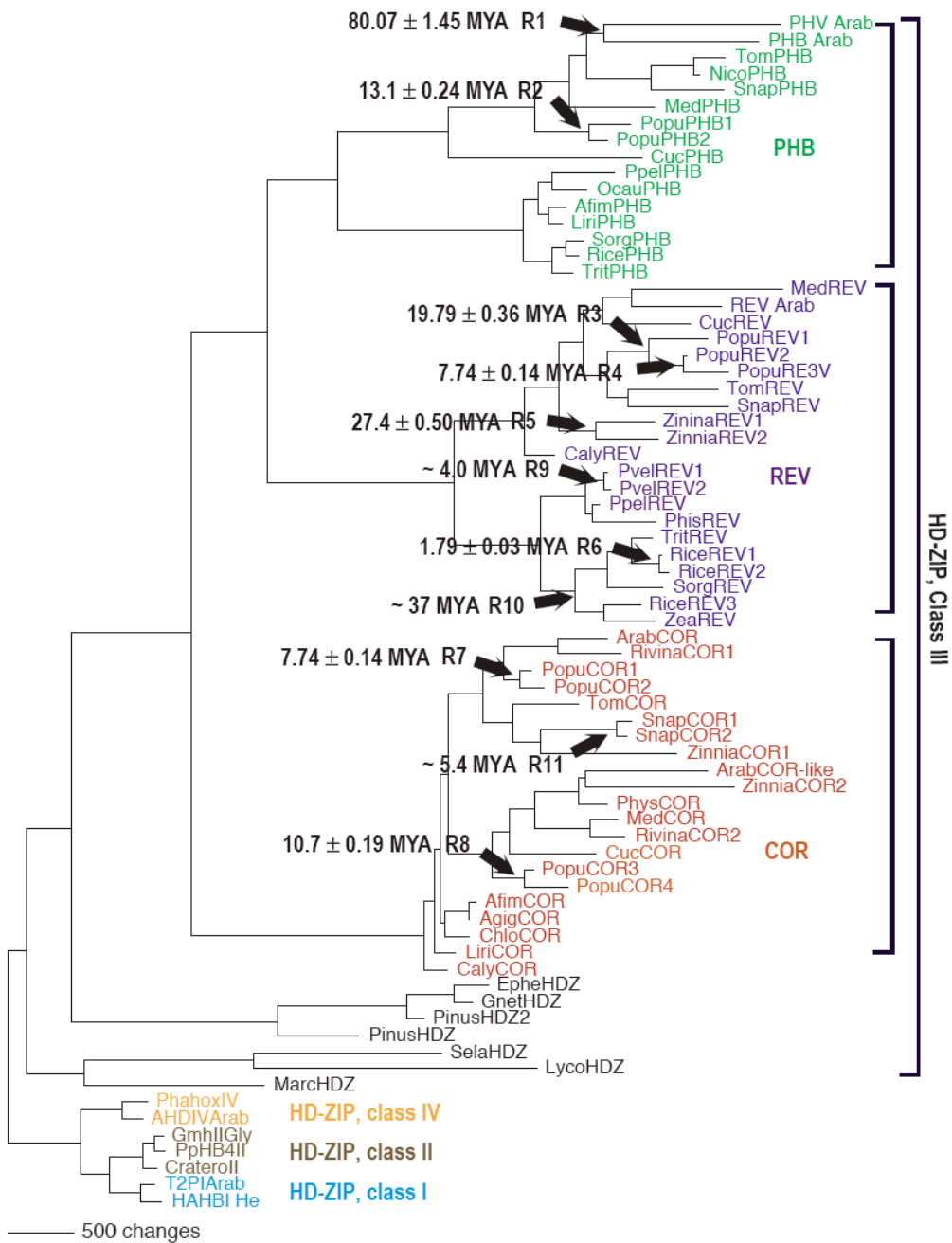


Figure 2.9. Minimum age estimates of eleven recent gene duplication events in the HDZ-III gene lineage based on BLOSSUM45 weighted-parsimony tree. Using the molecular clock calibrated for HDZ-III, minimum ages for gene duplication events in the PHB (2), REV (6) and COR (3) clades were estimated. Five of these duplication events are specific to *Populus*, two to rice (REV clade), and one each in *Antirrhinum* (COR clade), *Arabidopsis* (PHB clade), *Peperomia* (REV), and *Zinnia* (REV clade). The results indicate that gene duplication is an ongoing process in HDZ-III gene family and gene knock-out/knock-down studies in *Arabidopsis*, *Populus*, *Rice*, and *Zinnia* should account for potential cases of functional redundancy.



CHAPTER 3

COMPARATIVE STUDY OF GENES REGULATING LAMINAR DEVELOPMENT IN

LYCOPERSICON ESCULENTUM AND *MEDICAGO TRUNCULATA*

INTRODUCTION

In most plant species, shoot lateral appendages have asymmetrical features. This asymmetry has three axes: a proximodistal axis, leading from the base to the tip; an adaxial-abaxial axis, leading from the upper to the lower surface; and a centrolateral axis, leading from the midrib to the margin. Mature leaves project out from the stem and have an upper and a lower surface. When the leaf primordium emerges from the shoot apical meristem (SAM), the surface that faces the center of the meristem is called adaxial and, typically, will become the upper surface of the leaf, once it matures. The opposite surface, facing away from the center of the meristem, is called abaxial and generally develops into the lower surface of the leaf (Leyser and Day, 2003).

The SAM is at the very tip of the shoot and consists of the cells above the youngest primordium. The SAM has both a radial and a vertical structure. In terms of radial structure, the meristem typically has: 1) a central zone that has large, slowly dividing cells; and 2) a peripheral zone, which has smaller, more rapidly dividing cells and initiates leaves, buds and the outer layer of the stem. In terms of vertical structure, the meristem has two layers: tunica and corpus. The tunica layer has one or more cell layers and is at the meristem surface, whereas corpus is in deeper region of meristem with more randomly positioned cells. The ab/adaxialized positional relationship between the leaves and SAM was proposed to be the foundation of their asymmetric development. It was demonstrated in the mid 1950s that the SAM seems to provide the initial cues for the early establishment of this asymmetry because a young potato leaf could not form a flattened structure when it was experimentally isolated from the SAM through insertion of a silica chip (Sussex, 1955).

The root apical meristem is at the basal-most tip of the root, behind the root cap, which acts as a protective shield to meristematic cells as the elongating root penetrates the soil. Ongoing cell division in the outmost layer of the root apical meristem replaces cells in the root cap as it is worn down through contact with the substrate. Divisions in more apical cell layers of the root meristem contribute all other tissues of the root, moving toward the shoot. The root apical meristem also has a radial structure. Moving from the outside inward there are five tissue types: epidermis, cortex, endodermis, pericycle and vascular procambium.

Four gene families are integral to establishment of ab/adxial cell identity in developing shoot lateral organs. These include the MYB and HD-ZIP III gene families, which have some members that promote adaxial identity, and the GARP and YABBY gene families, which promote an abaxial cell fate. *PHANTASTICA* (*PHAN*) and its homologs encode transcription factors in the MYB gene family. In Snapdragon, *phan* mutants have abaxialized and radially symmetrical leaves (Waites *et al.* 1998). HD-ZIP, class III (HD-ZIP III) genes, in addition to directing the development of the apical meristem and vascular bundles also promotes adaxial cell fate in lateral organs.

In *Arabidopsis*, there are five class III HD-ZIP genes: *PHABULOSA* (*PHB*), *PHAVOLUTA* (*PHV*), *REVOLUTA* (*REV*), *CORONA* (*COR*) and *ATHB8*. *PHB*, *PHV* and *REV* are initially expressed diffusely in leaf anlagen (primordia prior to emergence), then their localization becomes progressively adaxialized after primordia emerge. The fact that dramatic gain-of-function alleles of *PHB* and *REV* have single-nucleotide substitutions or a small insertion because of altered splicing near the amino end of the START domain suggested that altered ligand perception might be the cause for the gain-of-function phenotype (McConnell, *et al.*, 2001). However, a more recent study identified the complementary sites of miRNA in the

mutated region of the START domain and the same gain-of-function phenotype could be obtained by changing bases in the miRNA binding site that did not alter the sequence of the encoded protein (Rhoades, *et al.*, 2002; Tang, *et al.*, 2003, Emery, *et al.*, 2003, Kidner and Martienssen 2004, 2005). This study provided evidence that altered miRNA binding might lead to the observed gain-of-function phenotype. Hence, it was proposed that the interaction between HD-ZIPIII genes and miRNA, whose abaxial localization is controlled by *ARGONAUTE1* (*AGO1*), in part specifies ab/adaxial polarity in developing leaves (Kidner *et al.* 2004). Additionally, Sieber *et al.* (2004) recently demonstrated that *PHB* is directly involved in outer integument development, which shares ab/adaxial features with plant lateral organs.

KANADII (*KAN1*), *KANADI2* (*KAN2*), *KANADI3* (*KAN3*) and *KANADI4* (*KAN4*) encode proteins in the GARP gene family (Riechmann *et al.* 2000; Eshed *et al.* 2001; Kerstetter *et al.* 2001). The YABBY gene family comprises six members: *CRABS CLAW* (*CRC*), *FILAMENTOUS FLOWER* (*FIL*), *YABBY2* (*YAB2*), *YABBY3* (*YAB3*), *INNER NO OUTER* (*INO*) and *YABBY5* (*YAB5*) (Bowman and Smyth 1999; Siegfried *et al.* 1999; reviewed in Bowman 2002). Each asymmetric above-ground lateral organ expresses at least one member of the YABBY family in a polar manner. Among these genes, *INO* and *CRC* have more specialized functions. In *Arabidopsis*, *CRC* contributes more directly to ovule placenta and carpel primordium initiation (Kumaran *et al.*, 2002). *INO* is exclusively expressed in the outer integument of developing ovules (Villanueva *et al.* 1999).

It was suggested that the mutual antagonism between homologs of *KAN1* and *PHB* leads to the establishment of adaxial and abaxial domains in developing lateral organs and this juxtaposition promotes laminar expansion (McConnell, *et al.*, 2001; Eshed, *et al.*, 2001; Kerstetter, *et al.*, 2001). This interaction helps specify the polar expression of YABBY genes,

which contributes to both adaxial/abaxial juxtaposition mediated lamina expansion and abaxial cell fate (Eshed *et al.* 2004).

The species, *Lycopersicon esculentum* and *Medicago trunculata*, chosen in this study fall into Asterids and Rosids within the eudicot clade. They have distinct leaf forms: tomato has pinnately compound leaves and medicago has non-peltately compound leafes. These two species provide us with an opportunity to explore how homologs of *PHB*, *REV* and *COR* are expressed in species that have differently shaped leaves and to gain insights into compound and simple leaf development. Additionally, comparing patterns of gene expression for duplicate gene pairs in diverse tissues of these two species may reveal cases of functional evolution in the gene family. Lastly, comparison of gene expression patterns in these two species will permit evaluation of models of plant development and gene function based on research in *Arabidopsis*.

MATERIALS AND METHODS

Tissue collection, fixation and embedding

SAM tissue for histology and paraffin sections was collected from 5-week old plants grown in the Plant Biology Department's greenhouse at the University of Georgia. Prior to appearance of the first flowers, shoot apices were severed from the plant and placed into FAA fixative. Additionally, Medicago and tomato seeds were grown on wet paper towels in a petri dish for ~ 2-7 days until immature roots and cotyledons sprouted. I collected these medicago and tomato seedlings for tissue fixation and embedding, using a razor blade to separate the shoot and root regions of the seedlings. Immediately upon collection, I placed all harvested tissue in FAA fixative and placed the collection vials on ice. After applying a slight vacuum (~ 20 in Hg) for ~ 1 hour, samples were agitated and squeeze to remove any air bubbles present. After the tissue

sank in the fixative, I gently rocked the tissues in FAA for ~ 3 hours in room temperature. Tissues were then dehydrated in an ethanol series, dyed with Eosin-O (Fisher), moved to HistoClear II (National Diagnostics) in gradations, infiltrated with Paraplast (Fisher) for three days, and cast into wax bricks.

Gene cloning

I used a degenerate PCR based approach to amplify initial fragments of *Medicago PHB* (*MedPHB*), *Medicago REV* (*MedREV*), *Medicago CORONA* (*MedCOR*), *Tomato PHB* (*TomPHB*), *Tomato REV* (*TomREV*) and *Tomato CORONA* (*TomCOR*) (using primers listed in Table 2.1). Then gene specific 3' and 5' RACE primers were designed to clone 3' and 5' ends of those genes using the Generacer kit (Invitrogen). I designed specific primers based on the full-length coding sequences complete with 5' and 3' UTRs to amplify and clone the full-length genes for subsequent subcloning (TOPO TA Cloning kit, Invitrogen, Carlsbad, Calif.).

Since homeodomain containing genes are abundant in plant species, I designed a forward primer for each full length gene just beyond the 3' end of the homeodomain and amplified and TA-cloned the remainder of the gene for subsequent ribo-probe synthesis. This resulted in a portion of the gene that averaged ~2,200 bp in length. Following TA cloning of the putative 3'-end of the gene, five white colonies were selected randomly from the central region of each LB plate. In order to check if the construct contains the right insertion and to determine the orientation of the inserted fragment, I re-amplified the insert from each selected colony using three combinations of primer pairs: 1) the probe cDNA-specific forward and reverse primers; 2) the M13-(-20) forward universal and probe-specific forward primer; and 3) the M13-reverse

universal and probe-specific forward primer. PCR products were resolved on a 1.2% agarose gel and evaluated for presence/absence and length.

Probe synthesis and precipitation

Ribo-probes were generated according to the following procedure. 1) Grow four overnight cultures in LB broth for each construct. Extract plasmids with QIAprep Spin Miniprep Kit (Qiagen). 2) Pool the four plasmid extractions and quantify them on a spectrophotometer (Bio-Rad SmartSpec 3000). 3) In a 100 μ l reaction, digest 20 μ g of each construct for ≥ 2 hours with 10 μ l 10X buffer (NEB), 10 μ l 10X BSA, and 3 U of BamHI or XhoI (BioLab) enzymes (leaving a 5' overhang). 4) Use QIAquick Gel extraction (Qiagen) to clean the digested construct. 5) Check the concentration of the linearized plasmid on a spectrophotometer (Bio-Rad SmartSpec 3000). 6) Add 5 μ g of digested construct to one 0.7 ml tube per synthesis reaction. 7) Bring the volume up to 13.125 μ l with RNase free water. 8) Make master mix for n +1 reactions following the one reaction formula using Roche products: 5 μ l 5X transcription buffer, 2.5 μ l DTT, 0.625 μ l RNasin, and 10X DIG labeling mix. 9) Mix 10.625 μ l of the master mix into the tubes containing DNA and water. Add 1.25 μ l T7 or Sp6 polymerase to each tube. 10) Incubate for 2 hours at 37°C, add 1 μ l RQ1 RNase free Dnase (Roche) and leave the tubes in 37°C water bath for 15 minutes. 11) Transfer the probes into a 1.5 ml centrifuge tube and add 2 μ l 10 μ g/ μ l yeast tRNA, 40 μ l 5M NH₄Oac, 33 μ l RNase free water and 250 μ l 100% ethanol. 12) Let the probes sit in room temperature for 1 hour. Then spin probes for 15 minutes at 14,000 rpm at room temperature, dump the supernatant, add 200 μ l 80% ethanol, spin for 15 minutes at 14,000 rpm at room temperature, pour off the ethanol, air dry the probes for 15 minutes, resuspend the probe by adding 210 μ l RNase free water. To precipitate the probe, add 23.6 μ l

3M, pH5.2 sodium acetate with 1.0 µl 20 mg/ml Glycogen (Roche), 2.0 µl 10 µg/µl yeast tRNA and 575.0 100% ethanol and let the mixture sit for at least 3 hours in -20 °C.

In situ hybridization

I used methods revised from a protocol used in the laboratories of John Bowman (Siegfried *et al.* 1999) and Marty Yanofsky (Ferrandiz *et al.*, 2000). Briefly, ribbons made by serial section at 10 µm are mounted on Probe-On Plus slides (Fisher), heated for ~24 hours at 43°C, hydrated in a water-ethanol series, treated with diluted HCL for 20 minutes, treated with Pronase (Sigma) for ~30 minutes at 37°C, fixed 3.7% formaldehyde in 1X PBS for 10 minutes, dehydrated in an ethanol series and placed in a vacuum for ~15 minutes.

Dilute synthesized probes in hybridization buffer at 50°C. Apply 300 µl of probe + hybridization buffer to slides and form “slide sandwiches” for overnight hybridization. When all the slides are done, move them to a pre-warmed, humidified box containing Kimwipes saturated in 2X SSC and 50% formamide. Incubate slides in humidified boxes for ~12 hours at 54°C, wash them twice in 0.2X SSC , 50% formamide, expose to 1:3000 anti-DIG AP-conjugated Ab (Roche), for one and a half hours, rewash, place the slides in detection buffer with a concentration of 1ml NBT+BCIP/50ml detection buffer for 6-12 hours, cover-slip and photograph with a ZEISS Axioskop microscope equipped with a Kodak MDS 290 digital camera.

Probe concentration and detection time

I detected the diffused expression of several probes at my beginning stage of in-situ experiment. In order to get differentiated expression, I experimented with different combinations

of probe concentration and length of detection time. Generally 0.5 to 2 μ l of probe per 300 μ l of hybridization buffer and 6 to 12 hours in detection medium appeared optimal.

RESULTS

Expression pattern in *Lycopersicon esculentum*

I detected abaxialized expression of *TomFIL* in the longitudinal and cross sectioned leaves of tomato throughout early stages of their development (Figure 3.1 A and B). This result is similar to the expression pattern for *FIL* in *Arabidopsis* (Siegfried *et al.* 1999, Villaneuva *et al.* 1999, Meister *et al.* 2002). *TomFIL* (Figure 3.2 A) also showed abaxialized expression in developing petals. This expression pattern was in sharp contrast to that of *TomCOR* (Figure 3.2 B) in the flower meristem, in which *TomCOR* showed much stronger adaxialized expression in petal primordia. Since the petal is a modified leaf, the complementary expression pattern detected in this study support the hypothesis of previous research that genes from the YABBY gene family and genes from the HD-ZIP, class III gene family promote abaxial and adaxial polarity in leaf development, respectively (McConnell, *et al.*, 2001; Eshed, *et al.*, 2001; Kerstetter, *et al.*, 2001). Additionally, *TomCOR* had distinct expression in vascular pro-cambial tissue and the apex (center) of the floral meristem. *TomREV* also has adaxialized expression in petal primordia and young petals. However, as the petal elongates away from the floral meristem, the adaxialized expression becomes progressively lighter (Figure 3.3 A). It is obvious that *TomCOR* had stronger expression than did either *TomFIL* or *TomREV*. *TomPHB* did not show any expression in either floral or leaf primordia, even though a wide range of experimental conditions were tried. This suggests that

TomCOR is playing an important role in young petal polarity determination and that the promoter of *TomCOR* is more active than that of *TomREV* or *TomPHB* in these tissues.

I detected an adaxialized expression pattern for *TomCOR* (Figure 3.3 C, D and E) and *TomREV* (Figure 3.3 B) in young leaf and leaf primordium in both longitudinal and cross sections of shoot apices. *TomCOR* is expressed diffusely in the young leaf primordia (Figure 3.3 D). As the leaf primordium develops, the expression of *TomCOR* becomes progressively adaxialized (Figure 3.3E) and eventually becomes exclusively adaxialized in young leaf (Figure 3.3 C). Here as well, no expression pattern for *TomPHB* was detected in mature SAM (Figure 3.4 A). In order to determine if *TomPHB* has any expression in the mature shoot apical meristem, I carried out the experiment more than four times by using probe concentration with 0.5-2.0 μ l/300 hyb-buffer and 8-12 hours detection time. However, I did not detect any expression of *TomPHB* in the SAM.

In immature root, *TomCOR* is expressed throughout the root tip as seen in longitudinal section (Figure 3.5 A). Stronger expression was shown in the root cap, root apical meristem, and root vascular procambium. However, I did not see any expression of *TomREV* (not shown) or *TomPHB* (Figure 3.5 B) in root tissues, even though the experiment was repeated several times as before.

In immature shoot, *TomCOR* (Figure 3.6 A) showed expression diffusely in the cross section of cotyledon and young shoot apical meristem. Stronger expression was seen in the vascular tissues. Slightly adaxial localized expression was seen in the immature shoot apical meristem. A similar expression pattern was seen for *TomREV* (Figure 3.6 B), but the apparent level of expression of *TomREV* is much lower under the same experiment conditions. All results

showed that expression of *TomPHB* is not distinguishable from that of the negative-control sense probe (Figure 3.6 C).

Expression pattern in *Medicago trunculata*

MedFIL (Figure 3.7 A) is expressed initially throughout the leaf anlagen, before any superficial outgrowth is visible, and later becomes progressively sequestered to the abaxial region of a developing leaf when the primordium emerges and expands (Figure 3.7 A and B). This abaxialized expression pattern is very distinct in older leaf primordium P5 (Figure 3.7 B). This reveals strikingly similar expression patterns to *FIL* in *Arabidopsis* (Siegfried *et al.* 1999, Villaneuva *et al.* 1999, Meister *et al.* 2002). No expression for *MedFIL* was detected in either the floral meristem or the petal primordium.

In longitudinal section of a mature SAM (Figure 3.8 B) and cross section of young leaf (Figure 3.8 A), I observed ubiquitous expression of both *MedPHB* (Figure 3.8 B) and *MedREV* (Figure 3.8 A). Similar expression was seen with *MedCOR* (not shown). In order to exclude the possibility that this equal expression pattern is caused by the high concentration of probe and long detection time, I tried several runs by reducing the probe concentration from 4 μ l/300 hyb-buffer to 0.5 μ l/300 hyb-buffer and shortening the detection time from 12 hours to 6 hours. However, the same ubiquitous expression pattern was observed under all conditions for *MedPHB*, *MedREV* and *MedCOR*. The expression of sense probe (Figure 3.8 C) is clear.

In immature roots, *MedPHB* had expression throughout the longitudinal section of root tip, but stronger expression of *MedPHB* is evident in the root cap, root vascular procambium, and root apical meristem (Figure 3.9 A). Similar expression was detected with *MedREV* (Figure 3.9

B), but with an apparently lower level expression. No expression of *MedCOR* was detected in immature roots.

In the immature shoot, expression of *MedPHB*, *MedREV* and *MedCOR* was detected in all cells (Figure 3.10 A and B), but stronger expression was seen in the vascular bundle and immature shoot apical meristem. The sense probe did not produce any apparent background. Among these three genes, *MedPHB* had the highest level of expression and even showed slightly adaxialized expression in leaf primordia at the tip of the immature shoot apex.

DISCUSSION

General expression pattern of *FIL* orthologs and *PHB* homologs

Previous study suggested that the signal coming from the shoot apical meristem an adaxial cell fate signal. This signal is thought to be mediated by PHB-like proteins (McConnell, *et al.*, 2001). *YABBY* and *KANADI* orthologs are expressed abaxially throughout lateral organ anlagen. Mutual suppression of *PHB*-like genes and *YABBY* and *KANADI* orthologs generates their spatially circumscribed patterns of expression (Siegfried, *et al.*, 1999; Sawa, *et al.*, 1999; Kerstetter, *et al.*, 2001, Prigge, 2004; Green, 2005). Accordingly, genes from the *YABBY* family generally have complementary expression patterns to genes from the HD-ZIP, class III gene family.

In-situ hybridization results reported here for tomato and medicago are generally consistent with the expectation that *YABBY* and HD-ZIP III genes will have complementary patterns of expression. *TomCOR* and *TomFIL* had complementary expression in petal primordium and young petals. However, *TomCOR* showed much stronger expression than *TomFIL*. *TomREV* also had an adaxial pattern of expression, though it was detected at a much

lower level in both petal primordia and young petals. These results suggest that *TomCOR* has an important role in establishing adaxial polarity identity in lateral organs. *TomCOR* may even play a major role in establishing adaxial identity in the petal primordia and in excluding the expression of abaxial polarity establishing genes from this region. In contrast to the role of its ortholog in *Arabidopsis*, an absence of detectable expression suggests that *TomPHB* may not be involved in lateral organ polarity determination in tomato.

MedFIL showed abaxialized expression in leaf anlagen and young leaf primordium. This localized expression is consistent with the results published from *Arabidopsis* (Siegfried *et al.* 1999, Villaneuva *et al.* 1999, Meister *et al.* 2002) and that of *TomFIL* investigated in this study. However, I did not detect any adaxialized expression of *PHB* homologs in mature shoot apical meristem or flower meristem. This suggests that there are other genes that interact with *MedFIL* to promote laminar outgrowth or function independently to promote the adaxial polarity in medicago lateral organ development.

Functional diversity among HD-ZIP, class III genes

Gene families are the major components in organism genomes. Gene duplication leading to the birth of paralogous genes is the primary known mechanism in the growth of a gene family. The six genes studied here are all the products of ancient gene duplication events. There are three possible fates for the duplicated gene pairs:

1) functional redundancy is least common and most likely with highly expressed housekeeping genes; 2) neofunctionalization is most common in single celled organism and involves an elevated rate of amino acid substitutions in conserved regions that were previously under stabilizing selection; and 3) subfunctionalization is most common in multicellular organisms and

could occur in either the protein coding region or promoter region of a duplicated gene (Zhang, 2003). If the subfunctionalization occurred in the protein-coding region, this could lead to the specialized protein function in daughter genes when one of the paralogous genes is better at performing one of the original functions of their ancestral gene (Zhang, *et al.*, 1998; reviewed in Zhang, *et al.*, 2003). If the subfunctionalization happened in the promoter region, this could cause the differentiated expression patterns in duplicated genes.

Adaxialized expression of *TomCOR* is strongly evident in both petal and leaf primordia and is expressed diffusely in roots. These results suggest that *TomCOR* is an important polarity determination factor in tomato lateral organ development. In contrast, *TomPHB* seems to have lost its function in the development of shoot lateral organs, since no expression of *TomPHB* was detected in these tissues. *TomREV* exhibits an intermediate level of expression, suggesting that it has a weaker promoter than that of *TomCOR*. Accordingly, this study provides evidence that sometime after their duplication events, the promoters of these three genes diverged in function.

In Chapter 2 (Figure 2.8), I determined that those three HD-ZIP, class III homologs are derived from ancient gene duplication events that occurred at least 200 million years ago. After their divergence from the most recent common ancestor of all the three genes, tomato *TomCOR* seemed to retain the dominant and presumably ancestral function of the gene family. In contrast, *TomPHB* seems to have lost its function in developing shoot lateral appendages. *TomREV* appears to have retained at least part of the ancestral HD-ZIP III function in shoot lateral appendages, but appears to have lost its activity in immature root development. Accordingly, there might be a case of promoter subfunctionalization or neofunctionalization following the duplication events that produced *TomCOR*, *TomPHB* and *TomREV*.

On the other hand, *MedPHB*, *MedREV* and *MedCOR* all have diffuse expression in mature and immature shoot apical meristem with slight differences in their expression levels in immature shoots. This suggests that these three genes have all deviated from the function of their most recent ancestor in shoot apical meristem development after the ancient duplication events and illustrates the potential for species-specific outcomes for the functional evolution of duplicated genes. Accordingly, this study underscores the importance of conducting studies of development in diverse organisms both to refine models of plant development based on select model research organisms and to provide a foundation for better understanding of the forces that drive plant developmental evolution.

REFERENCES

- Bowman, J. L. and Smyth, D. R. 1999. CRABS CLAW, a gene that regulates carpel and nectary development in Arabidopsis, encodes a novel protein with zinc finger and helix-loop-helix domains. *Development* **126**: 2387-2396.
- Emery, J. F., Floyd, S. K., Alvarez, J., Eshed, Y., Hawker, N. P., Izhaki, A., Baum, S. F. and Bowman, J. L. 2003 Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes. *Curr Biol.* **13**: 1768-1774.
- Eshed, Y., Baum, S. F., Perea, J. V. and Bowman, J. L. 2001. Establishment of polarity in lateral organs of plants. *Curr Biol.* **11**: 1251-1260.
- Eshed, Y., Izhaki, A., Baum, S. F., Floyd, S. K., Emery, J. F. and Bowman, J. L. 2004. Asymmetric leaf development and blade expansion in *Arabidopsis* are mediated by KANADI and YABBY activities. *Development* **131**: 2997–3006
- Ferrandiz, C., Gu, Q., Martienssen, R. and Yahofsky, M. F. 2000. Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 AND CAULIFLOWER. *Development* **127**: 725-734.
- Green, K. A., Prigge, M. J., Katzman, R. B. and Clarke, S. E. 2005. CORONA, a Member of the class III homeodomain leucine zipper gene family in *Arabidopsis*, Regulates stem cell specification and organogenesis. *The Plant Cell.* **17**: 691-704.
- Kerstetter, R. A., Bollman, K., Taylor, R. A., Bomblies, K. and Poethig, R. S. 2001. *KANADI* regulates organ polarity in *Arabidopsis*. *Nature* **411**: 706-709.
- Kidner, C. A. and Martienssen, R. A. 2004. Spatially restricted microRNA dwerects leaf

- polarity through AGRONAUTE1. *Nature* **428**: 81-84.
- Kidner, C. A. and Martienssen, R. A. 2005. The role of AGRONAUTE1 (AGO1) in meristem formation and identity. *Developmental Biology* **280**: 504-517.
- Kumaran, M. K., Bowman, J. L., Sundaresan, V. 2002. YABBY polarity genes mediate the repression of KNOX homeobox genes in Arabidopsis. *Plant Cell*. **14**: 2751-2760.
- Leyser, O. and Day, S. 2003. Mechanisms in plant development. Blackwell Publishing, Malden, US.
- Meister, R. J., Kotow, L. M., Gasser, C. S. 2002. SUPERMAN attenuates positive INNER NO OUTER autoregulation to maintain polar development of Arabidopsis ovule outer integuments. *Development* **129**: 4281-4289.
- McConnell, J. R., Emery, J. F., Eshed, Y., Bao, N., Bowman, J., and Barton, M. K. 2001. Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. *Nature* **411**: 709-713.
- Prigge, J. M., Otsuga, D., Alonso, J. M., Ecker, J. R., Drews, G. N. and Clark, S. E. Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in Arabidopsis development. 2004. *Plant Cell* **17**: 61-76.
- Pryer, K. M., Schneider, H., Smith, A. R., Cranfill, R., Wolf, P. G., Hunt, J. S. and Sipes, S. D. 2001. Horsetail and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* **409**: 618-622.
- Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B. and Bartel, D. P. 2002. Prediction of plant microRNA targets. *Cell* **110**: 513-520.

- Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C. Z. and Keddie, J. 2000. Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. *Science* **290**: 2105-2110.
- Sawa, S., Watanabe, K., Goto, K., Kanaya, E., Morita, E. H. and Okada, K. 1999. FILAMENTOUS FLOWER, a meristem and organ identity gene of Arabidopsis, encodes a protein with a zinc finger and HMG-related domains. *Genes Development* **13**: 1079-1088.
- Siegfried, K. R., Eshed, Y., Baum, S. F., Otsuga, D., Drews, G. N. and Bowman, J. L. 1999. Members of the YABBY gene family specify abaxial cell fate in Arabidopsis. *Development* **126**: 4117-4128.
- Sussex, I. M. 1955. Morphogenesis in *Solanum tuberosum* L.: orientation in the juvenile shoot. *Phytomorphology* **5**: 286-300.
- Tang, G., Reihart, B. J., Bartel, D. P. and Zamore, P. D. 2003. A biochemical framework for RNA silencing in plants. *Genes Development* **17**: 49-63.
- Villanueva, J. M., Broadhvest, J., Hauser, B. A., Meister, R. J., Schneitz, K. and Gasser, C. S. 1999. INNER NO OUTER regulates abaxial- adaxial patterning in Arabidopsis Ovules. *Genes Development* **13**: 3160-3169.
- Waites, R., Selvadurai, H. R. N. Oliver, I. R. and Hudson, A. 1998. The PHANTASTICA gene encodes a MYB transcription factor involved in growth and dorsiventrality of lateral organs in Antirrhinum. *Cell* **93**: 779-789.
- Zhang, J., H. F. Rosenberg, H. F. and Nei, M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci USA*. **95**: 3708-3713.

Zhang, J. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol.* **18** :292-298.

Table 3.1 In-situ hybridization conditions and results in this study. FAA is the fixation for all the tissues. Probe concentration is in μl probe/ 300 μl hybridization buffer. There are five gene expression pattern detected in this study: 1) Some, ab: expression is localized to the abaxial side of plant tissue; 2) Some, ad: expression is localized to the adaxial side of plant tissue; 3) All, equal: expression is throughout the plant tissue; 4) All, diff.: expression is throughout the plant tissue, but some ab/adaxialized region has stronger expression; 5) None: no expression pattern is detected.

Probe	Tissue type	* Probe Con.	Detection	Fixation	Expre. Pattern
<i>TomFIL</i>	leaf				some, ab
<i>TomCor.</i>	Immature Root	0.5/300hyb	8h	FAA	all, equal
	Immature Shoot	0.5/300hyb	8h	FAA	all, diff
	Mature Veg. Shoot	0.5ul/300hyb	8h	FAA	some, ad
<i>TomPHB</i>	Immature Root	2/300hyb	12h	FAA	None
	Immature Shoot	0.5/300hyb	8h	FAA	None
	Mature Veg. Shoot	2/300hyb	12h	FAA	None
<i>TomREV</i>	Immature Root	0.5/300hyb	8h	FAA	None
	Immature Shoot	2/300hyb	12h	FAA	all, diff
	Mature Veg. Shoot	2/300hyb	12h	FAA	some, ad
<i>TomSense</i>	Immature Root	0.5/300hyb	8h	FAA	none
	Immature Shoot	2/300hyb	12h	FAA	none
	Mature Veg. Shoot	2/300hyb	12h	FAA	none
<i>MedFIL</i>	immature shoot	2/300hyb	12h	FAA	all, diff
	Mature Veg. Shoot	1/300hyb	12h	FAA	some, ab
	Young leaf	1/300hyb	12h	FAA	some, ab
<i>MedCor</i>	Immature Root	0.5/300hyb	8h	FAA	all, equal
	Immature Shoot	0.5/300hyb	12h	FAA	all, diff
	Mature Veg. Shoot	0.5/300hyb	8h	FAA	all, equal
<i>MedPHB</i>	Immature Root	0.5/300hyb	8h	FAA	all, diff
	Immature Shoot	0.5/300hyb	8h	FAA	all, diff
	Mature Veg. Shoot	0.5/300hyb	8h	FAA	all, equal
<i>MedREV</i>	Immature Root	0.5/300hyb	6H	FAA	all, equal
	Immature Shoot	0.5/300hyb	8h	FAA	all, diff
	Mature Veg. Shoot	0.5/300hyb	8h	FAA	all, equal
<i>Sense</i>	Immature Root	2/300hyb			
	Immature Shoot	0.5/300hyb	12h	FAA	none
	Mature Veg. Shoot	0.5/300hyb	8h	FAA	None

Figure 3.1. Expression for *TomFIL* in longitudinal and cross sections of leaves. (Ad = adaxial, Ab = abaxial). A) *TomFIL* showed abaxialized expression pattern in longitudinal section of tomato young leaf; B) *TomFIL* showed abaxialized expression pattern in cross section of tomato young leaf.

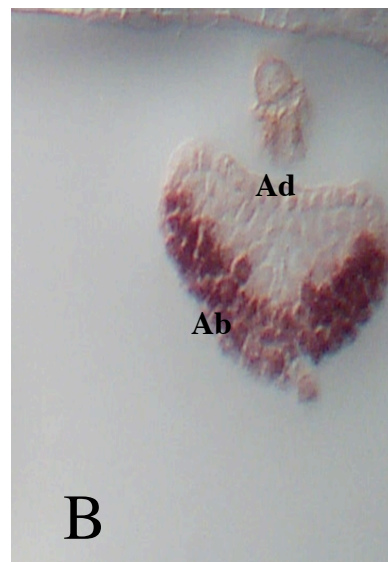
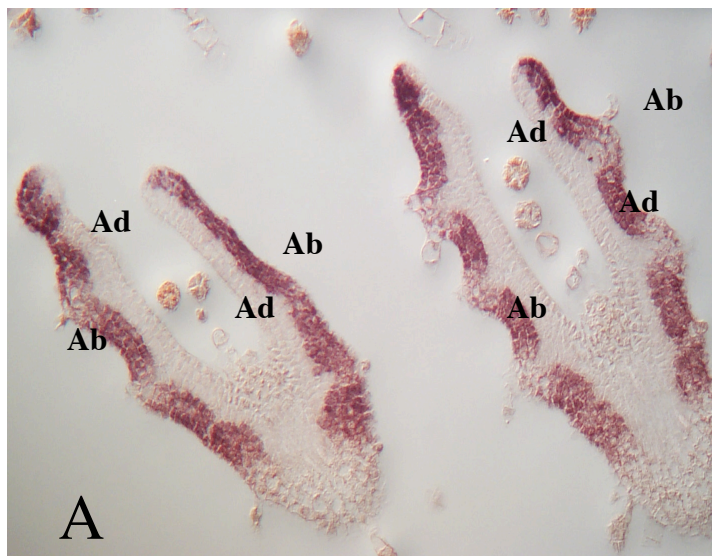


Figure 3.2. Expression for *TomFIL* and *TomCOR* in flower meristem, young petal and petal primordium. A) *TomFIL* showed abaxialized expression pattern in young petal;
 B) *TomCOR* has very strong adaxialized expression expression in petal primordium. Black row indicates the strong expression is prevascular tissue. Red arrow indicates strong expression was also seen in the center of flower meristem. Green arrows indicate the petal primordium. *TomCOR* has much stronger expression compared to the express seen with *TomFIL* when the experiment was carried out at the same conditions. This suggests that HD-ZIP, class III genes play stronger role in polarity patterning in flower tissue development (ab = abaxial, ad = adaxial, PE = petal, FM = flower meristem).

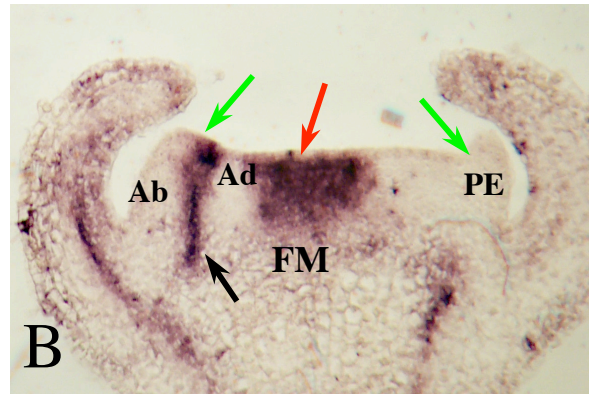
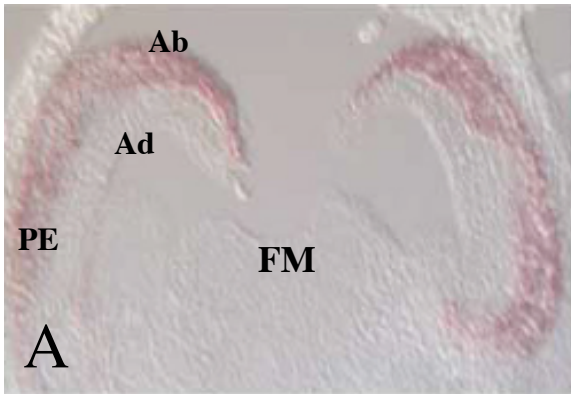


Figure 3.3 Expression for *TomREV* and *TomCOR* in FM, SAM, petal primordium and cross section of older leaf. A) *TomREV* has adaxialized expression in petal primordium and young petal. As the petal gets more independent to the flower meristem, the adaxialized expression becomes lighter. This suggests *TomREV* is involved in establishing the initial adaxial polarity in petal; B) In this cross section of tomato young leaf, *TomREV* showed adaxialized expression pattern; D) *TomCOR* has the expression throughout the young leaf primordium; E) As the leaf primordium becomes older, the expression of *TomCOR* is localized to the adaxialized side of the older leaf primordium.

C) *TomCOR* has abaxialized expression in young leaf. (ab = abaxial, ad = adaxial, SAM = Shoot apical meristem, PP = petal primordium, YP = young petal, P1 = young leaf primordium, P4 = older leaf primordium). C, D and E are longitudinal section of SAM, B is cross section of a young leaf).

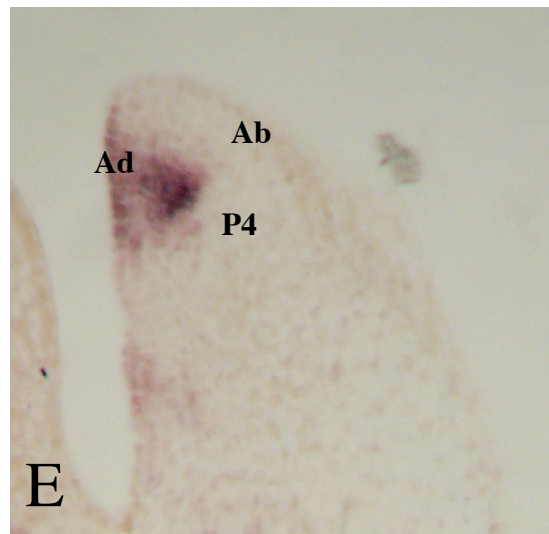
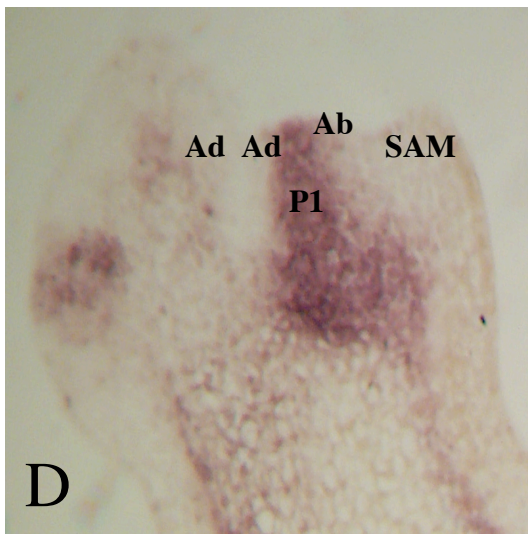
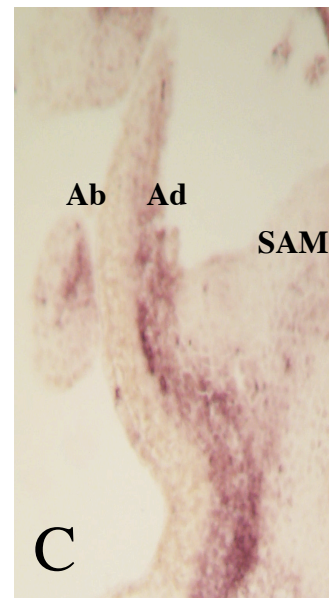
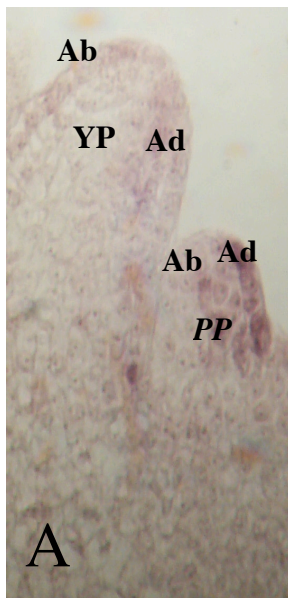


Figure 3.4. Expression of *TomPHB* in SAM compared to the expression of sense probe.
A) *TomPHB* did not show any expression in the longitudinal section of SAM, even though this experiment is repeated more than three times by using probe concentration with 0.5-0.2 μ l/300 hyb-buffer and 8-12 hours detection time. B) Sense probe just has very slightly pink background. This suggests that the clear expression seen with *TomPHB* is not an artificial effect (ab = abaxial, ad = adaxial, SAM = Shoot apical meristem, P1 = young primordium, P4 = older leaf primordium).

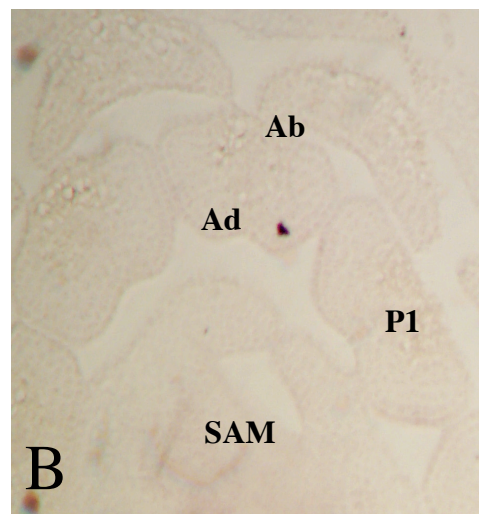
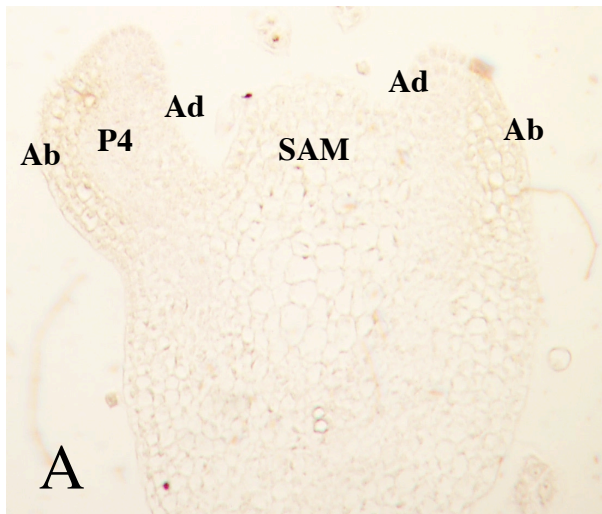


Figure 3.5. Expression for *TomCOR*, *TomPHB* in Tomato immature root compared to the expression of sense probe in immature root. A) *TomCOR* showed expression throughout longitudinal section of root. Stronger expression was seen in root cap and root apical meristem. Lighter expression is seen in the root vascular procambium. B) *TomPHB* did not show any expression in root. Since *TomPHB* shares the same expression to the sense probe when this experiment was repeated more than three times by using probe concentration with 0.5-0.2 μ l/300 hyb-buffer and 8-12 hours detection time, this suggests that *TomPHB* does not play roles in root development. *TomCOR* is probably actively involved in immature root and root vascular development. *TomREV* is the same as *TomPHB* (picture not shown). C) Sense probe expression pattern (RC = root cap, RAM = root apical meristem, RVP = root vascular procambium, ZOCE = zone of cell expansion).

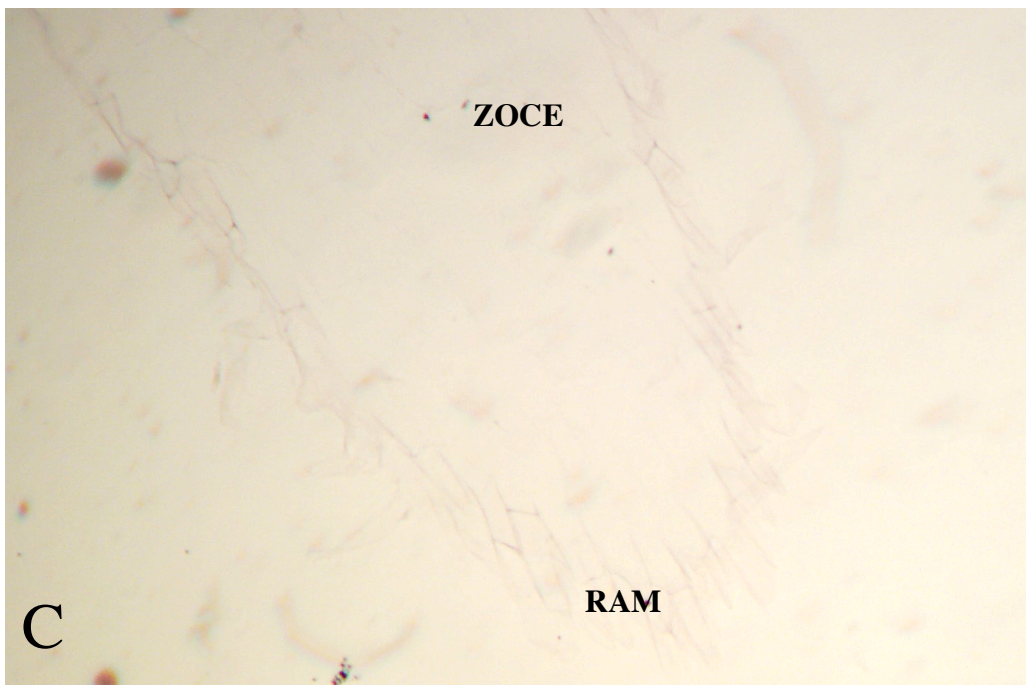
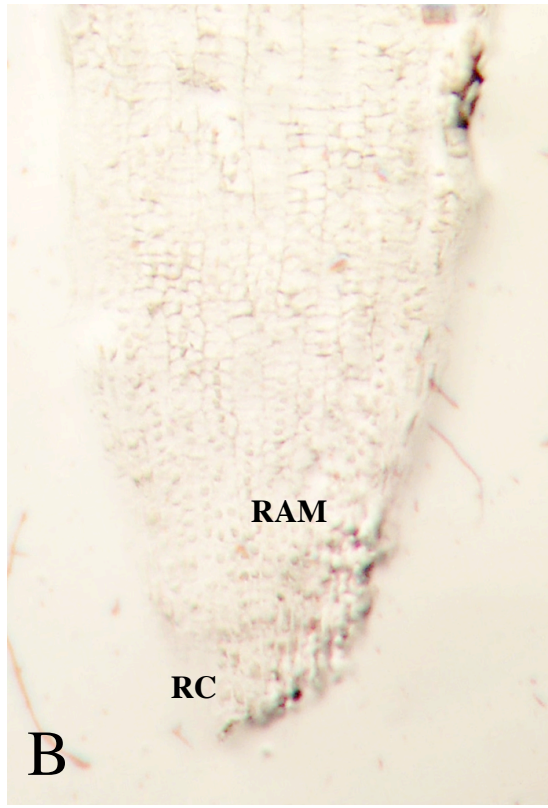
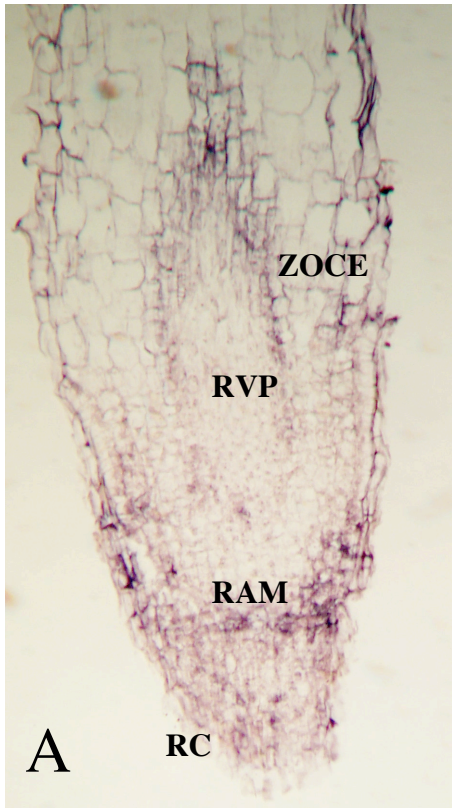


Figure 3.6. Expression of *TomCOR*, *TomREV* and *TomPHB* in immature shoot compared to the expression of sense probe. A) *TomCOR* showed expression throughout the cross section of cotyledon and young shoot apical meristem. Stronger expression was seen in the vascular tissues. Slightly adaxialized localized expression was seen in young shoot apical meristem. B) *TomREV* showed nearly the same expression pattern. But the level of stain is lighter than that of *TomCOR* given the same experiment conditions are followed. This suggests that *TomCOR* has more involved in shoot development. C) *TomPHB* did not show any expression as did D) the sense probe. (Arrows indicate the immature shoot apical meristem. ab = abaxial, ad = adaxial, SAM = Shoot apical meristem, VA = vascular bundle, Coty = cotyledon).

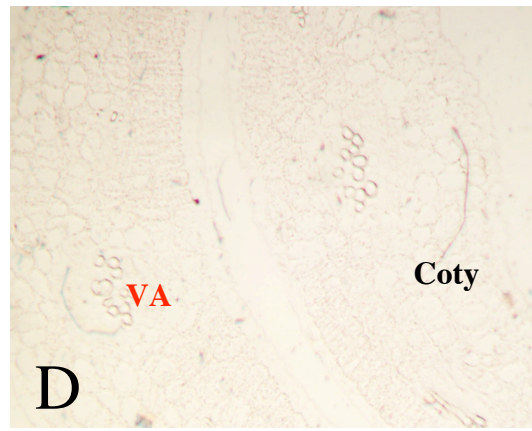
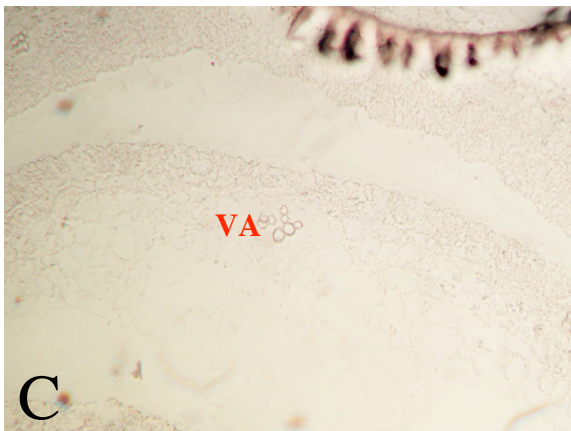
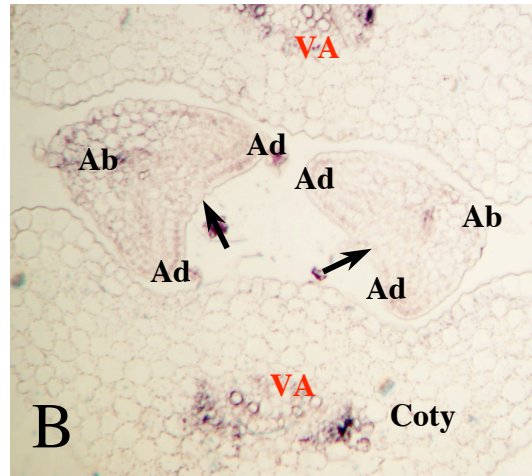
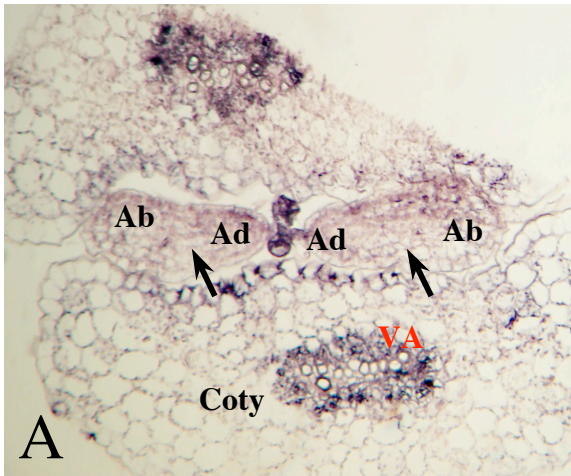


Figure 3.7. Expression for *MedFIL* in young leaf primordia and anlagen. A) *MedFIL* has expression all over the leaf anlagen. While the leaf primordium gets old, *MedFIL* slightly lost the adaxialized expression as seen in P1. *MedFIL* lost the adaxialized expression in older leaf and became completely abaxialized as seen in young leaf P5. B) *MedFIL* showed localized expression in the abaxial side of the young leaf. (Arrow indicates the space between two young leaves, ab = abaxial, ad = adaxial, SAM = Shoot apical meristem, AN = anlagen, P1 = youngest leaf primordium, P5 = young leaf).

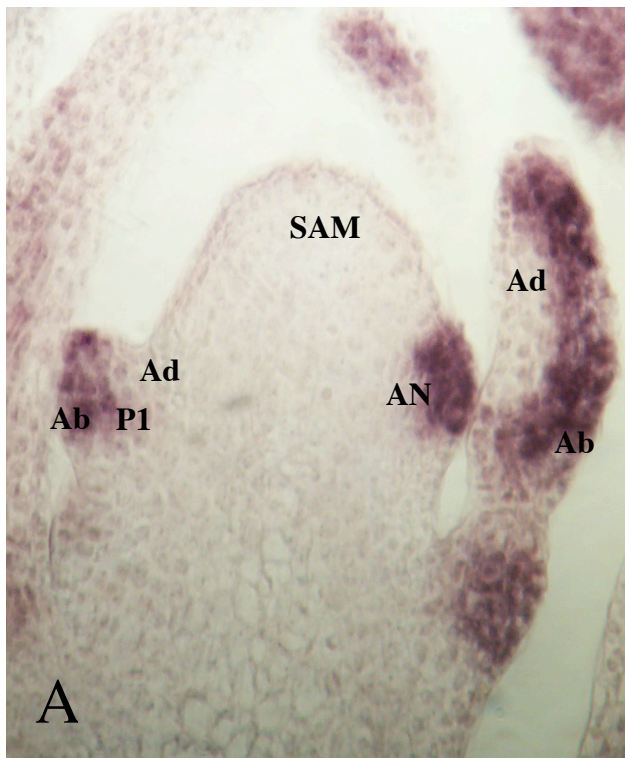


Figure 3.8. Expression of *MedREV* and *MedPHB* in mature shoot apical meristem (SAM) compared to the expression of sense probe. A) *MedREV* showed expression throughout the cross section of young leaves. B) *MedPHB* had expression throughout the shoot apical meristem. This expression was also seen with *MedCOR*. C) Sense probe just had very light pink background. This experiment was repeatedly more than four times by reducing the probe concentration from 4 μ l/300 hyb-buffer to 0.5 μ l/300 hyb-buffer and shortening the detection time from 12 hours to 6 hours. We got the same pattern of universal expression for all the medicago *PHB* homologs (ab = abaxial, ad = adaxial, SAM = shoot apical meristem, P4 = older leaf primordium).

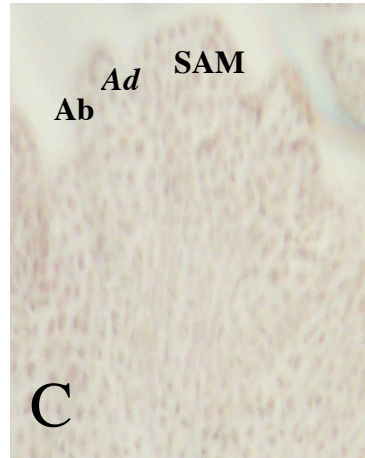
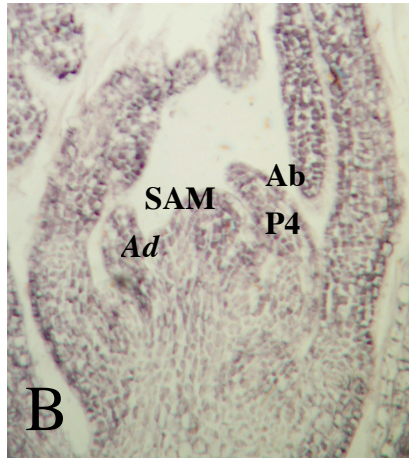
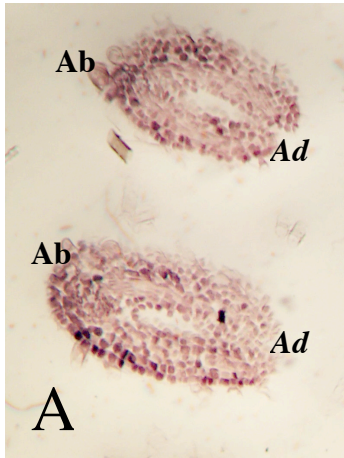


Figure 3.9. Expression of *MedPHB* and *MedREV* in immature root in contrast with the sense probe expression. A) *MedPHB* showed expression throughout the longitudinal section of the root. But stronger expression was seen in the root cap, root apical meristem and root vascular procambium. B) *MedREV* showed similar expression pattern to that of *MedPHB*. But the level of stain is much lighter with *MedREV*. This suggests that *MedPHB* is more heavily involved in immature root development given that this experiment was repeated more than four times. C) Sense probe did not show any expression in longitudinal section of root. Compared to the expression pattern of *MedPHB* and *MedREV*, this suggests that the diffused expression showed by *MedPHB* and *MedREV* was not artificial effect. (RC = root cap, RAM = root apical meristem, ZOCE = zone of cell expansion).

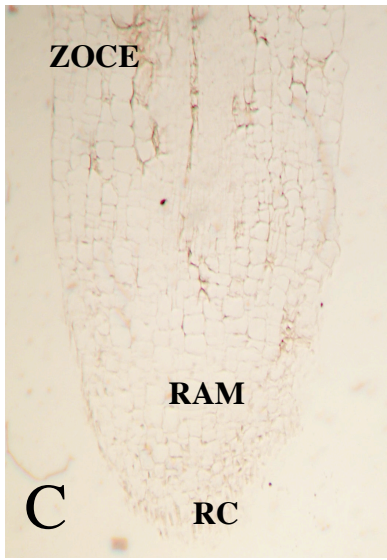
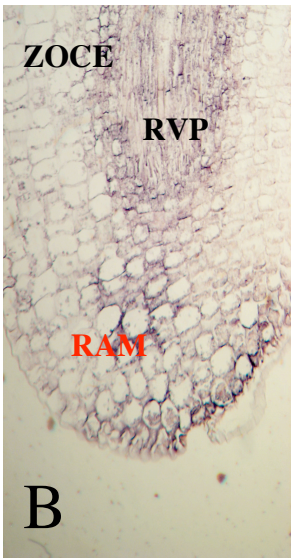
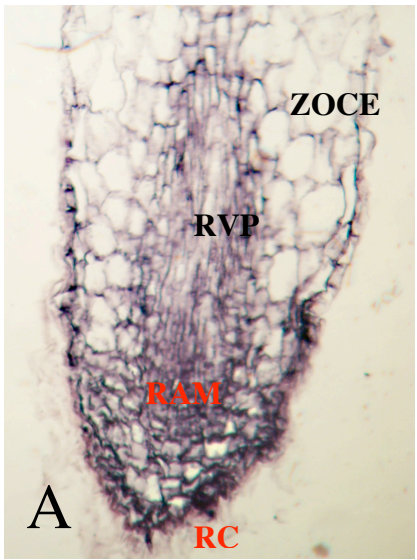
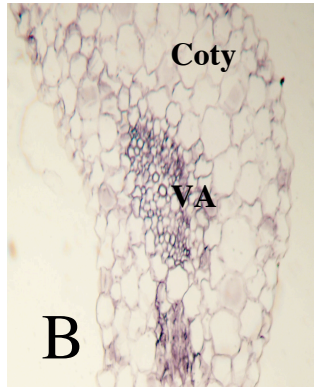
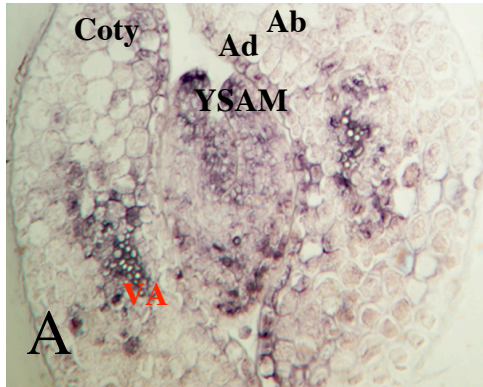


Figure 3.10. Expression of *MedPHB* and *MedCOR* in cross section of immature shoot.

A) *MedPHB* showed expression throughout the cross section of the cotyledon and young shoot apical meristem. But stronger expression was seen in the vascular bundle, the tip and adaxial region of the young shoot apical meristem. B) *MedCOR* had the similar expression pattern as *MedPHB*. But the level of stain was much lighter with *MedCOR*. *MedREV* had the same expression as *MedCOR* (picture not shown here). C) Sense probe did not have any expression. This suggests that the expression pattern seen with *MedPHB*, *MedCOR* and *MedREV* is not artificial effect (Ad = adaxial, Ab = abaxial, Coty = cotyledon, VA = vascular bundle, YSAM = young shoot apical meristem).



CHAPTER 4

CONCLUSION

Recent substantial study on reconstructing the lineal relationships among land plants has provided plant developmental biologists with a solid foundation for the study of developmental evolution. Conversely, progress in elucidating the pathways that govern plant development in model species provides evolutionary biologists with a solid foundation to approach the same questions, from an alternative perspective.

The study of plant development has gone a long way from investigating plant organs and tissue types using light or scanning electron microscopy based approach (e.g. Esau, 1977; Steeves and Sussex, 1989; Endress, 1996; Kaplan, 2001), to the elucidation the molecular determinants of plant morphogenesis (reviewed in Leyser and Day, 2003).

Exploring the establishment of dorsoventral organ polarity in shoot lateral organs, has been the subject of recent intensive research activity plant development biologist. Four gene families that are integral to determine the ab/axial identity include the: the MYB and HD-ZIP III gene families, which have some members that promote adaxial identity (Waites, *et al.*, 1998; Bowman and Smyth, 1999) and the GARP and YABBY gene families, which promote an abaxial cell fate (Eshed, *et al.*, 2001; Kerseetter, *et al.*, 2001).

The focus of this thesis is the diversification of HD-ZIP III, a subfamily within HD-ZIP gene family. Investigated protein products of this gene family members, including REVOLUTA (REV), PHBBULOSA (PHB), and PHAVOLUTA (PHV), are transcription factors, with three conserved domains, a homeodomain (HD), and beta-leucine domain (ZIP), and a StAR-related lipid-transfer (START) domain, which is the target of both hormonal and miRNA regulation through (Ponting and Arvind, 1999; Schrick *et al.*, 2004).

Previous study on the phylogeny of HD-ZIP gene family had only limited sampling of HD-ZIP, class III (Sakakibara, 2001). The present study is the first to provide a detailed phylogeny of HD-ZIP, class III. Based on a series of phylogenetic analyses using a variety of methods, including alternative amino acid transition matrices and character-weighting schemes, we conclude that angiosperm homologs of *REV*, *PHB* and *COR* form three closely related clades.

In my study on the expression pattern on HDZ-III genes and selected members from YABBY gene family, I found that *MedFIL* and *TomFIL* showed an abaxial expression pattern in young leaf primordium, young petal primordium and young petal. They share the similar expression pattern with the previous results in *Arabidopsis* (Siegfried *et al.* 1999, Villaneuva *et al.* 1999, Meister *et al.* 2002). *TomCOR* and *TomREV* showed expression on the adaxial sides of investigated tissues. These results with surveyed YABBY and HD-ZIP III genes are largely consistent with findings based on several previous studies in *Arabidopsis* (Siegfried, *et al.*, 1999; Sawa, *et al.*, 1999; Kerstetter, *et al.*, 2001).

It was hypothesized that *REV*, *PHB* and *PHV* are expressed in the adaxial domains of lateral organs, apical meristem and vascular tissue in general. Also, previous study showed that *COR* had high expression in vascular tissues and diffused expression in flower meristem in *Arabidopsis* (Prigge, 2004; Green, 2005). My results are concordant with the prior statement, but are in contrast with the later assertion. Additionally, I found evidence that after a pair of ancient gene duplication events, the function of the promoters of paralogous HD-ZIP III genes has diverged.

TomCOR and *TomREV* both had adaxialized expression in young petal primordium. But *TomCOR* had much stronger expression than did *TomREV*. This suggests that *TomCOR* is an important factor in establishing adaxial polarity in petal development. *TomPHB* did not have any

expression in all the studied tissues. This suggests that perhaps *TomPHB* has lost the function that it may have once shared *TomCOR* and their most recent common ancestor. Three HD-ZIP III homologs in *Medicago* showed ubiquitous expression in mature and immature shoot apical meristems. In immature root, *MedCOR* is the only gene that did not have detectable expression.

My comparative gene expression study also provides evidence of clade-specific divergence in expression patterns after the ancient gene duplication events. In immature root of tomato, *TomCOR* is only one that had expression. But in immature roots of *medicago*, *MedCOR* is the only one that did not have expression. A similar phenomenon is apparent between *TomPHB* and *MedPHB*: *TomPHB* did not have any expression in all the tissues studied, but *MedPHB* had diffused expression in all tissue types.

These findings underscore the importance both of comparative studies of gene expression in multiple species and of carefully resolving the molecular ancestry of the genes that are targeted in studies of plant developmental evolution.

REFERENCES

- Bowman, J. L., and Smyth, D. R. 1999. *CRABS CLAW*, a gene that regulates carpel and nectary development in Arabidopsis, encodes a novel protein with zinc finger and helix-loop-helix domains. *Development* **126**: 2387-2396.
- Endress, P. K. 1996. *Diversity and evolutionary biology of tropical flowers*. New York: Cambridge University Press.
- Esau, K. 1977. *Anatomy of Seed Plants*, 2nd ed. Somerset, NJ: John Wiley & Sons.
- Eshed, Y., Baum, S.F., Perea, J.V., and Bowman, J. L. 2001. Establishment of polarity in lateral organs of plants. *Curr. Biol.* **11**: 1251-1260.
- Green, K. A., Prigge, M. J., Katzman, R. B. and Clarke, S. E. 2005. CORONA, a Member of the class III homeodomain leucine zipper gene family in Arabidopsis, Regulates stem cell specification and organogenesis. *The Plant Cell.* **17**: 691-704.
- Kaplan, D. R. 2001. Fundamental concepts of leaf morphology and morphogenesis: a contribution to the interpretation of molecular genetics. *Int. J. Pl. Sci.* **162**: 465-474.
- Kersetter, R. A., Bollman, K., Taylor, R.A., Bomblied, K., and Peothig, R.S. 2001. *KANADI* regulates organ polarity in Arabidopsis. *Nature* **411**: 706-709.
- Meister, R. J., Kotow, L. M., Gasser, C. S. 2002. *SUPERMAN* attenuates positive *INNER NO OUTER* autoregulation to maintain polar development of Arabidopsis ovule outer integuments. *Development.* **129**: 4281-4289.
- Ponting, C.P., and Aravind, L. 1999. START: A lipid-binding domain in StAR, HD-ZIP and signalling proteins. *Trends Biochem. Sci.* **24**: 130-132.

- Prigge, M. J., Otsuga, D., Alonso, J. M., Ecker J. R., Drews, G. N., Clark, S. E. 2005. Class III homeodomain-leucine zipper gene family members have overlapping, Antagonistic, and distinct roles in Arabidopsis development. *Plant Cell*. **17**: 61-76.
- Sakakibara, K., Nishiyama, T., Kato, M. and Hasebe, M. 2001. Isolation of homeodomain-leucine zipper genes from the moss *Physcomitrella patens* and the evolution of homeodomain-leucine zipper genes in land plants. *Mol. Biol. Evol.* **18**(4): 491-502.
- Sawa, S., Watanabe, K., Goto, K., Kanaya, E., Morita, E. H. and Okada, K. 1999. FILAMENTOUS FLOWER, a meristem and organ identity gene of Arabidopsis, encodes a protein with a zinc finger and HMG-related domains. *Genes Development*. **13**: 1079-1088.
- Siegfried, K. R., Eshed, Y., Baum, S. F., Otsuga, Drews, G. N., Bowman, J. L. 1999. Members of the YABBY gene family specify abaxial cell fate in Arabidopsis. *Development*. **126**: 4117-4128.
- Schrack, K., Nguyen, D., Karlowski, W. M., Mayer, K. F. 2004. START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors. *Genome Biology*. **5**: R41
- Steeves, T. A. and Sussex, I. M. 1989. *Patterns in Plant Development*, 2nd ed. Cambridge University Press, Cambridge.
- Villanueva, J. M., Broadhvest, J., Hauser, B. A., Meister, R. J., Schneitz, K. and Gasser, C. S. 1999. INNER NO OUTER regulates abaxial- adaxial patterning in Arabidopsis ovules. *Genes Development*. **13**: 3160-3169.
- Waites, R. and Hudson, A. 1995. *phantastica*: a gene required for dorsiventrality in

leaves of *Antirrhinum majus*. *Development* **121**, 2143-2154.