MOLECULAR EVOLUTION AND FUNCTIONAL CHARACTERIZATION OF THE *ARABIDOPSIS THALIANA* ACTIN-DEPOLYMERIZING FACTOR GENE FAMILY

by

EILEEN MELANIE ROY-ZOKAN

(Under the Direction of Richard B. Meagher)

ABSTRACT

The actin cytoskeleton is an essential component of eukaryotic cells and has had a profound impact on the evolution of multicellular organisms. The Actin-Depolymerizing Factor/Cofilin (ADF/CFL) gene family encode for a group of proteins that serve in modulating actin filament dynamics. The ADF/CFL family of proteins are implicated in a variety of cellular processes such as lipid and membrane metabolism, mitochondrial dependent apoptosis, chemotaxis, and the cytonuclear trafficking of actin into the nucleus. Within Arabidopsis, there are 11 paralogous ADF proteins that partition into four ancient phylogenetic subclasses that are differentially regulated. The goal of the work presented here was to investigate the evolutionary processes that shaped the diversification of the Arabidopsis ADF gene family and to determine the degree of functional divergence between family members. By estimating selective pressure across the plant ADF and animal CFL gene phylogenies, I show that there are differing patterns of codon evolution specific to subclasses of plant ADFs as well as to the different classes of animal ADF/CFL variants. Efforts to characterize the functional role of the single subclass IV Arabidopsis ADF variant, ADF6, revealed that ADF6 most likely represents

a non-essential member of the ADF gene family. Additionally, crossing adf9-1 with adf6-1 revealed that ADF6 acts in an antagonistic way with ADF9 to regulate flowering time. This role of ADF6 in flowering time is completely dependent on ADF9, as *adf6-1* flowers normally and shows no alteration in expression of any of the key flowering time genes. Finally, performing a series of suppression studies with reconstructed ancestral proteins revealed that the Arabidopsis ADF family members have rapidly diverged in function post-duplication.

INDEX WORDS: Actin, Actin-Depolymerizing Factor, ADF, Ancestral State Reconstruction, *Arabidopsis thaliana*, Cofilin, CFL, Cytoskeletal Dynamics, Molecular Evolution, Neofunctionalization, Subfunctionilization.

MOLECULAR EVOLUTION AND FUNCTIONAL CHARACTERIZATION OF THE *ARABIDOPSIS THALIANA* ACTIN-DEPOLYMERIZING FACTOR GENE FAMILY

by

EILEEN MELANIE ROY-ZOKAN

BS, Georgia State University, 2002

MS, College of Charleston, 2007

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2012

MOLECULAR EVOLUTION AND FUNCTIONAL CHARACTERIZATION OF THE *ARABIDOPSIS THALIANA* ACTIN-DEPOLYMERIZING FACTOR GENE FAMILY

by

EILEEN MELANIE ROY-ZOKAN

Major Professor: Richard B. Meagher

Committee: Kelly Dyer
Jacek Gaertig
Nancy Manley
Michael McEachern

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2012

DEDICATION

To my parents, Denis and Ijlal Roy, and my husband, Marcus Zokan: Thank you for never giving up on me and supporting me throughout my scientific endeavors. I am eternally grateful.

ACKNOWLEDGEMENTS

First, I would like to extend the most gracious thanks to my adviser, Dr. Rich Meagher, for guiding me through my graduate career and helping me develop as a scientist. I also would like to thank my advisory committee, Dr. Kelly Dyer, Dr. Jacek Gaertig, Dr. Nancy Manley, and Dr. Michael McEachern, for their feedback, guidance, and support. A most profound thanks goes to Libby McKinney and Dr. Kandasamy for helping me develop as a molecular geneticist and providing me with the training I need to tackle any scientific question. To Lori King and Kris Mussar, thank you for all your support and friendship. I would also like to thank Yolanda Lay for help with managing plant lines and providing much needed laughter. Many thanks to Mark Fisher, Jenna Oberstaller, Emily Peeden, Mike McKain, Jill Rahnert, and Marcus Zokan for being my scientific soundboard and giving critical scientific advice. Finally, I would like to thank my parents, Denis and Ijlal Roy, and my husband, Marcus Zokan, for their unending support through this process.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

Introduction and Literature Review

The Dynamic Nature of the Actin Cytoskeleton

The actin cytoskeleton is one of the most dynamic features in a eukaryotic cell and is involved in a variety of important cellular processes including organelle movement, determining cell polarity, cell division, and chromatin remodeling. As such, actin dynamics plays an essential role in plant and animal development. Actin exists in two states within a cell: a monomeric form (G-actin) and a polymeric filamentous form (F-actin) (Staiger, 2000). The polymerization of monomeric G-actin into the filamentous form occurs in a unidirectional manner, resulting in a polarized filament characterized by a barbed end and a pointed end (Schafer and Cooper, 1995). While polymerization and depolymerization can occur at both ends of the filament, polymerization occurs at a higher frequency on the barbed end and depolymerization occurs more frequently on the pointed end (Bamburg, 1999). The regulation of actin filament turnover, and the corresponding versatility of the actin cytoskeleton, is due in large part to the multitude of actin-binding proteins available. Capping proteins, monomer sequestering proteins, filament nucleating factors, and filament severing proteins all facilitate the various cellular activity of the actin cytoskeleton. One of the most versatile and complex actin binding protein is the actin-depolymerizing factor (ADF) and very closely related (CFL).

For simplicity, I may sometimes refer to this class of proteins as ADF/CFL. Higher plants and animals express multiple ADF/CFL protein variants encoded by gene families.

Actin-depolymerizing factors and cofilins are traditionally described as proteins that bind and sever actin monomers from F-actin filaments (Bamburg, 1999); however, this activity has been shown to be highly dependent upon concentration levels. When the cellular concentration of ADF/CFL exceeds that of F-actin, we find that ADF/CFL has the opposite function for what it was initially named for; they actually promote the nucleation of filaments rather than its severing (Andrianantoandro and Pollard, 2006). It is at lower cellular concentrations that ADF/CFL exhibits its better known activity of ADP-actin filament severing (Andrianantoandro and Pollard, 2006), which typically occurs at the pointed end of the actin filament (Bamburg, 1999). These contradictory effects on F-actin are a result of the impact ADF/CFL has on the elasticity of F-actin. Binding of ADF/CFL to actin filaments makes the filaments more elastic by implementing twists and bends within the filament (McCullough et al., 2008). At lower concentrations, ADF/CFL binds in small clusters along the filament, resulting in filaments that are only partially decorated by ADF/CFL. The boundaries between unbound and bound actin along these filaments set up local differences in mechanical strain that can promote severing (McCullough et al., 2008). As one can easily imagine, with increasing concentrations of ADF/CFL (and thus, increased binding along the filament), fewer boundaries will exist along the filament and therefore less severing will occur (**Figure 1.1**) (McCullough et al., 2008).

As more research has been conducted on the ADF/CFL gene family, it is becoming quite evident that the roles of this class of actin -binding proteins are far more

complex and diverse than initially described. We now know that their function within the eukaryotic cell extends well beyond actin filament turnover. ADF/CFL proteins have been implicated in cellular processes that range from membrane and lipid metabolism to mitochondrial dependent apoptosis (Chua et al., 2003; Klamt et al., 2009; Bernstein and Bamburg, 2010). These proteins have also been shown to be an essential component in cell motility (Tammana et al., 2008; Cooper and Schafer, 2000). A study conducted by Chan et al. (2009) revealed that ADF/CFL specifically targets Arp2/3 branch points on F-actin for filament severing at the leading edge of moving yeast cells. Not only does ADF/CFL reduce the number of available binding sites for Arp2/3, but the binding of ADF/CFL to the filament results in the propagation of conformational changes along the filament that rapidly dissociate other Arp2/3 complexes nearby (Chan et al., 2009). This interaction between ADF/CFL and Arp2/3 is what sets up the localized rapid transition between dense filaments to long, straight filaments that is crucial for shifts in motility (Chan et al., 2009).

The ADF/CFL protein variants have also been shown to be involved in intracellular trafficking. For example, yeast cofilin participates along with actin in the movement of vesicle-mediated endocytosis and it has been demonstrated that decreased levels of yeast cofilin can have adverse effects on the endocytic pathway (Ogreclak et al., 2007). The regulation of ADF/CFL is also crucial for proper trafficking from the trans-golgi network to other parts of the cell (i.e., plasma membrane, vacuole). By using either a constitutively active form of cofilin or through the mutation of Lim kinase 1 (LIMK; an important inhibitor of cofilin activity), Salvarezza et al. (2009) discovered that protein exit from the trans-golgi network was severely inhibited within canine kidney cells.

Interestingly, it was shown to be protein specific as only certain proteins were hindered from exiting the trans-golgi network (Salvarezza et al., 2009). The ADF/CFL protein variants have also been implicated in the cytonuclear trafficking of actin. Accumulating evidence strongly suggests that actin has an active role in nuclear processes such as transcription and chromatin remodeling (Bettinger et al., 2004). Actin itself lacks a nuclear localization signal (NLS); however, its binding partner ADF/CFL does possess an NLS. Thus, it has been hypothesized that it is the ADF/CFL proteins that are responsible in binding actin and transporting it into the nucleus (Ohta et al., 1989; Bettinger et al., 2004). The multifaceted roles of the ADF/CFL gene family members have resulted in some (e.g. Bernstein and Bamburg, 2010) to rightly dub these proteins as "functional nodes in cellular biology."

Actin-depolymerizing Factor/Cofilin Gene Family Evolution

As complex as the cellular roles of ADF/CFLs proteins are, the evolutionary history of the ADF/CFL gene family is proving to be equally intricate. The ADF/CFL proteins are relatively small (13-19 kDa) and are marked by a modest degree of sequence conservation between species. All members of the ADF/CFL gene family belong to a group of proteins that are classified by the presence of an actin-depolymerizing factor homology (ADF-H) domain. The ADF-H domain is an actin-binding domain that is made up of five β-strands that lie internal to four or more α-helices (Poukkula et al., 2011). Other members in this group include twinfilin, glia maturation factors, coactosin, and drebrin (Lappalainen et al., 1998; Poukkula et al., 2011). Out of all the ADF-H proteins, the ADF/CFL variants are the smallest, being comprised of only this single

protein motif (Lappalainen et al., 1998). The ADF-H family of proteins are all similar in protein folding structures but share much less sequence homology outside of the ADF-H domain (Poukkula et al., 2011).

Although vertebrates and chordates have gene families encoding ADF/CFL variants most invertebrate and single-cellular eukaryotic organisms possess only a single ADF/CFL variant (Maciver and Hussey, 2002; Bowman et al., 2000). Despite having only a single copy, these proteins are intricately regulated to facilitate their various cellular activities. For instance, the nematode *Caenorhabditis elegans* uses alternative splicing to differentially regulate its ADF/CFL protein, Unc60 (Bamburg, 1999). The resulting two variants, Unc60A and Unc60B, have fundamentally distinct developmental roles within *C. elegans* as Unc60A is involved in muscle cell differentiation and Unc60B in embryonic muscle development and that they bind actin differentially (Waterston et al., 1980; McKim et al., 1988; McKim et al., 1994). When we switch our focus over to vertebrate lineages, we find that the ADF/CFL gene family has undergone a considerable expansion in comparison to their invertebrate counterparts. Classically speaking, the vertebrate ADF/CLF variants are classified into two major classes: non-muscle (ADF/Destrin) and muscle-specific proteins (CFL2) (Bamburg, 1999). Mammalian lineages also have a third class of CFL proteins, CFL1, which is also a non-muscle form (Bamburg, 1999). As seen in the invertebrates, these classes of vertebrate ADF/CFL proteins also exhibit a high degree of spatial and temporal regulation. This regulation can even be seen on the single cellular level; for example, the two non-muscle isoforms within mammals are often expressed within the same cells but in different locations and at different developmental stages (Van Troysa et al., 2008).

This extensive partitioning of expression is also seen within ADF protein variants across plant lineages. As gene and genome duplications have occurred more frequently and more recently in the evolutionary history of higher plants than in animals, it comes as no surprise that the ADF gene family is far more expansive than that seen in animal lineages. Phylogenetic analyses reveal that the plant ADF protein variants group into four ancient subclasses that have been conserved in angiosperms for an estimated 250 million years (**Figure 1.1**) (Ruzicka et al., 2007; Feng et al., 2006; Maciver and Hussey, 2002). While the number of ADF variants varies between plant species, all angiosperm species possess at least one protein variant within each subclass (Ruzicka et al., 2007). Most of the information that has been gained on plant ADFs has come from research focusing on the model organism *Arabidopsis thaliana*. Using Arabidopsis, Ruzicka et al. (2007) demonstrated that each ADF subclass has its own unique expression profile. Within Arabidopsis, there are four ADF proteins that group within subclass I: AtADF1, AtADF2, AtADF3, and AtADF4. Subclass I ADF protein variants were found to be the most highly expressed members of the gene family with AtADF3 being the most strongly expressed of the four. These proteins exhibit their strongest expression in 10-day old seedlings, mature leaves, and mature flowers and are more moderately expressed in root, callus tissue, young seedlings, and in reproductive tissue with the exception of pollen (**Figure 1.3**) (Ruzicka et al., 2007). Plant ADF protein variants within subclass II partition into reproductive and vegetative subgroups, as designated by Ruzicka et al. (2007). Subgroup IIa contains AtADF7 and AtADF10 and subgroup IIb contains AtADF 8 and AtADF11 (**Figure 1.2**). The variants within subgroup IIa are expressed at moderate levels in pollen whereas subgroup IIb ADFs are expressed primarily in root

hairs and trichoblasts (**Figure 1.4**). What they share is expression in two cell types with rapid cell tip growth. Quantification of expression of the two subclass IIa ADF variants suggests that they are developmentally regulated as AtADF7 is expressed at higher levels than AtADF10 within mature flower tissue and ADF10 has stronger expression in immature flowers (**Figure 1.4a**). By fluorescently labeling these proteins with either CFP or YFP, Daher et al. (2011) demonstrated that this differential regulation could even be detected within earlier stages of a pollen grain. During the microspore stage within pollen, AtADF7 expression begins prior to AtADF10 expression AtADF7 where it localizes mainly to the nucleus. AtADF7 expression diminishes after this stage and does not reappear until the mature pollen stage when it localizes to the vegetative nucleus and along actin filaments in the elongating pollen tube. This nuclear localization is unique to AtADF7 as AtADF10 does not localize to the nucleus during any of the developmental stages within the male gametophyte. AtADF10 does not appear within pollen until the polarized microspore stage and is associated mainly with filamentous actin all through pollen development (Daher et al., 2011).

The *Arabidopsis* ADF variants within subclass III (AtADF5 and AtADF9) are primarily expressed in fast growing and differentiating tissue, although both genes are expressed at 5- to 10-fold lower levels than subclass I ADFs (**Figure 1.5**). AtADF5 is most strongly expressed in cotyledons, root vascular tissue, the root tip meristem, and emerging leaves but is also very weakly expressed in reproductive tissue, adult leaves, and mature roots (Ruzicka et al., 2007). AtADF9 is found in completely different tissue than that of AtADF5 as it is expressed quite strongly in callus tissue but has only weak expression in young seedlings, the root tip meristem, the root elongation zone, root

vascular tissue, the apical meristem, leaf edges, trichomes, the style, and in anthers and stamens (Ruzicka et al., 2007). The single subclass IV ADF, AtADF6, is the only ubiquitously expressed ADF (**Figure 1.6**): it is moderately to weakly expressed in all tissue types examined although expression seems to be strongest in the root vascular tissue, the root tip, and in cotyledons. AtADF6 is expressed at moderate levels in filaments, anthers, pollen, stigma, and vegetative epidermal tissue but is weakly expressed in carpel, sepal, and hypocotyls tissue (Ruzicka et al., 2007).

The tissue- and temporally-specific partitioning of expression patterns demonstrates that the ADF/CFL variants have, at minimum, subfunctionalized after duplication events. However, there is compelling evidence that indicates that the divergence of family members is not solely based on regulatory differences and gives credence to the idea that the ADF/CFL protein variants have also diverged significantly in function (neofunctionalized). Focusing first on the muscle (CFL2) and non-muscle specific variants, one would be tempted to argue that since both variants associate with filamentous actin that these proteins would still perform identical function but in different cellular compartments. To the contrary, recent finding have shown that these two proteins differ in the way they interact with F-actin, as CFL2 has a higher affinity for filaments than CFL1, suggesting that these two proteins may have differing roles in actin filament dynamics (Nakashima et al., 2005). Recent findings have also highlighted the fundamental biochemical differences between the two non-muscle specific mammalian protein variants, CFL1 and ADF/Destrin, which are highly suggestive of functional divergence (Bernstein and Bamburg, 2010). For one, ADF/Destrin was shown to be far more effective at sequestering actin monomers and in F-actin depolymerization, while

CFL1 is a more efficient nucleator (Yeoh et al., 2002; Estournes et al., 2007). Additionally, knocking out CFL1 is embryonically lethal whereas mice that are ADF/Destrin deficient show only minor developmental defects (Bernstein and Bamburg, 2010).

Although research into the function of plant ADFs is lagging in comparison to their mammalian homologs, the evidence for neofuncitonalization is equally as strong. Research into Arabidopsis subclass I protein variants have shown that they play a fundamental, and highly specialized, role in plant immunity. For instance, when looking at the role of the ADF proteins in eliciting an immune response to *Pseudomonas*, it was discovered that AtADF4 plays a vital role in providing a defense against a specific strain of *Pseudomonas* (Tian et al., 2004). Knocking down AtADF4 resulted in plants being highly susceptible to infection of the AvrPphB strain of *Pseudomonas syringae*; no other ADF was shown to have this response indicating that AtADF4 has evolved a targeted response to a very specific pathogen (Tian et al., 2004). Some of the most compelling evidence for neofunctionalization comes from work focusing on the Arabidopsis subclass III ADFs, AtADF5 and AtADF9. These two closely related paralogs are two of the least expressed ADF variants, but despite this provides for the most striking phenotypes when knocked out. What's really fascinating about this is that neither protein can suppress the phenotype of its paralog. For instance, AtADF9 is moderately expressed in the shoot apical meristem but exhibits low to no expression in most other tissues. Arabidopsis plants mutant in AtADF9 exhibited various aberrant phenotypes, but the most notable phenotype was that *adf9-1* mutants flowered early during long-day light cycles but not during short day light cycles (Burgos-Rivera, 2008). Investigations into the expression

levels of FLOWERING LOCUS C (FLC), a master repressor of the transition to flowering, revealed a significant decrease in transcript levels in the *adf9* mutants. Also, transcript levels of downstream flowering activators showed consistent expression profiles to that which would be expected with the downregulation of FLC. Most importantly, there is a significant increase in expression of CONSTANS, an activator of flowering. These findings are compelling as they suggest that ADF9 may play a role in the regulation of gene expression within the flowering time pathway. Whether it's a direct or indirect role remains to be seen.

Models of protein evolution

The complexity of the ADF/CFL gene family conjures up a number of compelling questions about protein evolution. The classical model of protein evolution predicts that the majority of duplicated genes undergo degeneration, with rare instances of neofunctionalization. According to this model, duplication events result in the relaxation of selective constraints on one of the paralogous genes (Ohno, 1970). With the relaxation of selection, this paralog is free to accumulate mutations, some of those being deleterious. Over time, this paralog will accumulate enough deleterious mutations to render it functionless (i.e., pseudogene). In rare instances, instead of accumulating deleterious mutations, the paralog could accumulate rare, beneficial mutations that allows for the paralog to take on a novel function. Since the probability of acquiring deleterious mutations is greater than the probability of acquiring beneficial/adaptive mutations, it was argued that the most common fate of a duplicate gene copy is to become pseudogenized (Ohno, 1970; Lynch and Conery, 2000). In 1999, Force et al. countered this argument by

putting forth their own model, which they coined the duplication-degeneration-complementation (DDC) model. The DDC model of protein evolution maintains that *"degenerative mutations in regulatory elements can increase rather than reduce the probability of duplicate gene preservation and the usual mechanism of duplicate gene preservation is the partitioning of ancestral functions (subfunctionalization) rather than the evolution of new functions"* (Force et al., 1999). A prime example of this concept, which the authors site as support for their model, is the *Hox1a* and *Hox1b* paralogs within *Mus musculus*. These duplicates have similar functions in segment identity, but their retention in the genome has been facilitated by the partitioning of their expression between different tissue segments (Force et al., 1999).

Following the publication of the DDC model, there were a number of studies that demonstrated that the process of subfunctionalization could also occur through means of adaptive mutations (see Conant and Wolfe, 2008). In this evolutionary process, coined the escape from adaptive conflict (EAC) model, paralogs from a recent gene duplication will be subjected to positive selection, acquiring adaptive mutations that facilitate the partitioning of ancestral protein function (Hughes, 2005; Des Marais and Rausher, 2008; Conant and Wolfe, 2008). A fourth model that has been recently proposed is the innovation, amplification, and divergence (IAD) model and is based on dosage selection on duplicated genes (Bergthorsson et al., 2007). This model centers on genes that have one primary function but simultaneously maintain a few accessory functions that are not selectively advantageous or disadvantageous. If an environmental change suddenly confers a benefit to one of these side functions, then a subsequent duplication event would result in a selectively advantageous amplification of function of this ancestral

gene. The two duplicates could then acquire adaptive or neutral mutations that would render each copy non-functional for one of the particular functions. Thus, the only copies remaining in the genome are a paralog that maintains the ancestral function and one that acquires the newly advantageous function. The difference in the IAD model versus the other models thus far discussed is that selection acts at all stages of the process (Bergthorsson et al., 2007; Conant and Wolfe, 2008).

Purpose of study

The ADF/CFL gene family is the ideal group of proteins to study protein evolution for the simple fact that there is preliminary evidence for particular variants having diverged and specialized in function while others have partitioned their ancestral function among new duplicates. While I have highlighted evidence that shows how these proteins have neofunctionalized, we do know from work done on Arabidopsis that neofunctionalization is not the ultimate the result for all duplication events within the ADF gene family. The best example of this comes from work conducted in the Meagher laboratory focusing on the Arabidopsis subclass II ADF variants. Knocking down AtADF11, a root specific ADF variant, results in retarded root hair development in Arabidopsis plants (unpublished data). The interesting part about this is that, unlike the case between AtADF9 and AtADF5, this root hair phenotype is successfully suppressed with the overexpression of AtADF11's closest paralog, AtADF8. In addition, root hair development was also restored with the overexpression of the pollen specific subclass II paralog, AtADF10 (unpublished data). It is quite clear that while there are Arabidopsis ADF family members that have become functionally distinct, such as the case with

AtADF9, there are also family members that have retained their ancestral state and have subdivided their functions between closely related members. Therefore, this group of proteins allows one to ask important evolutionary questions such as what drives proteins to gain novel functions, and how is this achieved over evolutionary time? These questions are where the focus of my dissertation lies, trying to decipher which ADF protein variants have neofunctionalized and to investigate how these divergences have occurred using the Arabidopsis ADF gene family as a model. My first strategy involved taking a computational approach to investigate the evolutionary forces responsible for the hypothesized diversification within the ADF/CFL gene family (Chapter 2). For this project, I expanded my scope of investigation beyond the Arabidopsis ADF gene family, and included the animal ADF/CFL gene family in the analyses. In Chapter 2, I describe how I used estimates of selective pressure to identify protein variants that experienced accelerated rates of evolution, and linked them to corresponding protein domains that may be involved in functional divergence.

One of the first projects that I tackled when I joined the Meagher laboratory was to characterize the functional role of the single subclass IV Arabidopsis ADF variant, AtADF6. What piqued my interest about AtADF6 was that while great strides had been made in understanding the function of ADF variants in the other three subclasses, the same couldn't be said for subclass IV. For instance, as described in the literature review, we know that members within subclass I have fundamental roles in immunity, that the subclass II variants are involved in tip elongation during development, and that the subclass III ADF variants may have specialized roles in gene expression regulation. However, virtually nothing is known about the protein variants within subclass IV. My

goals for this project (Chapter 3) were to detail the regulation of ADF6 expression throughout development and then characterize any morphological and developmental defects in plants deficient in AtADF6. I also attempt to identify any role AtADF6 may play in eliciting stress responses to various abiotic and biotic stressors. Understanding the functional role of AtADF6 was key to my ultimate goal of dissecting the evolution of this gene family.

One of the more ambitious undertakings in this dissertation was my attempt to identify the actual mutations and biochemical changes between family members that have resulted in functional diversification. In Chapter 4, I take advantage of extant yet ancient ADF protein sequences, as well as reconstructed ancestral proteins from the plant ADF phylogeny, to answer these questions. The results from all three research projects are summarized in Chapter 5 where I discuss how my results have impacted what we know about the evolutionary history of the ADF/CFL gene family. I also highlight future work that I believe would enhance our understanding of this complex group of proteins.

References Cited

Andrianantoandro, E., and Pollard, T.D. (2006) Mechanism of actin filament turnover by severing and nucleation at different concentrations of ADF/Cofilin. Mol. Cell. 24: 13–23.

Bamburg, J R. (1999) Proteins of the ADF/Cofilin Family: Essential Regulators of Actin Dynamics. Annual Review of Cellular Developmental Biology. 15:185-230.

Bergthorsson, U., Andersson, D.I., and Roth, J.R. (2007) Ohno's dilemma: Evolution of new genes under continuous selection. PNAS. 104(43):17004-17009.

Bernstein, B.W., and Bamburg, J.R. (2010) ADF/Cofilin: a functional node in cell biology. Trends in Cell Biology. 20(4): 187-195.

Bettinger, B.T., Gilbert, D.M., and Amberg, D.C. (2004) Actin up in the nucleus. Nature Reviews Molecular Cell Biology. 5: 410-415.

Bowman, G.D., Nodelman, I.M., Hong, Y., Chua, N.H., Lindberg, U., and Schutt, C.E.. (2000) A comparative structural analysis of the ADF/cofilin family. Proteins. 41:374-384.

Burgos-Rivera, B., Ruzicka, D.R., Deal, R.B., McKinney, E.C., King-Ried, L., and Meagher, R.B. (2008) Actin depolymerizing factor9 controls development and gene expression in *Arabidopsis*. Plant Mol Biol. 68(6):619-632.

Chan, C., Beltzner, C.C., and Pollard, T.D. (2009) Cofilin dissociates Arp2/3 complex and branches from actin filaments. Current Biology. 19(7): 537-545.

Chua, B.T., Volbracht, C., Tan, K.O., Li, R., Yu, V.C., and Li, P. (2003) Mitochondrial translocation of cofilin is an early step in apoptosis induction. Nat. Cell Biol. 5:1083–1089.

Conant, G.C., and Wolfe, K.H. (2008) Turning a hobby into a job: How duplicated genes find new functions. Nature Reviews Genetics. 9:938-950

Cooper, J.A., and Schafer, D.A. (2000) Control of actin assembly and disassembly at filament ends. Current Opinion in Cell Biology. 12:97–103.

Daher, F.B., van Oostende, C., and Geitmann, A. (2011) Spatial and temporal expression of actin depolymerizing factors ADF7 and ADF10 during male gametophyte development in *Arabidopsis thaliana*. Plant Cell Physiol. 52(7): 1177–1192.

Des Marais, D.L., and Rausher, M.D. (2008) Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature. 454:762-765.

Estornes, Y., Gay, F., Gevrey, J.C., Navoizat, S., Nejjari, M., Scoazec, J.Y., Chayvialle, J.A., Saurin, J.C., and Abello, J. (2007) Differential involvement of destrin and cofilin-1 in the control of invasive properties of Isreco1 human colon cancer cells. International Journal of Cancer. 121:2162–2171.

Feng, Y., Liu, Q., and Xue, Q. (2006) Comparative study of rice and Arabidopsis Actin-depolymerizing factors gene families. Journal of Plant Physiology 163:69-79

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y., and Postlethwait, J. (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 151:1531-1545.

Hughes, A.L. (2005) Gene duplication and the origin of novel proteins. PNAS. 102:8791-8792.

Klamt, F., Zdanov, S., Levine, R.L., Pariser, A., Zhang, Y, Zhang, B., Veenstra, T.D., and Shacter, E. (2009) Oxidant-induced apoptosis is mediated by oxidation of the actin-regulatory protein cofilin. Nat. Cell Biol. 11:1241–1246

Lappalainen, P., Kessels, M.M., Jamie, M., Cope, T.V., and Drubin, D.G. (1998) The ADF homology (ADF-H) domain: A highly exploited actin-binding module. Molecular Biology of the Cell. 9:1951–1959.

Lynch M., and Conery, M.S. (2000) The evolutionary fate and consequences of duplicate genes. Science. 290(5494):1151-1155.

Maciver, S.K., and Hussey, P.J. (2002) The ADF/cofilin family: actin-remodeling proteins. Genome Biol. 3(5):3007.1-3001.12

McCullough, B.R., Blanchoin, L., Martiel, J.L., and De La Cruz, E.M. (2008) Cofilin increases the bending flexibility of actin filaments: Implications for severing and cell mechanics. Journal of Molecular Biology. 381(3): 550-558

McKim, K.S., Heschl, M.F.P., Rosenbluth, R.E., and Baillie, D.L. (1988) Genetic organization of.the *unc-60* region in *Caenorhabditus elegans.* Genetics. 118:49-59.

McKim, K. M., Matheson, C., Marra, M. A., Wakarchuk, M. F., and Baillie, D. L. (1994) The *Caenorhabditus elegans unc-60* gene encodes proteins homologous to a family of actin binding proteins. *Mot. Gen. Genet.* 242:346-357.

Nakashima, K., Sato, N., Nakagaki, T., Abe, H., Ono, S., and Obinata, T. (2005) Two mouse cofilin isoforms, muscle-type (MCF) and non-muscle type (NMCF), interact with F-actin with different efficiencies. J. Biochem. 138:519–526.

Ohno, S. (1970) Evolution by Gene Duplication. Springer-Vergag, Heidelberg, Germany.

Ohta, Y., Nishida, E., Sakai, H., and Miyamoto, E. (1989) Dephosphorylation of cofilin accompanies heat shock-induced nuclear accumulation of cofilin. The Journal of Biological Chemistry. 264:16143-16148.

Okreglak, V., and Drubin, D.G. (2007) Cofilin recruitment and function during actin-mediated endocytosis dictated by actin nucleotide state. The Journal of Cell Biology. 178(7):1251-1264

Poukkula, M., Kremneva, E., Serlachius, M., and Lappalainen, P. (2011) Actin-Depolymerizing Factor Homology Domain: A conserved fold performing diverse roles in cytoskeletal dynamics. Cytoskeleton. 68:471-490.

Ruzicka, D.R., Kandasamy, M.K., McKinney, E.C., Burgos-Rivera, B., and Meagher, R B. (2007) The ancient subclasses of Arabidopsis ACTIN DEPOLYMERIZING FACTOR genes exhibit novel and differential expression. The Plant Journal. 52: 460-472.

Salvarezza, S.B., Deborde, S., Schreiner, R., Campagne, F., Kessels, M.M., Qualmann, B., Caceres, A., Kreitzer, G., and Rodriguez-Boulan E. 2009. LIM Kinase 1 and Cofilin Regulate Actin Filament Population Required for Dynamin-dependent Apical Carrier Fission from the Trans-Golgi Network. Molecular Biology of the Cell. 20(1): 438-451.

Schafer, D.A., and Cooper, J.A. (1995) Control of actin assembly at filament ends. Annu. Rev. Cell Dev. Bio. 11:497-518.

Staiger, C.J. (2000) Signaling to the actin cytoskeleton in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 51:257–88 .

Tammana, T.V.S., Sahasrabuddhe, A.A., Mitra, K., Bajpai, V.K., and Gupta, C.M. (2008) Actin-depolymerizing factor, ADF/cofilin, is essentially required in assembly of Leishmania flagellum. Molecular Microbiology. 70(4):837-852

Tian, M.Y., Chaudhry, F., Ruzicka, D.R., Meagher, R.B., Staiger, C.J., and Day, B. (2009) *Arabidopsis* Actin-Depolymerizing Factor AtADF4 Mediates Defense Signal Transduction Triggered by the *Pseudomonas syringae* Effector AvrPphB. Plant Physiology. 150(2): 815-824.

Van Troysa, M., Huycka, L., Leymana, S., Dhaesea, S., Vandekerkhovea, J., and Ampea, C. (2008) Ins and outs of ADF/cofilin activity and regulation. European Journal of Cell Biology 87:649–667

Waterston, R.H., Thomson, J.N., and Brenner, S. (1980) Mutants with altered muscle structure in *Caenorhabditus elegans.* Dev. Biol. 77:271-302.

Yeoh, S., Pope, B., Mannherz, H.G., and Weeds, A. (2002) Determining the differences in actin binding by human ADF and cofilin. J. Mol. Biol. 315:911–925

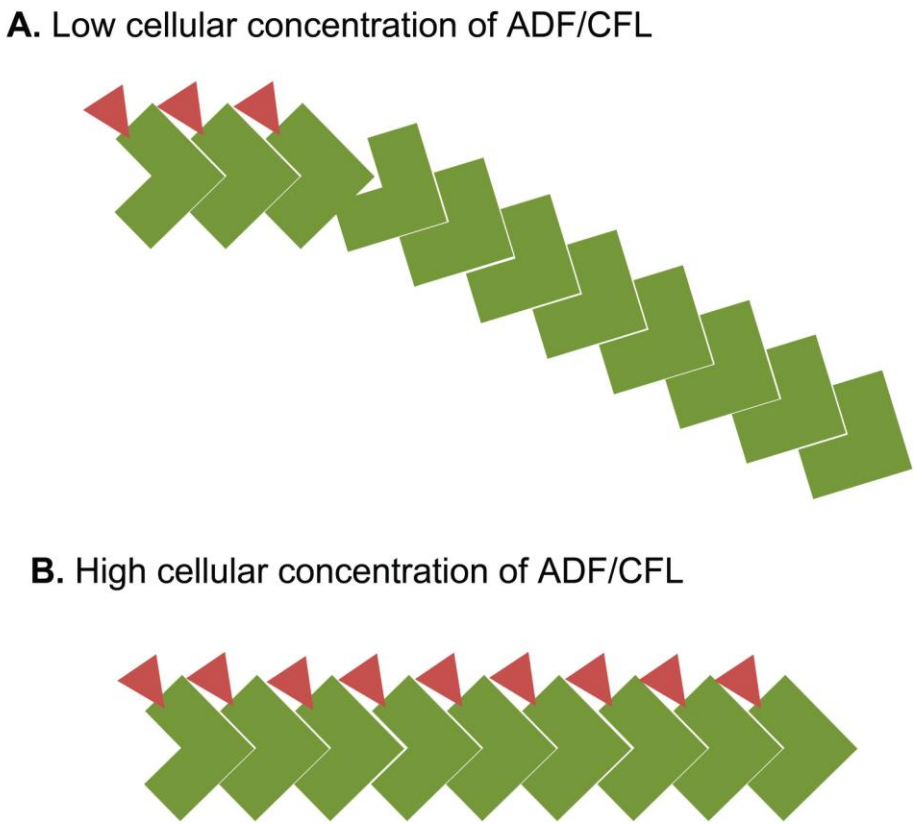


Figure 1.1: ADF/CFL proteins affect actin filament turnover in a concentration dependent manner. Green geometric shapes represent actin and pink triangles represent ADF/CFL proteins. The binding of ADF/CFL to F-actin results in increased flexibility in the actin filament. **(A)** When cellular concentrations of ADF/CFL are low, this increased flexibility results in a bend in the filament which occurs between ADF/CFL bound and unbound F-actin. This bend causes a destabilization in the interaction between actin subunits within the filament and ultimately results in severing. **(B)** At high cellular concentrations, ADF/CFL binds cooperatively along the filament, causing the filament to become stabilized. This stabilization arises from the reduced number of bound to unbound boundaries between actin subunits along the filament. Without these boundaries, the torsional strain on the filament is not established.

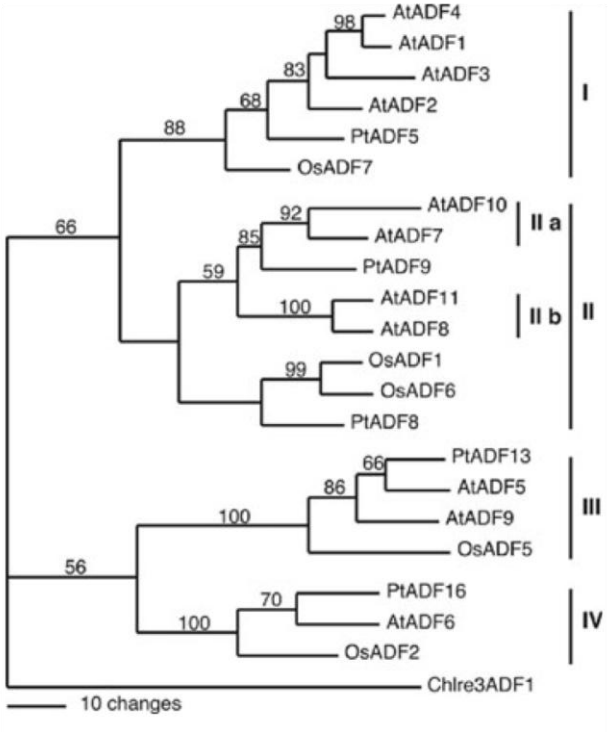


Figure 1.2: Parsimony phylogenetic analysis of the plant ADF gene family as published by Ruzicka et al. (2007). The plant ADF gene family partitions into four ancient subclasses that have been conserved for an estimated 250 million years. The eleven Arabidopsis ADF variants are underlined in red. Species designations are as follows: At, Arabidopsis thaliana; Pt, Populus trichocarpa; Os, Oryza sativa (japonica); Chlre, Chlamydomonas reinhardtii.

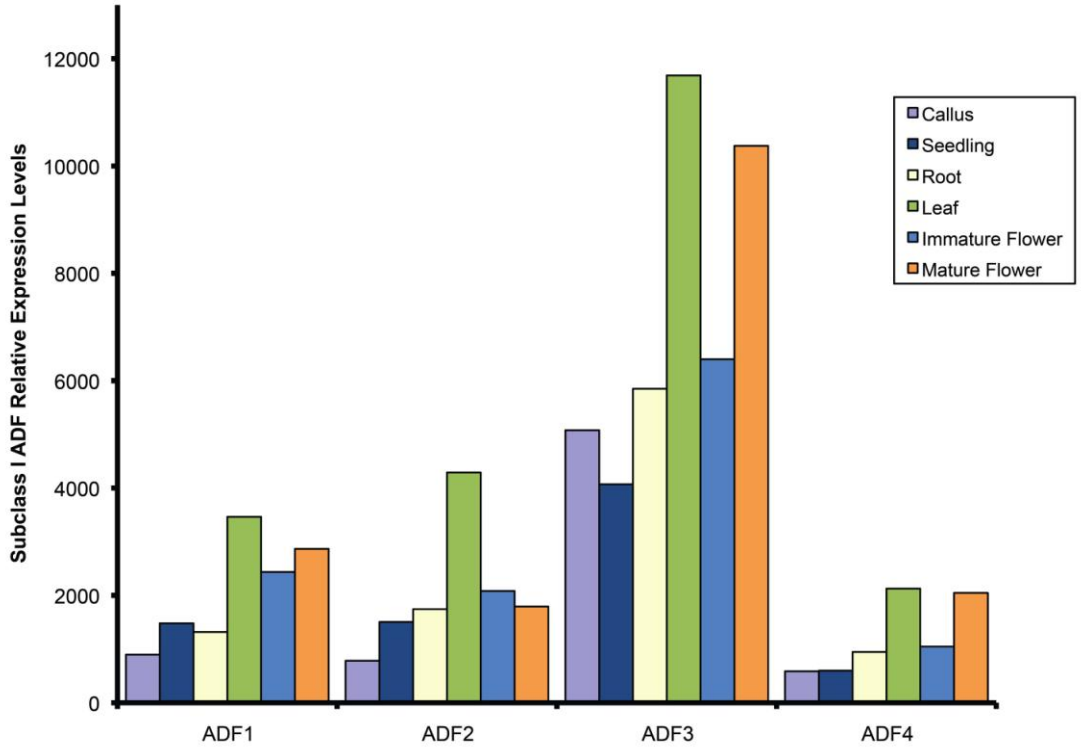


Figure 1.3: Expression patterns of the four Arabidopsis subclass I ADF variants as quantified by qRT-PCR (Ruzicka et al., 2007). Subclass I ADF variants are expressed at moderate to high levels in all tissue examined, except for mature pollen. Out of all four variants, ADF3 is the most strongly expressed protein. Tissue types and developmental stages examined include callus, 10 day old seedling, root, mature rosette leaf, immature flower, and mature flower.

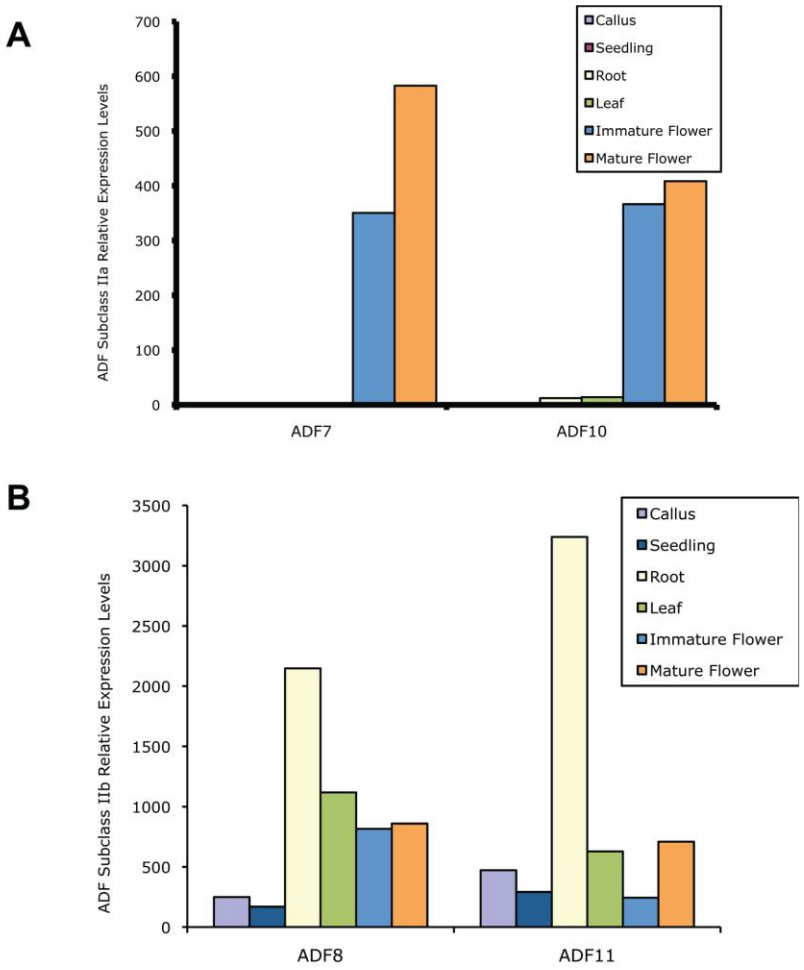


Figure 1.4: Expression patterns of the four Arabidopsis subclass II ADF variants as quantified by qRT-PCR (Ruzicka et al., 2007). Tissue types and developmental stages examined included callus, 10 day old seedling, root, mature rosette leaf, immature flower, and mature flower. (A) Expression patterns of the two subgroup IIa protein variants, ADF7 and ADF10. While qRT-PCR analysis revealed that these two variants are expressed in reproductive tissue, GUS staining revealed that expression is limited to pollen grains only. (B) Expression patterns for subgroup IIb protein variants, ADF8 and ADF11. ADF8 and ADF11 were strongest in root tissue, as revealed by qRT-PCR analysis. Further analysis using GUS staining showed that expression was completely restricted to root hairs and root epidermal cell files.

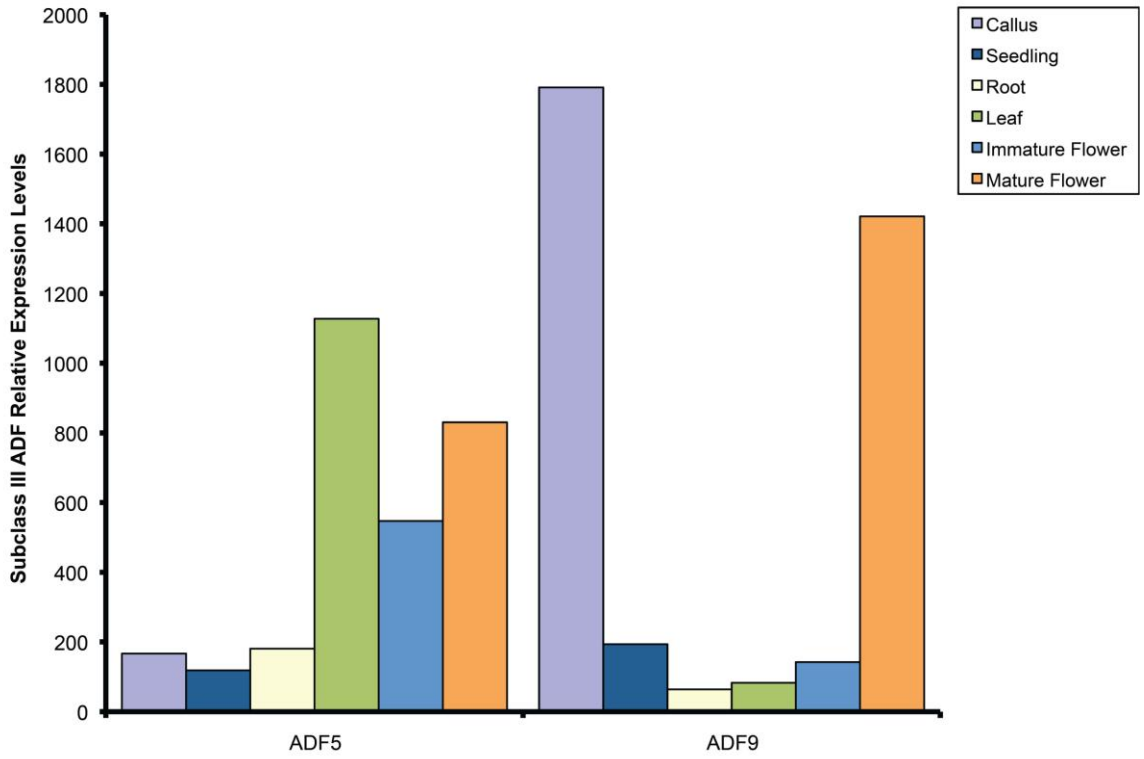


Figure 1.5: Expression patterns of the two Arabidopsis subclass III ADF variants as quantified by qRT-PCR (Ruzicka et al., 2007). Subclass III ADF(ADF5 and ADF9) variants are expressed at the lowest levels in comparison to the ADF protein variants in the other three subclasses. ADF5 had the highest expression in leaf and reproductive tissue whereas ADF9 exhibited its highest expression levels in callus tissue as well as mature flowers. While the qRT-PCR data suggests these protein variants have a ubiquitous expression pattern, GUS staining revealed that expression is highly localized in each tissue type examined. For instance, although ADF5 is expressed in root, it is localized only at the root tip.

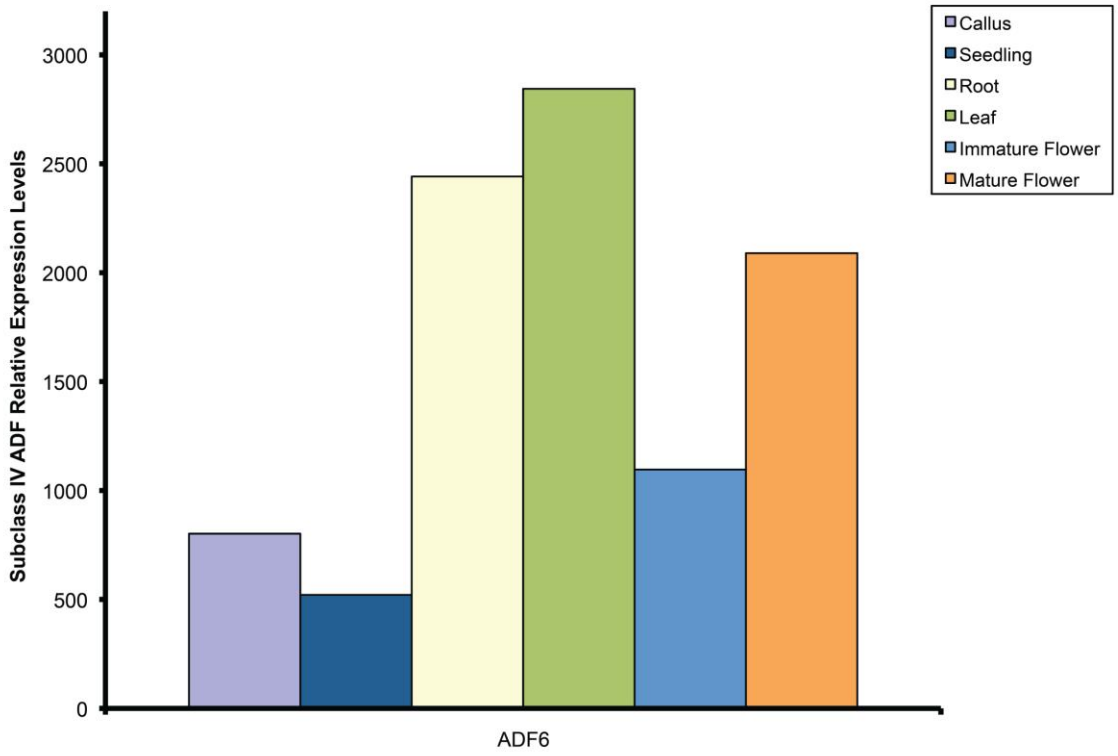


Figure 1.6: Expression pattern of the single Arabidopsis subclass IV ADF variant as quantified by qRT-PCR (Ruzicka et al., 2007). ADF6 is the sole Arabidopsis representative in subclass IV and is the only ubiquitously expressed ADF, as demonstrated by qRT-PCR analysis and confirmed through GUS-staining.

CHAPTER 2

Subclass specific patterns of codon evolution in the ACTIN-DEPOLYMERIZING FACTOR/COFILIN (ADF/CFL) GENE FAMILY[1]

[1]Eileen M. Roy-Zokan and Richard B. Meagher. To be submitted to the Journal of Molecular Evolution

Abstract

Genes involved in adaptive evolution, such as stress response genes, may be strongly influenced by positive "Darwinian" selection. However, little is known about the role selection has played on the evolution of proteins involved in mediating these responses, such as actin and actin-binding proteins. The actin-depolymerizing factor/cofilin (ADF/CFL) gene family encode a group of relatively small proteins that serve in a variety of cellular processes. Once known strictly as modulators of actin filament dynamics, recent research has demonstrated that these proteins are involved in a variety of cellular processes, from signal transduction to the cytonuclear trafficking of actin. In both plant and animal lineages, the encoded protein variants have partitioned expression pattern across tissue types and developmental stages with strong evidence for neofunctionalization between family members. The goal of this study is to take a comparative and computational approach to investigate the evolutionary forces responsible for the diversification within the ADF/CFL gene family. Estimating the ratio of rates of non-synonymous to synonymous mutations (dN/dS) across phylogenetic lineages revealed that the majority of ADF/CFL codon positions are under tight selective constraint with rare, episodic events of accelerated rates of evolution. Codon positions that had the highest probability of accelerated rates of evolution were subclass specific for both plants and animals and selection patterns differed among taxonomic groups. Mapping these sites onto predicted protein structures gives strong support to the view that particular amino acid residues may have played a fundamental role in facilitating the functional diversification of subsets of ADF/CFL proteins.

Introduction

The importance of the actin cytoskeleton within eukaryotes, both for development and cellular function, has been well documented (Kandasamy et al., 2012). The actin cytoskeleton is one of the most dynamic features in a eukaryotic cell, being involved in a number of important cellular processes as diverse as organelle movement, exo- and endocytosis, cell division, establishment of cell polarity, nuclear trafficking, and chromatin remodeling. There are a variety of actin binding proteins that facilitate the dynamic nature of the F-actin cytoskeleton, one of which is the actin-depolymerizing factor (ADF)/Cofilin (CFL) gene family. ADF/CFLs are encoded by a gene family in higher plants and animals and comprise a group of relatively small proteins of about 13 – 19 kDa in size (Bamburg, 1999). All members of the ADF/CFL gene family possess a highly conserved protein motif known as the actin-depolymerizing factor homology (ADF-H) domain (Lappalainen et al. 1998; Poukkula et al. 2011). The ADF-H domain is characterized by the presence of five β-strands that lie internal to four or more α-helices (Poukkula et al. 2011). While other classes of actin-binding proteins, such as twinfilin and coactosin, possess multiple copies of this motif, what sets the ADF/CFL gene family members apart is that they only contain a single ADF-H motif (Lappalainen et al. 1998). The ADF-H domain possessing actin-binding proteins share little in sequence homology; however, they have homologous protein folded structures within this ADF-H domain suggesting conservation of an ancient actin binding domain among gene families (Poukkula et al. 2011).

The members of the ADF/CFL family are traditionally described as proteins that, at low cellular concentrations, bind and sever actin monomers from F-actin filaments

(Bamburg 1999; Bernstein and Bamburg, 2010). However, at intermediate concentrations, ADF/CFLs will stabilize filaments and at high concentrations will promote the nucleation of filaments rather than severing (Andrianantoandro and Pollard 2006). It is at lower cellular concentrations that ADF/cofilin exhibits its better-known function of ADP-actin filament severing (Andrianantoandro and Pollard 2006). As more research is conducted on the ADF/cofilin gene family, it is becoming quite evident that the story of these actin-binding proteins is far more complex than initially described. Their role within the eukaryotic cell extends well beyond their actin depolymerizing, stabilizing, and nucleating activities. For instance, ADF/CFL have been implicated as an essential component in cell motility and intracellular trafficking (Chan et al. 2009; Salvarezza et al. 2009; Okreglak and Drubin 2007). This role in intracellular trafficking contributes the metastatic spread of breast cancer cells (Wang et al. 2007). Cofilin/ADF also participates in the cytonuclear trafficking of actin. Actin has an active role in nuclear processes such as transcription and chromatin remodeling (Bettinger et al. 2004). Actin lacks a nuclear localization sequence (NLS) but most ADF/CFL protein variants do possess a potential NLS. It is hypothesized that it is the ADF/CFL protein variants that are responsible in binding to actin and transporting it into the nucleus, and this activity is enhanced in response to stress (Ohta et al. 1989; Bettinger et al. 2004). The multifaceted and important roles of the ADF/cofilin gene family members have in controlling F-actin dynamics resulted in Bernstein and Bamburg (2010) to respectfully characterize these proteins as “a functional node in cellular biology.”

As complex as the cellular role of this gene family is, the evolutionary history of this gene family is proving to be equally compelling. Across invertebrate lineages,

species typically possess only a single ADF/CFL variant; however, these proteins are intricately regulated. For example, the nematode *Caenorhabditis elegans* differentially regulates its ADF/CFL protein, Unc60, through alternative splicing (Bamburg, 1999). Alternative splicing yields two variants of the Unc60 protein, Unc60A and Unc60B, which have distinct fundamental developmental roles within *C. elegans* (Waterston et al. 1980; McKim et al. 1988; McKim et al. 1994). The ADF/CFL gene family underwent an expansion in vertebrate lineages, with three classes of protein variants being found: CFL1, CFL2, and ADF (also referred to as Destrin). However, it should be noted that not all vertebrate species possess an ADF/CFL variant from each class. The CFL2 class of ADF/CFLs is the muscle specific CFL protein variant while CFL1 and ADF/Destrin are non-muscle forms. All three classes of ADF/CFL proteins are spatially and temporally regulated (Bamburg 1999) and within mammals, this regulation can even be seen within a single cell (Van Troysa et al., 2008).

In contrast, the plant ADF gene family is much larger than that seen in animals. Previous phylogenetic analyses have shown that plant ADF protein variants group into four ancient subclasses (Maciver and Hussey 2002; Feng et al. 2006; Ruzicka et al. 2007) that have been conserved in angiosperms for approximately 250 million years. While the number of ADF variants varies between species, all angiosperm species possess at least one protein variant within each of the four subclasses. The partitioning of expression patterns is even more obvious for ADF protein variants across plant lineages. Studies in Arabidopsis show that each subclass has its own unique expression profile (Ruzicka et al., 2007). Subclass I ADF protein variants are the most highly expressed members of the gene family; they are expressed in almost all tissues and organs except for a few

reproductive tissues such as mature pollen. Subclass II protein variants are involved primarily in tip elongation in pollen tubes and root tips; as such, their expression is confined to root and reproductive tissues (Ruzicka et al. 2007; Daher et al. 2011). The protein variants in subclass III are expressed at a minimal level, and are found in tissue undergoing rapid growth such as meristematic tissue. By contrast, subclass IV variants are ubiquitously expressed throughout the plant at moderate levels (Ruzicka et al., 2007).

The tissue- and temporal-specific partitioning of expression patterns suggests that ADF/CFL protein variants have at least subfunctionalized during their evolutionary history. But to what degree have the actual protein variants diverged in function from one another? The evidence garnered so far focusing on mammalian CFL and plant ADF protein variants gives strong support for functional diversification among gene family members. For one, CFL and ADF/Destrin have fundamental biochemical differences that are highly suggestive of functional divergence (Bernstein and Bamburg, 2010). ADF/Destrin has been shown to be greater at monomer sequestering and actin depolymerization while CFL1 is a more efficient nucleator (Yeoh et al., 2002; Estournes et al., 2007; Bernstein and Bamburg, 2010). Knocking out CFL results in embryonic lethality in mice whereas ADF/Destrin knockout mice show only minor developmental defects (Bernstein and Bamburg, 2010). Additionally, muscle (CFL2) and non-muscle (CFL1) cofilins have been found to interact differently with F-actin, suggesting that these two protein variants may have differing roles in actin filament dynamics (Nakashima et al., 2005). The evidence for neofunctionalization is equally as strong for the plant ADF gene family. The closely related subclass III Arabidopsis ADF paralogs, ADF5 and ADF9, have contrasting and unique phenotypes that cannot be suppressed through the

exogenous expression of its respective paralog. For instance, the early flowering phenotype that is exhibited in AtADF9 mutants cannot be suppressed through the overexpression of ADF5 (Burgos-Riviera et al., 2009). This suggests that there must be key differences between these closely related protein variants, whether it is solely structural and biochemical, or whether it is due to different binding partners cannot be concluded at this point. There also have been a few studies that have highlighted the varying and specialized roles of subclass I protein variants in pathogen defense. Arabidopsis plants that are defective in ADF4 are significantly more susceptible to *Pseudomonas* infection than wildtype plants (Tian et al., 2004). Interestingly, the authors demonstrated that this role in *Pseudomonas* immunity is not conserved across all subclass I Arabidopsis variants and is unique to AtADF4. Along those same lines, it has been demonstrated that AtADF2 plays a crucial role in the establishment of nematode infection in Arabidopsis (Clement et al., 2009). So while investigation into plant ADFs has only just begun, it is obvious that the diversity in cellular function of these proteins is great.

Taken together, the evidence for neofunctionalization in the ADF/CFL gene family is compelling. But exactly which protein variants have undergone neofuctionalization? And which regions of the proteins are implicated in facilitating the functional diversification and subsequent versatility of this gene family? To answer these and other questions about neofunctionalization of the ADF/CFL gene family, we have investigated the evolutionary processes that have maintained these protein variants within their respective genomes. Estimations of selection can be useful when investigating genes that have undergone duplication, as duplicates that have neofunctionalized may experience accelerated rates of evolution (Yang, 2007). This method of estimating

selective pressure across a gene phylogeny has been used extensively to decipher to molecular evolution of important gene families such as the triplicate alpha-globin genes in rodents and MADS-box transcription factors in plants (Storz et al., 2008; Shan et al., 2009). Here, we describe a more detailed phylogenetic analysis of the animal ADF/CFL and plant ADF gene families, separately. Next, we show that even though the majority of the ADF/CFL protein variants are under tight selective constraint, the analyses revealed that there are episodes of accelerated rates of evolution across the gene family. More specifically, we find that there are subclass- and class-specific patterns of codon evolution occurring in the plant and animal ADF/CFL gene families, respectively, indicating that these regions have been particularly important in the functional diversification of these proteins.

Materials and Methods:

Sequence Acquisition and Alignment

The nucleotide coding sequences (CDS) for ADF and CFL proteins were acquired through NCBI and TAIR online databases (Supplementary Material, Table 2.1). The following plant species were chosen for analysis: *Physcomitrella patens, Selaginella moellendorffii, Zea mays, Oryza sativa japonica, Vitis vinifera, Mimulus guttatus, Populus trichocarpa,* and *Arabidopsis thaliana* (see Supplementary Table 2.2 for divergence times and classifications). The addition of more plant species along with the addition of *Physcomitrella patens* as an outgroup was done in order to obtain a phylogeny with greater resolution than that put forth by Ruzicka et al. (2007). This particular range

of species was chosen in order to have a broad representation of both monocot and dicot lineages (Supplementary Table 2.2). This more refined resolution is important for estimating selection on the ADF variants (see below). For the remainder of the eukaryotes, seven vertebrate, four invertebrate, and one protist species were chosen as representatives: *Monosiga brevicolis*, *Ciona intestinalis*, *Branchiostoma floridae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Salmo salar*, *Danio rerio*, *Xenopus tropicalis*, *Gallus gallus*, *Sus scrofa*, *Mus musculus*, and *Homo sapiens* (see Supplementary Table 2.3 for divergence times).

The CDS of the ADF/CFL variants were translated to amino acid sequences for alignment to avoid aberrant insertions into the coding sequence during alignment. Sequences were aligned using ClustalW (Larkin et al. 2007) program within MEGA 5.05 (Tamura et al. 2011). The BLOSUM protein weight matrix was used for aligning the sequences; once aligned, the translated sequences were converted back to nucleotide sequences and the alignment was manually adjusted using Se-Al v2.0a11. The final alignment used for phylogenetic and selection analyses corroborated with that published by Bowman et al (2000) with vertebrate specific insertions being included. Sequence divergence between protein variants were calculated using nucleotide substitution p-distance in MEGA 5.05.

Phylogenetic Analyses of Plant and Animal ADF/CFLs

Three separate phylogenetic analyses were performed for this study. The first ADF/CFL tree constructed included both plant and animal ADF/CFL protein variants in order to assess the overall relationship between plant and animal proteins. The next two

phylogenetic analyses focused on the plant and animal ADF/CFL families separately. For these two analyses, the number of species was increased in order to take a much more inclusive look at the evolutionary relationships of the ADF/CFL proteins across plant and animal lineages. Due to the high degree of sequence divergence, the analysis containing both plant and animal ADF/CFL proteins (herein referred to as the across kingdom dataset) were performed using amino acid sequences. The program MrModelTest (Nylander 2004)) was used to determine the best-fit substitution model for the animal and plant ADF/CFL datasets. For the across kingdom dataset, a mixed amino acid model run was performed using MrBayes (Ronquist and Huelsenbeck 2003). The model with the highest posterior probability was chosen to be used in analyzing the across kingdom dataset. Bayesian phylogenetic analyses were conducted on all three phylogenies using the program MrBayes 3.1. A maximum-likelihood and neighbor-joining phylogenetic analyses were also performed in Mega 5.05 to check for congruence. For the plant ADF dataset, the general-time-reversible (GTR) model of nucleotide substitution (Yang 1994; Zharkikh 1994) was used with a gamma distribution. Two independent computational runs were implemented for 10,000,000 generations with a tree sample frequency of every 100 generations. The resultant tree was compiled after the first 25% of sampled trees were discarded. For the animal ADF/CFL dataset, the GTR model of nucleotide substitution was also used for analysis but with a gamma distribution that allows a proportion of invariable sites. The two independent computational runs were implemented for 5,000,000 generations with a tree sample frequency of every 100 generations. The final tree was compiled after a burnin of 12,500 trees (25%). As mentioned above, the across kingdom phylogeny was built using amino acid instead of

nucleotide sequences. A preliminary mixed model run in MrBayes indicated that the Wag model of protein evolution was the best-fit model for the data. Therefore, the Wag model was implemented on the across kingdom dataset for 2,000,000 generations with a sample frequency of 100 generations. Again, the resultant tree was compiled after the first 5,000 sampled trees were discarded. All three trees were converted to newick format using the ape library (Paradis et al. 2004) in the R64 statistical package and subsequently visualized using Mega5.05. All trees were rooted post-analysis during the visualization process.

Selection Analyses

To investigate the selective pressure across the gene phylogeny, we estimated the rate of nonsynonymous mutations to the rate of synonymous mutations both across the protein sequence as well as across the phylogeny. The ratio of the rate of nonsynonymous mutation to the rate of synonymous mutations, commonly referred to as dN/dS or ω, has been widely used to investigate the molecular evolution of gene duplicates. Typically, when the rate of nonsynonymous mutations is less than the rate of synonymous mutations ($\omega < 1$), this is said to be a signature of negative/purifying selection (i.e., conserved residues). If a protein is evolving neutrally, then the rate of nonsynonymous mutation is equal to the rate of synonymous mutations ($\omega = 1$). Positive/diversifying selection (i.e., residues with accelerated rates of evolution) leaves a signature of a higher rate of nonsynonymous mutations than synonymous mutations ($\omega > 1$). The main goal of this paper is not to test whether gene family members have evolved through positive selection, but rather to identify protein variants and/or regions and

residues within the protein that have experienced accelerated rates of evolution (i.e., relaxed selective constraints). To do this, we utilized the program fitModeL (Guindon et al. 2004) to estimate the degree of selective pressure exerted on the ADF/CFL protein variants. As with the program PAML (Yang 2007), fitModeL is a maximum-likelihood based program that enables the user to investigate patterns of evolution. However, it has one major advantage over PAML in that it allows for selection to vary across phylogenetic lineages as well as between amino acid sites without making *a priori* assignments (Guindon et al. 2004). A series of nested models of codon evolution were performed to determine the best-fit model for each dataset (plant ADF and animal ADF/CFL, separately). In total, four models of codon evolution were tested: M0, M3, M3+S1, and M3+S2. The M0 model represents the null model and assumes a single ω value for every amino acid position across all phylogenetic lineages. This null model was compared to the discrete model of codon evolution (M3), which allows selective pressure to vary across the protein sequence. The M3 model assumes three selection rate categories (ω_1, ω_2, and ω_3) with the only restriction being that $\omega_1 < \omega_2 < \omega_3$ (Yang 2007). Even though the M3 model of codon evolution allows for selection to vary across the protein sequence, the estimates are held constant across phylogenetic lineages. In other words, every single protein variant in the phylogeny shares the same estimated ω category for each codon position. The next two models (M3+S1 and M3+S2) allows for selection to vary across phylogenetic lineages. For both the M3+S1 and M3+S2 models, each amino acid/codon position for all proteins in the phylogeny is assigned a probability of being in a particular rate category (ω_1 vs. ω_2 vs. ω_3). The M3+S1 model tests whether there is unbiased switching between rate categories across phylogenetic lineages for a

particular codon position. The M3+S2 model tests the added parameter of unequal switching between rate categories, allowing for certain sites to switch more frequently from one selection rate category (e.g., ω_1) to another selection rate category (e.g., ω_3). A series of nested log-likelihood ratio tests (LRT) were performed to determine the best-fit model for the data under a χ^2 distribution. Results were mapped onto the respective phylogenies based on which selection category a particular site had the highest probability of belonging to.

Three dimensional protein modeling was performed by using pdb template identification and 3D modeling programs in Swiss-Model Workspace (swissmodel.expasy.org). The Arabidopsis ADF1 pdb structure 1f7sA was used for modeling ZmADF5, human CFL1 3josW for itself, and chicken cofilin 1tvjA for DrCFL2. The N- and C-terminal ends of AtADF11 and AtADF6 were too divergent to be mapped to the best template. Structures were visualized in pymol (www.pymol.org).

Results:

Sequence Acquisition and Alignment

In total, 60 plant ADF protein variants from 8 plant species and 24 animal ADF/CFL protein variants from 12 species were examined for phylogenetic inferences and estimates of selection (Supplementary Tables 2.1 and 2.2). Per species, plants had a far greater number of ADF/CFL genes and protein variants than animals (Table 2.1). Species were chosen to represent important steps within the evolutionary history of this gene family. For the plant ADFs, the bryophyte *P. patens* was chosen as the outgroup/root of the tree. *Physcomitrella patens* has only a single ADF variant that has

been well characterized in the literature (Augustine et al. 2008). The second outgroup species used was the lycophyte *Selaginella moellendorffii*. The *S. moellendorffii* genome is comprised of two haplotypes that are nearly identical in nucleotide sequence (Banks et al., 2011) and initially four ADF protein variants were identified from the annotated *S. moellendorffii* genome: Sm146459, Sm230142, Sm270871, and Sm233521. Upon further inspection it was discovered that Sm270871 and Sm233521 were nearly identical in nucleotide sequence to Sm148459 and Sm230142, respectively. These two pairs of ADF protein variants are most likely a product of the two genomic haplotypes; therefore, only one pair of ADF sequences were used for analysis (Sm146459 and Sm230142). To verify that both *S. moellendorffii* protein variants are expressed, RNA was extracted from *S. moellendorffii* tissue and ADF RNA levels were quantified using qRT-PCR. Results showed that both protein variants are indeed expressed, with Sm146459 being expressed at 2-fold higher levels than Sm230142 in the frond tissue examined (Figure S2.1). Two model monocot species were chosen as well as four dicot species to sample the evolutionary history of angiosperms. Of all the plant species included, *O. sativa* by far contained the largest number of annotated ADF sequences. For the sake of simplicity, any fully redundant or nearly identical protein variants (i.e., < 5 nucleotide differences) were discarded from the analysis because they were unlikely to contribute significantly to the overall outcome of the selection analyses. Gymnosperms were not included simply due to the fact that there were no gymnosperm species that had their complete ADF gene family annotated at the time of our analysis.

For the animal dataset, a greater of number species were included because the animal ADF/CFL gene families were much smaller relative to plants (Supplementary

Table 2.3). Four invertebrate species were included in the analysis and each contained only a single ADF/CFL sequence. The only exception was amphioxus (*Branchiostoma floridae*), which had three annotated CFL/ADF sequences. Only one of the three CFL/ADF variants was included for analysis as inclusion of all three variants proved problematic for selection analyses as the runs would fail to optimize when all three variants were included. In total, seven species were chosen as the vertebrate representatives including two fish species (*S. salar* and *D. rerio*), an amphibian (*X. tropicalis*), a bird (*G. gallus*), and three mammalian species (*S. scrofa, M. musculus*, and *H. sapiens*). Of the vertebrates, the two fish species (*S salar* and *D rerio*) and mammalian species possess the most variants with 3 CFL variants each (Table 2.1). All other species possess only two CFL/ADF variants that have been well documented in the literature. Initially four *S. salar* sequences were found in our search, but a BLAST search of the RNA transcripts revealed that one of the variants was much closer in identity to protist RNA transcripts of ADF/CFL. This variant was discarded from the analysis as it might represent an error in annotation. While *S. salar* ADF/CFLs variants have not been previously characterized, the inclusion of only three variants does coincide with the fact that *D. rerio* is known to possess three ADF/CFL variants (Lin et al., 2010) and a preliminary phylogenetic analysis grouped this variant with the invertebrate ADF/CFL variants (data not shown). In addition to the invertebrate and vertebrate species, a single protist species was included: *Monosiga brevicolis*. This protist species has been suggested to be the closest unicellular relative to animals (King, 2008).

Looking across kingdoms, the average nucleotide divergence based on p-distance (proportion of variable nucleotide positions) ranged from 50 – 60% between plant and

animal ADF/CFL variants (Figure S2.2). Likewise, there was an equally high degree of sequence divergence between vertebrate and invertebrate ADF/CFL variants, again ranging from 50 – 60% (Figure S2.2). Interestingly, invertebrate ADF/CFL variants were closer in sequence divergence to plant ADF variants than they were to animal CFL/ADF variants. The exception is amphioxus, which seemed to be equal in sequence divergence between plant and animal ADF/CFL variants. We did not see this high degree of sequence divergence between ADF/CFL variants within plant lineages and within vertebrate lineages, respectively (Figure S2.2); however, invertebrate ADF/CFL variants ranged in divergence by 40 – 60% (Figure S2.2).

Phylogenetic Analyses of Plant and Animal ADF/CFLs

Previously published phylogenetic analyses have shown that the plant and animal ADF/CFL gene families partition into two distinct groups (Figure 2.1) (Bowman et al. 2000; Maciver and Hussey 2002). Due to this deep divergence between the plant and animal ADF/CFL gene families, the two groups were analyzed separately. The phylogenetic analysis concentrating on the plant ADF variants recovered the four ancient subclass previously described (Feng et al. 2006; Ruzicka et al. 2007). Of special note in our analysis was the placement of the moss *P. patens* and the lycophyte *S. moellendorffii* ADF variants within the phylogeny. The single *P. patens* ADF variant grouped outside of the four subclasses while one *S. moellendorffii* ADF variant (Sm146459) grouped with subclass I and subclass II and the other *S. moellendorffii* ADF variant (Sm230142) grouped with subclass III and subclass IV (Figure 2.2a). This suggests that there was an ancient split that occurred in the plant ADF gene family in plant ancestry postdating

common ancestry with moss 600 mya (Supplementary Table 2.2) and predating the common ancestor with lycophytes 450 mya. Also seen in the phylogeny is a further divergence between monocot and dicot ADF variants within subclasses II, III, and IV, as monocot and dicot variants form distinct subgroups within the three subclasses. However, this divergence between monocot and dicot variants was not seen within subclass I. The two most diverse subclasses were subclass I and II, with more ADF variants per species being represented in these two subclasses. In contrast, subclass III and subclass IV contain far fewer ADF variants per species; however, each angiosperm species had at least one ADF variant within these two subclasses (Figure 2.2a).

The phylogenetic analysis focusing on the animal ADF/CFLs revealed that the invertebrate and vertebrate ADF/CFL sequences form two distinct clades (Figure 2.2b). The vertebrate ADF/CFL variants further diverged into three classes: CFL1, CFL2, and ADF/Destrin. Of these, only mammals contained representatives in all three classes. In fact, it was only in the mammalian lineage that we found the non-muscle CFL1 variants. All other species had representatives only in the CFL2 and ADF/Destrin classes. Our phylogenetic analysis provided greater resolution for two particular lineages: the amphibian *X. tropicalis* and the two bony fish species, *D. rerio* and *S. salar*. As with mammals, fish possess three distinct ADF/CFL variants. However, only *S. salar* CFL2 and *D. rerio* CFL2 grouped within the other vertebrate ADF/CFL sequences, the muscle-specific CFL2 class. The remaining fish ADF/CFL variants grouped outside of the three vertebrate classes suggesting that these protein variants have progressed on an independent evolutionary trajectory that is unique to fishes (Figure 2.2b). *Xenopus tropicalis* has two ADF/CFL variants that have historically been shown to group with

each other, outside of the three conserved classes of vertebrate ADF/CFLs. In our analysis, the two *X. tropicalis* variants grouped well within the vertebrate sequences, with one variant grouping with the vertebrate ADF/Destrin group and the second variant grouping with the muscle-specific CFL2 group (Figure 2.2b).

Selection Analyses

Next, we tested whether selective constraint varies across the ADF/CFL plant and animal gene families. The main goal was to determine whether there are episodic instances of accelerated rates of evolution occurring both across evolutionary history as well as across individual protein sequences. For this analysis, a series of likelihood ratio tests (LRT) were performed using analyses from the program fitModeL (Guindon et al., 2004). Focusing on the plant ADF gene family, the LRT analyses revealed that the Guindon's M3+S2 model of codon evolution best fit the data, suggesting that selection varies both across the protein sequence as well as across the gene phylogeny. The results from the M3+S2 analysis revealed that while most codons were under tight selective constraint, there were rare instances where selection was relaxed. Sixty-four percent of codon positions were found in the ω_1 rate category with an estimate of 0.01 while 0.9% of sites were found in the ω_3 rate category that had a dN/dS estimate far greater than 1 ($\omega_3 = 19.99$, Table 2.2). A very similar pattern of codon evolution was seen for the animal CFL/ADF gene family as the M3+S2 model of codon evolution was also shown to best fit the animal dataset by LRT analysis. Furthermore, as with the plant ADFs, while the animal CFL/ADF protein variants were largely under tight selective constraints ($\omega_1 = 0.01$, $p1 = 0.65$), there were rare periods of significant accelerated rates of

evolution with a ω_3 estimate of 19.53 ($p3 = 0.015$, Table 2). For both the plant and animal ADF/CFL proteins, there was a greater degree of switching from the ω_2 to the ω_3 selection category. While traditionally the ω_2 rate category has been classified as the neutral selection category, it should be noted that the estimate for ω_2 was still considerably lower than 1. This suggests that sites were switching from a highly constrained selective process to a dramatically increased rate of evolution (e.g., for plant ADFs, switching occurs from an ω_2 of 0.13 to an ω_3 of 19.99, Table 2.2).

To discern where along the protein sequence these sites occurred, the number of branches that had the highest probability of containing a particular amino acid position within the ω_3 rate category was plotted out across the protein sequence (Figure 2.3A-B). Focusing solely on the plant ADF gene family, 10 of the 11 amino acid positions that had the highest probability of being grouped within the ω_3 rate class were located outside of any known G-actin and F-actin binding domains and not within the NLS (Figure 2.3A). Two sites in particular had a large number of branches in the ω_3 rate class: codon position 9 and codon position 157. Plotting these branches out on the phylogeny, we found that there were subclass specific, as well as monocot and dicot specific, patterns of relaxed selection (Figure 2.4A-C). Codon position 9 was found to have the highest probability of being in the ω_3 rate category for dicots within subclass IV only (Figure 2.4C). Likewise, codon position 157 was found to be in the ω_3 rate category for Subclass II dicots only (Figure 2.4A). Coincidentally, there was a third site that displayed a striking pattern: codon position 26 was found to have experienced accelerated rates of evolution in the subclass III monocot ADF variants only (Figure 2.4B). Mapping codon position 26 onto a 3D model of *Z. mays* ADF5, a subclass III monocot ADF variant,

showed that this particular site is located on an alpha helix that lies on the exterior of the protein, with its side chain exposed to the extracellular matrix (Figure 2.5A)

For the animal CFL/ADF protein variants, 8 of the 10 sites with the highest probability of being in the ω_3 rate category were found outside of all known conserved binding regions and secondary structural elements (Figure 2.3B). There were two notable exceptions: codon position 63, which is located within a β sheet, and codon position 173, which is located within a conserved F-actin binding domain. Codon position 63 was found to be in the ω_3 rate category for a single *S. salar* protein variant while codon 173 had the highest probability of being in the ω_3 rate category for the branch leading to be fish specific CFL group and an *S. salar* sequence within that group (data not shown). Plotting the remaining branches that had amino acid positions within the ω_3 rate category across the phylogeny once again showed some interesting patterns of codon evolution. Codon position 125 was found to have the highest probability of experiencing accelerated rates of evolution for the mammalian CFL1 protein variants (Figure 2.4D). In addition, codon position 69 was found to be in the ω_3 rate category for the *S. salar* and *D. rerio* protein variants that group with the CFL2 protein variants (Figure 2.4E). Interestingly, this site was found to be in the ω_1 rate category for the two *S. salar* and *D. rerio* protein variants that are basal to the rest of the CFL/ADF protein variants (Figure 2.4E). Codon position 125 was mapped onto the 3D structure of the *H. sapiens* CFL1 protein variant and showed that this site lies on an outer loop of the protein towards the C-terminal end. Likewise, when codon position 69 was mapped onto a 3D model of *D. rerio* CFL2, it was shown to also lie on an outer loop of the protein.

Discussion:

Gene duplications in the actin gene family and actin binding protein families like ADF/CFLs in Angiosperms and mammals parallel the evolution of structural and developmental complexity as these organisms evolved from single celled algal and choanoflagelate ancestors, respectively. Once believed to be solely involved in the depolymerization of actin filaments, these small proteins have been shown to be involved in many intricate cellular processes. It is likely that it was these duplication events that facilitated their diversity and subsequent versatility within eukaryotic organisms. Our new phylogenetic analyses coupled with the results from our estimates of selective pressure provides some interesting insights into the evolutionary history of this important class of proteins.

The inclusion of the two moss species, the bryophyte *P. patens* and the lycophyte *S. moellendorffii*, allow us to gain a better understanding of the divergence of the four plant ADF subclasses. The placement of the *P. patens* and *S. moellendorffii* ADF variants within the new phylogeny suggests that there was ancient split between the plant ADF protein variants that predates the further diversification seen in the angiosperm lineage. This is further emphasized by the gene expression results for the two *S. moellendorffii* protein variants examined. The majority of this paper has emphasized protein evolution, but regulatory changes can also have a huge impact on the diversification of gene duplicates (Force et al., 1999). The quantification of mRNA levels of the two *S. moellendorffii* ADF variants suggests that expression level, and therefore regulation in general, may be a significant factor in the divergence of the plant ADF proteins. In the frond tissue that was screened, the RNA levels of Sm146459 was

much higher than that of the other paralog, Sm230142. This finding is striking in that the Sm146459 variant groups with subclass I and II, which are the two most highly expressed ADF subclasses. Plant ADF protein variants within subclasses III and IV, which Sm230142 groups with, are expressed at a much lower level than that of subclasses I and II. No strong conclusions can be drawn at this point, as only one tissue type was examined and the remaining two protein variants in the *S. moellendorffii* genome would need to be screened, but the expression levels of the these two ancient variants seems to correlate with the divergence in expression levels in the angiosperm subclasses.

The updated plant ADF phylogeny also revealed another unique pattern in that there is a divergence between monocot and dicot ADF variants within three of the subclasses. There has been discussion of this divergence in previous studies, but our analysis provided a more robust finding than previously published (Feng et al. 2006). In subclass II, III, and IV, we see that the monocot ADF variants form distinct phylogenetic groups apart from dicot ADF variants. What is even more interesting about these splits is that there are also dicot and monocot specific patterns of codon evolution occurring (discussed below). Subclass I presented a unique case, as there was not any detectable divergence between the monocot and dicot ADF variants. In fact, subclass I seems to be further split into two separate subgroups, with monocot and dicot representatives in each, but not all plant species examined contained a representative within each subgroup. It is hard to make any reasonable conjectures about this except to say that there is possibly convergent evolution occurring between the monocot and dicot ADF variants within this subclass.

Our updated animal ADF/CFL phylogeny also provided some needed resolution, especially for the vertebrate lineage. As has been noted in previous studies, there is a major split between the invertebrate and vertebrate ADF/CFL variants (Bowman et al. 2000; Maciver and Hussey 2002). While most invertebrates possess only a single ADF/CFL protein variant, there are three classes of ADF/CFL variants within vertebrates: CFL1, CFL2, and ADF/Destrin (Bowman et al. 2000; Maciver and Hussey 2002). In our analysis, we find that only mammals have a representative in all three vertebrate ADF/CFL classes while lower order vertebrates have only CFL2 and ADF/Destrin representatives. The one exception is the two fish species included in this study (zebrafish and salmon), as they also contain three ADF/CFL variants but only one of their variants (S. salar CFL2 and D. rerio CFL2) grouped within one of the three conserved vertebrate classes (CLF2). Also of special note is the resolution our new phylogeny provided for the *X. tropicalis* CFL/ADF sequences. Previous analyses had the two well-characterized *Xenopus* protein variants forming their own phylogenetic group outside of the three known vertebrate ADF/CFL classes, resembling the fish ADF/CFL variants (Bowman et al. 2000; Maciver and Hussey 2002). Our new analysis was able to resolve their placement within the phylogeny, showing that one of the *X. tropicalis* variant groups with the ADF/Destrin class while the second variant groups with the CFL2 class of CFL protein class.

The main goal of this paper was to assess the role that selection has played in the evolution of the ADF/CFL gene family. Our analyses clearly show that selective pressure has varied not only across the protein sequence, but also across phylogenetic lineages. This was true not only for the plant ADF gene family, but also for the

vertebrate ADF/CFL gene family. First, we found that even though the ADF/CFL proteins have been under tight selective constraint for the majority of the time, there have been rare, episodic events of relaxed selective constraints. For the plant ADF phylogeny, we find that 0.9% of the time, or 0.9% of the amino acids in the entire phylogeny have experienced significantly high rates of evolution ($\omega_3 = 19.99$). The same was found for the animal ADF/CFL family with slightly more codons having experienced elevated rates of evolution (1.5% of codons in ω_3 category). The fact that these proteins have been under tight selective constraints for the majority of their evolution is not surprising as ADF/CFL proteins are extremely small and basically made up of conserved actin binding domains. However, it was unexpected to see such significantly high rates of evolution occurring in the ADF/CFL gene family, albeit rare. The switching pattern between selection categories reveals that codon positions have a propensity to switch from purifying/negative selection to accelerated rates of evolution. The rate of switching between the ω_2 to ω_3 selection category was significantly higher than switching between any other categories for both the plant and animal datasets. You will notice, however, that the estimates for ω_2 for both datasets are far from that of neutrality as it was significantly lower than 1. What we can conclude from this is that these sites are not simply switching from neutral to relaxed selection; we are actually seeing these sites switching from relatively strong selective constraints to that of estimates seen more commonly for diversifying selection. Plotting the number of phylogenetic lineages that had a particular site in the ω_3 rate category across the protein sequence showed that these sites were widely scattered across the protein indicating that there is not any particular binding domain that can be implicated in functional diversification. In fact, most codons

that had a high probability of being in the ω_3 rate category were located outside of any known conserved binding regions or any known conserved secondary structures. There were two notable exceptions within the vertebrate ADF/CFL proteins, codon positions 63 and 173, but neither site showed any clear pattern except that they specific for two unrelated fish ADF/CFL variants.

By far, the most striking result was seen when plotting the codon positions in the ω_3 rate category (i.e., accelerated rates of evolution) across the phylogeny, revealing subclass and class specific patterns of codon evolution. Within the plant ADF protein variants, there were three particular sites that had a high probability of belonging to the ω_3 rate category, but each were contained within different subclasses. Codon position 157, which lies just outside of the conserved G-actin binding domain of the C-terminus, was found to have accelerated rates of evolution in dicot ADF protein variants in subclass II. Based on mutational and structural studies, this site was positioned outside of any known binding domains or conserved secondary structures. Along with accelerated rates of evolution, there was a clear pattern for this site as to which branches of the phylogeny had this site within the ω_1 rate category (high selective constraint). While codon 157 was estimated to be in the ω_3 rate category, it was found to be in the ω_1 rate category for the subclass III dicot ADF variants and also for two monocot variants in subclass III. So while codon 157 may have been important in the divergence of the subclass II ADF variants, it was equally important to remain conserved across the dicot variants in subclass III. Moving to codon position 26, this site was found to be in the ω_3 rate category for the monocot variants in subclass III. Interestingly, this site is within a known α-helix of the ADF-H domain. This α-helix lies on the external portion of the

protein with its side chain exposed to the extracellular matrix. The location and position of codon 26 in *Z. mays* suggests that this site could be important in forming new binding interactions. The third site, codon position 9, was found to be in the ω_3 rate category for the dicot variants in subclass IV while being in the ω_1 rate category for all of the ADF variants in subclass III. This site is located within a region of the protein that is unique to subclass IV variants, an extended N-terminus region not found in any other ADF subclasses. This region/site could have had a fundamental role in the diversification between subclass III and subclass IV dicot ADF variants.

These patterns of codon evolution also extended into the animal ADF/CFL gene family. Codon position 125, which does not occur in any conserved regions, was found to have experienced accelerated rates of evolution in the CFL1 class of protein variants only. This is quite interesting as the CFL1 protein variants are only found within mammals, as far as we know. This particular region could be involved in the divergence between CFL1 and the other two classes of ADF/CFL protein variants. Interestingly, this codon position mapped to an outer loop of the *H. sapiens* protein model, suggesting that it could be involved in new protein interactions. Codon position 69 was found to be important in the evolution of a class of CFL fish protein variants. This particular site was found to be in the ω_3 rate category for the CFL2 fish protein variants. What is interesting is that while most of the vertebrate ADF variants had this site in the ω_2 rate category ($\omega_2 = 0.05$), the fish CFL variants that are basal to all of the vertebrate ADF/CFL variants (known as CFL1 in *D. rerio* literature) had this site in the ω_1 rate category ($\omega_1 = 0.003$). This comes with the caveat that the two fish variants had relatively low posterior probabilities of this site being in the ω_3 rate category. In fact, this site had close to equal

probability of being in the ω_2 rate category. With such low posterior probabilities, it's difficult to say whether this site is truly experiencing accelerated rates of evolution, but it may be worth not completely dismissing the site as a site of interest. What we do know about these two variants from *D. rerio* literature is that CFL2 is expressed solely in the muscle tissue of *D. rerio* while the CFL1 protein variant is an essential component of gastrulation (Lin et al., 2010). Even though this site is relatively constrained across all vertebrate ADF/CFL variants, it seems to be even more constrained in these two basal fish CFL1 variants. This is highly suggestive that this site played a fundamental role in the divergence between fish ADF/CFL variants. The three dimensional model gives even further support of this as this site also mapped to an outer loop of the CFL2 protein, which has the potential to act as a new interactive surface.

Taken together, the results from our study provide tantalizing clues to the evolutionary processes and possible mutations that facilitated functional divergence between gene family members. Our analyses show that selective pressure has varied during key points in the evolutionary history of the ADF/CFL proteins that could have implications in functional divergence. We see that there are clear patterns of codon evolution occurring within plant ADF subclasses and animal ADF/CFL classes. Furthermore, protein locations of three amino acid positions with accelerated rates of evolution (codon 26 in plants and codons 125 and 69 in animals) suggests these sites may have important functional implications and are good targets for future molecular work. It is our hope that the results from our study will be able to further guide future studies in the molecular characterization of the members of the ADF/CFL gene family.

Acknowledgments

The authors would like to thank Kelly A. Dyer for her help and guidance with the phylogenetic and selection analyses and Zac Wood for aid in protein structure modeling. This research was supported by a National Institutes of Health research grant to R.B.M. (GM 36397).

Literature Cited

Andrianantoandro E, Pollard TD. (2006) Mechanism of actin filament turnover by severing and nucleation at different concentrations of ADF/Cofilin. Mol. Cell. 24: 13–23.

Augustine RC, Vidali L, Kleinman KP, Bezanilla M. (2008) Actin depolymerizing factor is essential for viability in plants, and its phosphoregulation is important for tip growth. The Plant Journal. 54, 863-875.

Bamburg, JR. 1999. Proteins of the ADF/Cofilin Family: Essential Regulators of Actin Dynamics. Annual Review of Cellular Developmental Biology. 15: 185-230.

Banks, JA et al. (102 Coauthors). 2011. The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. Science. 332:960-963.

Bernstein, B.W., and Bamburg, J.R. 2010. ADF/Cofilin: a functional node in cell biology. Trends in Cell Biology. 20(4), 187-195.

Bettinger, BT, Gilbert, DM, Amberg, DC. 2004. Actin up in the nucleus. Nature Reviews Molecular Cell Biology. 5: 410-415.

Bowman, GD, Nodelman, IM, Hong, Y, Chua, NH, Lindberg, U, Schutt, CE. 2000. A comparative structural analysis of the ADF/cofilin family. Proteins. 41:374-384.

Burgos-Rivera, B., Ruzicka, D.R., Deal, R.B., McKinney, E.C., King-Ried, L., and Meagher, R.B. (2008) Actin depolymerizing factor9 controls development and gene expression in *Arabidopsis*. Plant Mol Biol. 68 (6), 619-632.

Chan, C., Beltzner, C.C., and Pollard, T.D. 2009. Cofilin Dissociates Arp2/3 Complex and Branches from Actin Filaments. Current Biology. 19(7): 537-545.

Clement, M., Ketelaar, T., Rodiuc, N., Banora, M.Y., Smertenko, A., Engler, G., Abad, P., Hussey, P., and Engler, J.D. 2009. Actin-Depolymerizing Factor2-Mediated

Actin Dynamics Are Essential for Root-Knot Nematode Infection of *Arabidopsis*. Plant Cell. 21(9), 2963-2979.

Daher, FB, van Oostende, C, Geitmann, A. 2011. Spatial and temporal expression of actin depolymerizing factors ADF7 and ADF10 during male gametophyte development in *Arabidopsis thaliana*. Plant Cell Physiol. 52(7): 1177–1192.

Estornes, Y, Gay, F, Gevrey, JC, Navoizat, S, Nejjari, M, Scoazec, JY, Chayvialle, JA, Saurin, JC, Abello, J. 2007. Differential involvement of destrin and cofilin-1 in the control of invasive properties of Isreco1 human colon cancer cells. International Journal of Cancer. 121: 2162–2171.

Feng, Y, Liu, Q, Xue, Q. 2006. Comparative study of rice and Arabidopsis Actin-depolymerizing factors gene families. Journal of Plant Physiology 163:69-79

Force, A, Lynch, M, Pickett, FB, Amores, A, Yan, Y, and Postlethwait, J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 151:1531-1545.

Galkin, VE, Orlovaa, A, Kudryashov, DS, Solodukhina, A, Reislerb, E, Schröderc, GF, Egelmana, EH. 2011. Remodeling of actin filaments by ADF/cofilin proteins. PNAS. 108(51):20568–20572

Guindon, S., Rodrigo, A.G., Dyer, K.A., and Huelsenbeck, J.P. (2004) Modeling the site-specific variation of selection patterns along lineages. PNAS. 101 (35), 12957-12962.

Heckman, DS, Geiser, DM, Eidell, BR, Stauffer, RL, Kardos, NL, Hedges, SB. 2011. Molecular evidence for the early colonization of land by fungi and plants. Science. 293:1129-1133.

Lappalainen, P, Kessels, MM, Jamie, M, Cope, TV, Drubin, DG. 1998. The ADF homology (ADF-H) domain: A highly exploited actin-binding module. Molecular Biology of the Cell. 9:1951–1959.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. and Higgins D.G. (2007) ClustalW and ClustalX version 2. Bioinformatics 23(21): 2947-2948.

Lin CW, Yen ST, Chang HT, Chen SJ, Lai SL, Liu YC, Chan TH, WL, Lee SJ. (210) Loss of cofilin 1 disturbs actin dynamics, adhesion between enveloping and deep cell layers and cell movements during gastrulation in zebrafish. PLOS One. 5(12).

Maciver, SK, Hussey, PJ. 2002. The ADF/cofilin family: actin-remodeling proteins. Genome Biol. 3(5):3007.1-3001.12

McKim, K. M., C. Matheson, M. A. Marra, M. F. Wakarchuk, and D. L. Baillie. 1994. The *Caenorhabditus elegans unc-60* gene encodes proteins homologous to a family of actin binding proteins. *Mot. Gen. Genet.* 242:346-357.

McKim, KS, Heschl, MFP, Rosenbluth, RE, Baillie, DL. 1988. Genetic organization of the *unc-60* region in *Caenorhabditus elegans.* Genetics. 118:49-59.

Moore, MJ, Bell, CD, Soltis, PS, Soltis, DE. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. PNAS. 104(49): 19363-19368.

Nakashima, K, Sato, N, Nakagaki, T, Abe, H, Ono, S, Obinata, T. (2005) Two mouse cofilin isoforms, muscle-type (MCF) and non-muscle type (NMCF), interact with F-actin with different efficiencies. J. Biochem. 138:519–526.

Nylander JAA. (2004) MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Ohta, Y, Nishida, E, Sakai, H, Miyamoto, E. 1989. Dephosphorylation of Cofilin Accompanies Heat Shock-induced Nuclear Accumulation of Cofilin. The Journal of Biological Chemistry 264: 16143-16148.

Okreglak, V, Drubin, DG. 2007. Cofilin recruitment and function during actin-mediated endocytosis dictated by actin nucleotide state. The Journal of Cell Biology. 178 (7): 1251-1264.

Paavilainen, VO, Oksanen, E, Goldman, A, Lappalainen, P. 2008. Structure of the actin-depolymerizing factor homology domain in complex with actin. J. Cell Biol. 182(1):51-59.

Paradis E, Claude J, Strimmer K. (2004) APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 20:289–290.

Poukkula, M, Kremneva, E, Serlachius, M, Lappalainen, P. 2011. Actin-Depolymerizing Factor Homology Domain: A Conserved Fold Performing Diverse Roles in Cytoskeletal Dynamics. Cytoskeleton. 68:471-490.

R Core Team. (2012) R: A Language and Environment for Statistical Computing. ISBN 3-900051-07-0. http://www.R-project.org.

Ronquist, F, Huelsenbeck, JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572-1574.

Ruzicka, D R, Kandasamy, M K, McKinney, E C, Burgos-Rivera, B, and Meagher, R B. (2007) The ancient subclasses of Arabidopsis ACTIN DEPOLYMERIZING

FACTOR genes exhibit novel and differential expression. The Plant Journal. 52: 460-472.

Salvarezza, S.B., Deborde, S., Schreiner, R., Campagne, F., Kessels, M.M., Qualmann, B., Caceres, A., Kreitzer, G., and Rodriguez-Boulan E. 2009. LIM Kinase 1 and Cofilin Regulate Actin Filament Population Required for Dynamin-dependent Apical Carrier Fission from the Trans-Golgi Network. Molecular Biology of the Cell. 20(1): 438-451.

Shan, H, Zahn, L, Guindon, S, Wall, PK, Kong, H, Ma, H, dePamphilis, CW, Leebens-Mack, J. 2009. Evolution of plant MADS Box transcription factors: Evidence for shifts in selection associated with early angiosperm diversification and concerted gene duplications. Mol. Biol. Evol. 26(10):2229–2244.

Storz, JF, Hoffmann, FG, Opazo, JC, and Moriyama, H. 2008. Adaptive functional divergence among triplicated α-Globin genes in rodents. Genetics. 178: 1623-1638.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol. Biol. Evol. 28(10):2731–2739.

Tian, M.Y., Chaudhry, F., Ruzicka, D.R., Meagher, R.B., Staiger, C.J., and Day, B. 2009. *Arabidopsis* Actin-Depolymerizing Factor AtADF4 Mediates Defense Signal Transduction Triggered by the *Pseudomonas syringae* Effector AvrPphB. Plant Physiology. 150(2), 815-824.

Van Troysa, M, Huycka, L, Leymana, S, Dhaesea, S, Vandekerkhovea, J, Christophe Ampea, C. 2008. Ins and outs of ADF/cofilin activity and regulation. European Journal of Cell Biology 87:649–667

Wang, W, Eddy, R, Condeelis, J. 2007. The cofilin pathway in breast cancer invasion and metastasis. Nature Reviews Cancer. 7:429-440.

Waterston, RH, Thomson, JN, Brenner, S. 1980. Mutants with altered muscle structure in *Caenorhabditus elegans.* Dev. Biol. 77:271-302.

Yang, Z. (2007) PAML4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution. 24(8), 1586-1591.

Yeoh, S, Pope, B, Mannherz, HG, Weeds, A. 2002. Determining the differences in actin binding by human ADF and cofilin. J. Mol. Biol. 315, 911–925

Yoon, HS, Hackett, JD, Ciniglia, C, Pinto, G, Bhattacharya, D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Mol. Biol. Evol. 21(5):809-

818.

Zharkikh, A. (1994) Estimation of evolutionary distances between nucleotide sequences. Journal of Molecular Evolution. 39:315-329.

Zimmer, A, Lang, D, Richardt, S, Frank, W, Reski, R, Rensing, SA. 2007. Dating the early evolution of plants: detection and molecular clock analyses of orthologs. Mol. Genet. Genomics. 278(4):393-402.

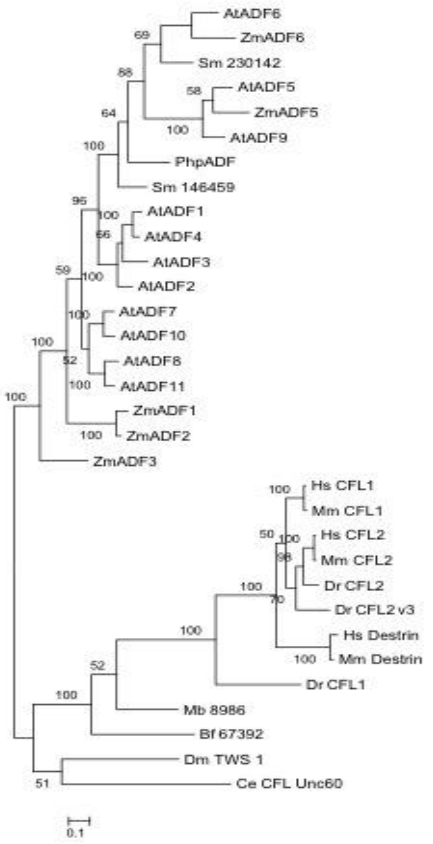


Fig. 2.1 Actin-depolymerization factor/cofilin phylogenetic analysis across multiple kingdoms. A Bayesian phylogenetic analysis of two plant species, 2 moss species, three invertebrate species, three vertebrate species, and a single protist species was performed using the Wag model of protein evolution and rooted post-run at the branch leading to the animal ADF/CFL group. The analysis highlights the deep divergence between plant and animal ADF/CFL proteins, as previously published (Maciver and Hussey, 2002; Bowman et al., 2000). Species designation are as follows: At, *Arabidopsis thaliana*; Zm, *Zea mays*; Sm, *Selaginella moellendorffii*; Php, *Physcomitrella patens*; Dm, *Drosophila melanogaster*; Ce, *Caenorhabditis elegans* ; Bf, *Branchiostoma floridae*; Mb, *Monosiga brevicolis*; Dr, *Danio rerio*; Mm, *Mus musculus*; and Hs, *Homo sapiens*.

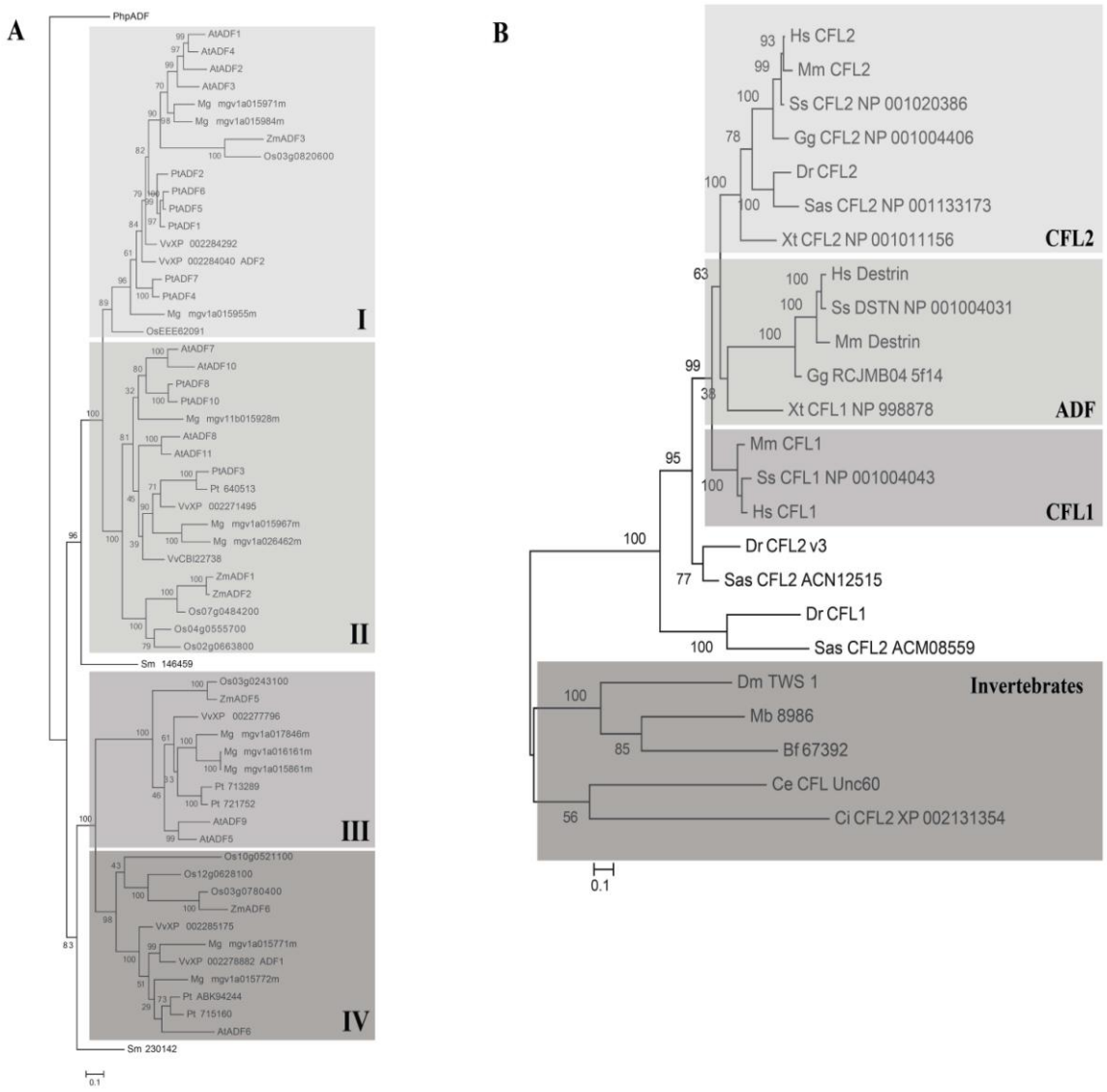

A
PhpADF
AtADF1
AtADF4
AtADF2
AtADF3
Mg mgv1a015971m
Mg mgv1a015984m
ZmADF3
Os03g0820600
PtADF2
PtADF6
PtADF5
PtADF1
VvXP 002284292
VvXP 002284040 ADF2
PtADF7
PtADF4
Mg mgv1a015955m
OsEEE62091
I
AtADF7
AtADF10
PtADF8
PtADF10
Mg mgv11b015928m
AtADF8
AtADF11
PtADF3
Pt 640513
VvXP 002271495
Mg mgv1a015967m
Mg mgv1a026462m
VvCBI22738
ZmADF1
ZmADF2
Os07g0484200
Os04g0555700
Os02g0663800
II
Sm 146459
Os03g0243100
ZmADF5
VvXP 002277796
Mg mgv1a017846m
Mg mgv1a016161m
Mg mgv1a015861m
Pt 713289
Pt 721752
AtADF9
AtADF5
III
Os10g0521100
Os12g0628100
Os03g0780400
ZmADF6
VvXP 002285175
Mg mgv1a015771m
VvXP 002278882 ADF1
Mg mgv1a015772m
Pt ABK94244
Pt 715160
AtADF6
IV
Sm 230142
0.1
B
Hs CFL2
Mm CFL2
Ss CFL2 NP 001020386
Gg CFL2 NP 001004406
Dr CFL2
Sas CFL2 NP 001133173
Xt CFL2 NP 001011156
CFL2
Hs Destrin
Ss DSTN NP 001004031
Mm Destrin
Gg RCJMB04 5f14
Xt CFL1 NP 998878
ADF
Mm CFL1
Ss CFL1 NP 001004043
Hs CFL1
CFL1
Dr CFL2 v3
Sas CFL2 ACN12515
Dr CFL1
Sas CFL2 ACM08559
Dm TWS 1
Mb 8986
Bf 67392
Ce CFL Unc60
Ci CFL2 XP 002131354
Invertebrates
0.1

Fig. 2.2 Bayesian phylogenetic analyses of the plant ADF and animal ADF/CFL gene families. (**a**) The updated plant ADF phylogeny recovers the four subclasses previously described (Ruzicka et al, 2007). The only exception are the ADF variants from the two moss species examined, *P. patens* and *S. moellendorffii*. The single *P. patens* ADF groups outside of all four subclasses while one *S. moellendorffii* variant is basal to subclass I and II and the other is basal to subclass III and IV. Also of special note is the divergences seen between the monocot and dicot ADF variants within each subclass except for subclass I. Species designations are as follows: At, *Arabidopsis thaliana*; Pt, *Populus trichocarpa*; Mg, *Mimulus guttatus*; Vv, *Vitis vinifera*; Os, *Oryza sativa*; Zm, *Zea mays*; Sm, *Selaginella moellendorffii*; and Php, *Physcomitrella patens*. (**b**) Bayesian phylogenetic analysis of ADF/CFL gene family within invertebrates and vertebrates. The invertebrate and vertebrate ADF/CFL protein variants form two distinct phylogenetic groups with the vertebrate proteins being further subdivided into three conserved classes: the muscle specific CFL2 group, and the two non-muscle groups CFL1 and ADF/Destrin. Species designation are as follows: Hs, *Homo sapiens*; Mm, *Mus musculus*; Ss, *Sus scrofa*; Gg, *Gallus gallus*; Xt, *Xenopus tropicalis*; Dr, *Danio rerio*; Sas, *Salmo salar*; Bf, *Branchiostoma floridae*; Ci, *Ciona intestinalis*; Dm, *Drosophila melanogaster*; Ce, *Caenorhabditis elegans*; Mb, *Monosiga brevicollis*.

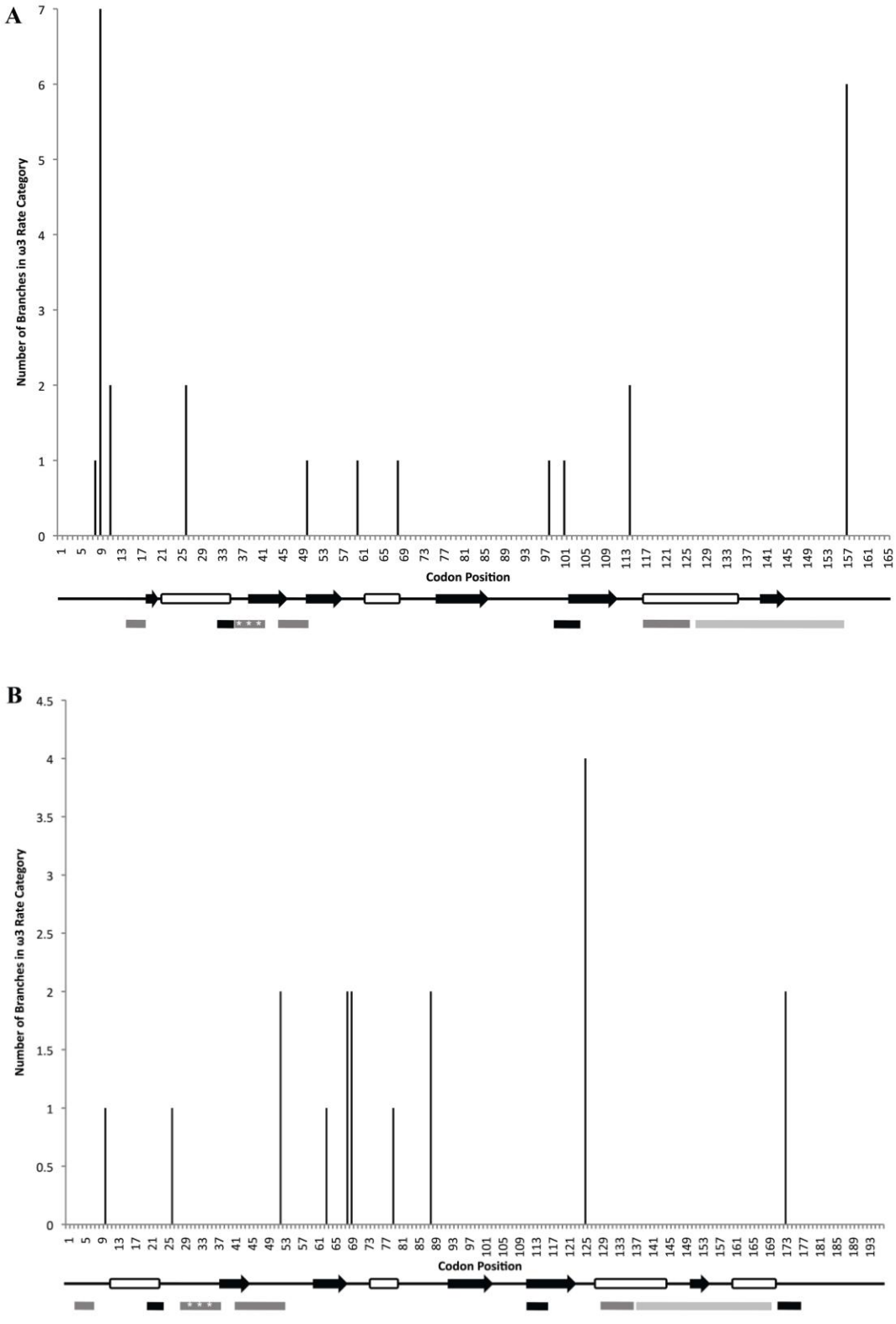

A
Number of Branches in ω3 Rate Category
Codon Position
B
Number of Branches in ω3 Rate Category
Codon Position

Fig. 2.3 The number of branches with the highest probability of being in the ω3 rate category plotted across each codon position for the (**a**) plant ADF and (**b**) animal CFL/ADF protein variants. Secondary structures are depicted directly below the graph: open cylinders represent α-helices and black arrows represent β-sheets. Conserved binding regions are depicted directly below the secondary structures: black bars represent F-actin binding regions, dark gray bars represent regions that are involved in both G-actin and F-actin binding, light gray bars are regions involved in only G-actin binding, and the dark gray bar with asterisks is the nuclear localization sequence. Tertiary structure and binding domains are based on Paavilaine et al (2008) and Galkin et al. (2011).

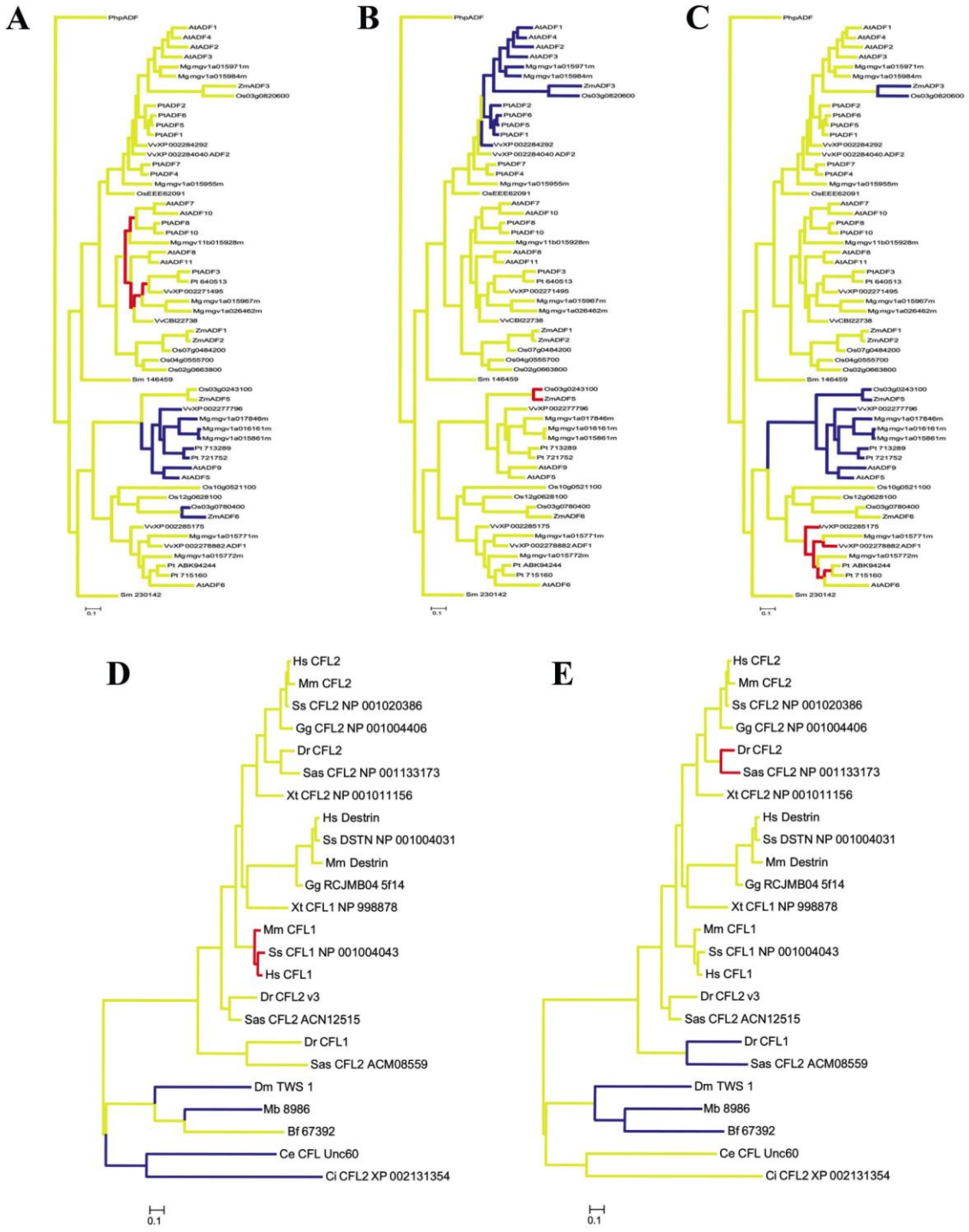

A
B
C
D
E
Hs CFL2
Mm CFL2
Ss CFL2 NP 001020386
Gg CFL2 NP 001004406
Dr CFL2
Sas CFL2 NP 001133173
Xt CFL2 NP 001011156
Hs Destrin
Ss DSTN NP 001004031
Mm Destrin
Gg RCJMB04 5f14
Xt CFL1 NP 998878
Mm CFL1
Ss CFL1 NP 001004043
Hs CFL1
Dr CFL2 v3
Sas CFL2 ACN12515
Dr CFL1
Sas CFL2 ACM08559
Dm TWS 1
Mb 8986
Bf 67392
Ce CFL Unc60
Ci CFL2 XP 002131354
0.1

Fig 2.4 Sites with highest probability of being in the ω_3 rate category mapped across the plant ADF (**a-c**) and animal ADF/CFL (**d-e**) gene family phylogenies. Red branches have the highest probability of being in the ω_3 rate category, yellow branches represents branches with the highest probability in the ω_2 rate category, and blue branches represents branches with the highest probability of being in the ω_1 rate category. The following codon positions were found to be in the ω_3 rate category in the plant ADF phylogeny: (**a**) codon position 157 (**b**) codon position 26 and (**c**) codon position 9. The following codon positions were found in the ω_3 rate category in the animal ADF/CFL phylogeny: (**d**) codon position 125 and (**e**) codon position 69.

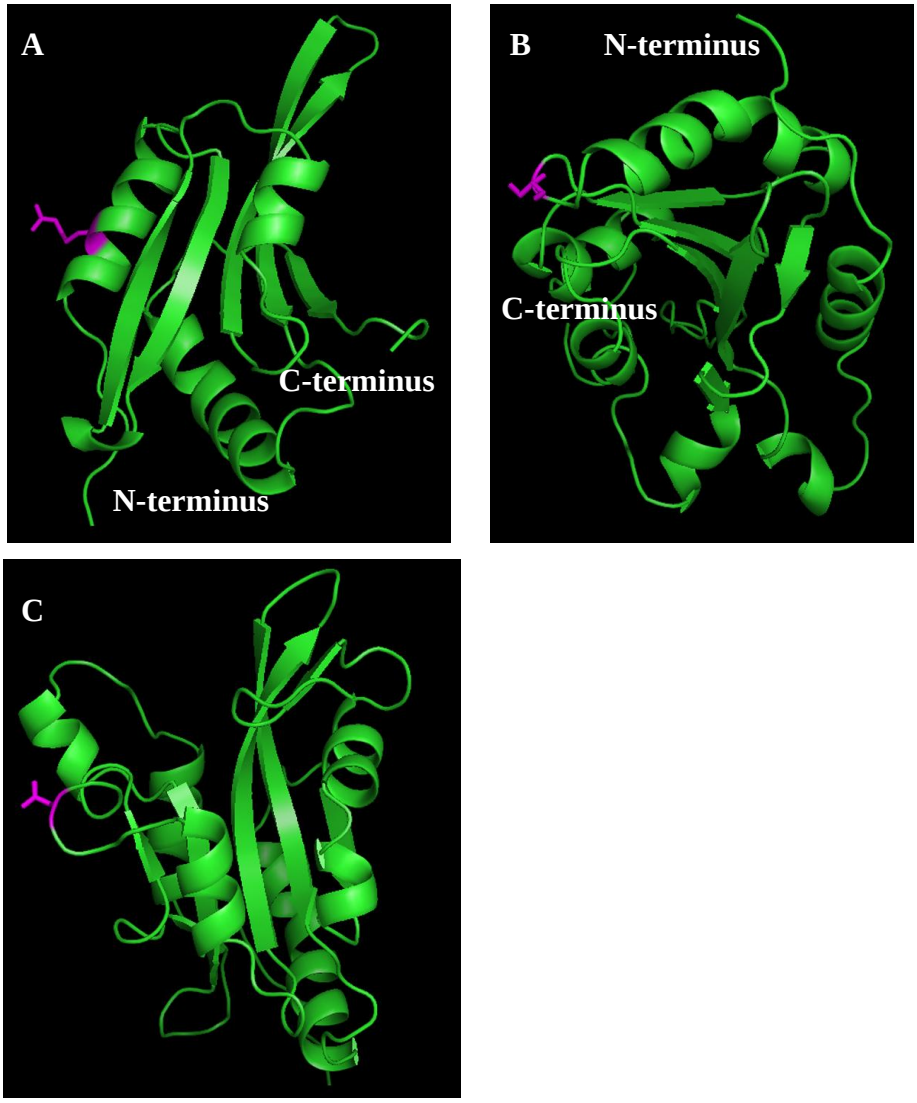


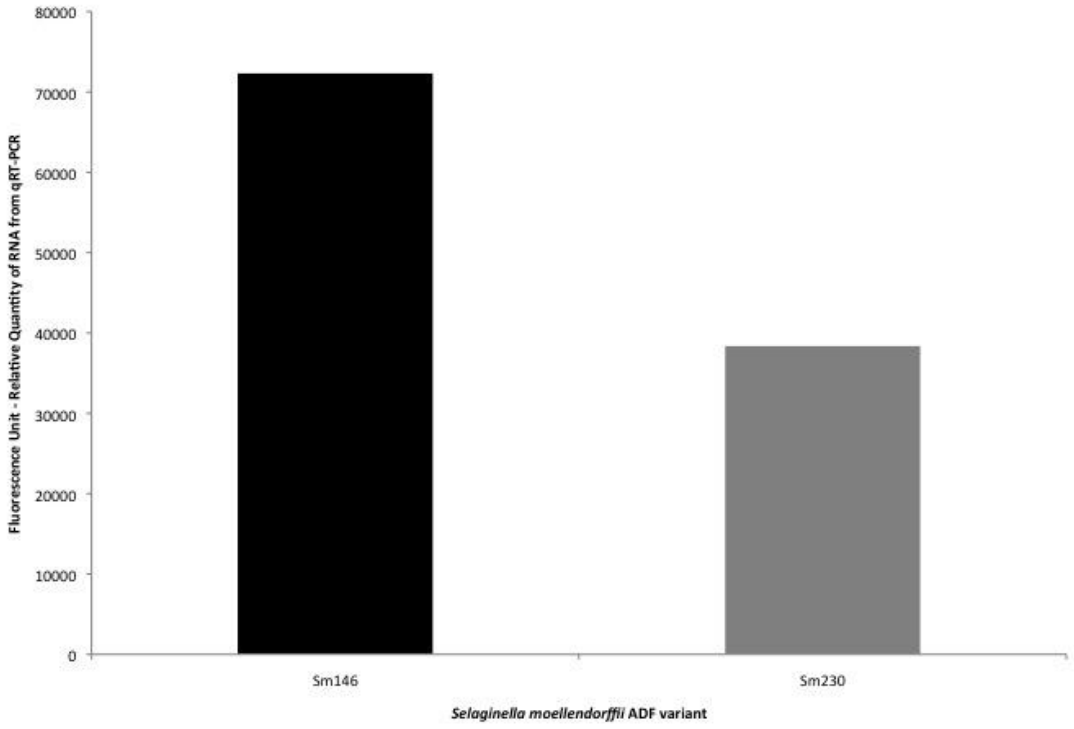
Fig 2.5 Three dimensional protein models of ADF/CFL proteins containing sites with accelerated rates of evolution (ω_3 rate category). (A) Zea mays ADF5 proteins with codon position 26 highlighted in magenta. (B) Homo sapiens CFL1 protein with codon position 125 highlighted in magenta. (C) Danio rerio CFL2 protein with codon position 69 highlighted in magenta.

Table 2.1 The number of ADF/CFL protein variants used for each species analyzed

	Species	# of Variants		Species	# of Variants
Plant ADF	*Physcomitrella patens*	1	**Animal CFL/ ADF**	*Monosiga brevicolis*	1
	Selaginella moellendorffii	2		*Ciona intestinalis*	1
	Zea mays	5		*Branchiostoma floridae*	1
	Oryza sativa (japonica)	9		*Drosophila melanogaster*	1
	Mimulus guttatus	11		*Caenorhabditis elegans*	1
	Vitis vinifera	7		*Salmo salar*	4
	Populus trichocarpa	14		*Danio rerio*	4
	Arabidopsis thaliana	11		*Xenopus tropicalis*	2
				Gallus gallus	2
				Sus scrofa	3
				Mus musculus	3
				Homo sapiens	3

Table 2.2 Estimates of selective pressure across the plant ADF and animal ADF/CFL gene family

	Model	ln(Likelihood)	Comparison	LRT Statistic	df	p-value	ω Estimates (Proportion of sites)
Plant ADF	M0	-13191.73					ω1= 0.07(1.0)
	M3	-12916.57	M0 vs M3	550	4	p<0.0001	ω1= 0.02(0.48); ω2= 0.12(0.38); ω3= 0.26(0.14)
	M3	-12916.57					
	M3S1	-12772.53	M3 vs M3S1	288	1	p<0.0001	ω1= 0.007(0.64); ω2= 0.15(0.30); ω3= 1.15(0.06)
	M3S2	-12737.16	**M3S1 vs M3S2**	70.74	2	**p<0.0001**	ω1= 0.010(0.64); ω2= 0.13(0.35); **ω3= 19.999(0.009)**
							R(1,2) = 0.17; R(1,3) = 0.001; R(2,3) = 12.31
Animal ADF/CFL	M0	-7809.34					ω1= 0.03(1.0)
	M3	-7617.07	M0 vs M3	192	4	p<0.0001	ω1= 0.005(0.5); ω2= 0.025(0.37); ω3= 0.21(0.13)
	M3	-7617.07					
	M3S1	-7496.88	M3 vs M3S1	120	1	p<0.0001	ω1=0.00001(0.66); ω2= 0.03(0.22); ω3= 0.40(0.12)
	M3S2	-7473.81	**M3S1 vs M3S2**	23.07	2	**p<0.0001**	ω1= 0.003(0.65); ω2= 0.05(0.33); **ω3= 19.53(0.015)**
							R(1,2) = 0.06; R(1,3) = 0.001; R(2,3) = 3.91



Supplementary Figure 2.1 (Fig S2.1): RNA levels of two *S. moellendorffii* ADF variants, Sm 146459 (Sm146) and Sm 230142 (Sm230).

Supplementary Table 2.1 Nucleotide divergence between protein variants based on p-distance

AtADF6																															
AtADF1	0.340																														
AtADF2	0.354	0.146																													
AtADF3	0.372	0.163	0.194																												
AtADF4	0.344	0.118	0.149	0.160																											
AtADF5	0.340	0.344	0.309	0.358	0.333																										
AtADF7	0.392	0.281	0.278	0.274	0.278	0.313																									
AtADF8	0.365	0.288	0.295	0.306	0.292	0.313	0.247																								
AtADF9	0.375	0.358	0.361	0.365	0.365	0.184	0.361	0.372																							
AtADF10	0.372	0.257	0.274	0.285	0.264	0.326	0.125	0.229	0.344																						
AtADF11	0.365	0.302	0.316	0.302	0.299	0.326	0.219	0.118	0.358	0.201																					
ZmADF5	0.365	0.333	0.354	0.365	0.344	0.257	0.309	0.340	0.257	0.326	0.344																				
ZmADF3	0.410	0.302	0.309	0.337	0.306	0.392	0.313	0.351	0.431	0.313	0.347	0.378																			
ZmADF1	0.403	0.295	0.316	0.323	0.333	0.365	0.264	0.302	0.351	0.264	0.281	0.274	0.326																		
ZmADF6	0.309	0.382	0.382	0.378	0.368	0.385	0.389	0.413	0.375	0.396	0.413	0.365	0.399	0.365																	
ZmADF2	0.406	0.288	0.302	0.316	0.326	0.354	0.260	0.285	0.340	0.250	0.267	0.274	0.337	0.038	0.361																
Sm_230142	0.354	0.306	0.306	0.295	0.299	0.330	0.313	0.319	0.340	0.340	0.326	0.316	0.354	0.337	0.354	0.337															
Sm_146459	0.347	0.278	0.302	0.285	0.302	0.354	0.323	0.333	0.344	0.330	0.330	0.319	0.375	0.330	0.372	0.330	0.285														
PhpADF	0.340	0.323	0.316	0.330	0.319	0.372	0.323	0.333	0.361	0.333	0.330	0.351	0.340	0.313	0.347	0.309	0.292	0.337													
Hs_CFL1	0.569	0.538	0.542	0.556	0.549	0.552	0.545	0.503	0.542	0.521	0.510	0.542	0.514	0.483	0.528	0.479	0.545	0.510	0.517												
Hs_CFL2	0.566	0.545	0.528	0.566	0.528	0.573	0.556	0.500	0.569	0.556	0.545	0.604	0.545	0.545	0.576	0.552	0.566	0.549	0.559	0.281											
Hs_Destrin	0.573	0.535	0.545	0.535	0.531	0.528	0.549	0.524	0.590	0.556	0.535	0.618	0.580	0.587	0.576	0.587	0.542	0.559	0.566	0.337	0.295										
Mm_Destrin	0.576	0.545	0.559	0.552	0.549	0.524	0.542	0.521	0.569	0.514	0.503	0.580	0.569	0.538	0.566	0.545	0.552	0.549	0.545	0.306	0.319	0.111									
Mm_CFL1	0.583	0.535	0.545	0.563	0.556	0.545	0.542	0.521	0.549	0.535	0.517	0.556	0.535	0.483	0.538	0.483	0.552	0.514	0.535	0.056	0.281	0.323	0.295								
Mm_CFL2	0.573	0.542	0.531	0.566	0.528	0.573	0.545	0.500	0.563	0.549	0.535	0.597	0.538	0.535	0.563	0.542	0.559	0.545	0.552	0.267	0.028	0.306	0.323	0.267							
Dr_CFL1	0.563	0.563	0.580	0.590	0.563	0.587	0.552	0.538	0.569	0.535	0.521	0.594	0.563	0.521	0.556	0.514	0.559	0.569	0.531	0.424	0.465	0.500	0.462	0.424	0.462						
Dr_CFL2	0.587	0.569	0.559	0.580	0.556	0.583	0.594	0.542	0.552	0.580	0.542	0.583	0.542	0.545	0.576	0.535	0.566	0.545	0.552	0.247	0.177	0.330	0.344	0.250	0.160	0.458					
Dr_CFL2_v3	0.573	0.536	0.556	0.563	0.559	0.556	0.542	0.531	0.556	0.517	0.545	0.556	0.514	0.497	0.556	0.503	0.549	0.531	0.559	0.229	0.260	0.330	0.288	0.229	0.250	0.448	0.236				
Dm_TWS_1	0.500	0.448	0.451	0.486	0.465	0.472	0.455	0.458	0.479	0.486	0.476	0.490	0.472	0.472	0.483	0.465	0.413	0.458	0.444	0.517	0.552	0.583	0.573	0.510	0.535	0.535	0.545	0.535			
Ce_CFL_Unc60	0.549	0.497	0.545	0.528	0.524	0.521	0.486	0.514	0.507	0.476	0.531	0.503	0.486	0.455	0.517	0.458	0.479	0.524	0.514	0.594	0.587	0.628	0.587	0.604	0.583	0.549	0.601	0.542	0.497		
Mb_8986	0.497	0.469	0.503	0.472	0.486	0.503	0.493	0.507	0.448	0.493	0.514	0.441	0.455	0.417	0.438	0.424	0.469	0.431	0.438	0.441	0.517	0.559	0.545	0.448	0.510	0.514	0.500	0.486	0.472	0.500	
Bf_67392	0.545	0.538	0.552	0.566	0.559	0.566	0.573	0.538	0.552	0.556	0.552	0.531	0.528	0.503	0.538	0.521	0.559	0.531	0.549	0.514	0.535	0.576	0.528	0.510	0.524	0.563	0.549	0.517	0.514	0.507	0.413

Supplementary Table 2.2 Plant species sampled for ADF/CFL sequences and their divergence times from a common ancestor with Arabidopsis

Species	Informative Classification	Estimated Divergence Times
Physcomitrella patens	Non-vascular plant, Bryophyta	600 mya
Selaginella moellendorffii	Vascular plant, Lycopodiophyta	450 mya
Zea mays	Angiosperm, monocot, Poaceae	250 mya
Oryza sativa japonica (rice)	Angiosperm, monocot, Poaceae	250 mya
Vitis vinifera	Angiosperm, dicot, Vitaceae	110 mya
Mimulus guttatus	Angiosperm, dicot, Scrophulariaceae	110 mya
Populus trichocarpa	Angiosperm, dicot, Salicaceae	70 mya
Arabidopsis thaliana	Angiosperm, dicot, Brassicaceae	0 mya

Estimates of divergence times among species were extrapolated from the following references: Heckman et al., 2001; Yoon et al., 2004; Moore et al., 2007; Zimmer et al., 2007.

Supplementary Table 2.3 Animal species sampled for ADF/CFL sequences and their divergence times from a common ancestor with *H. sapiens*

Species	Informative Classification	Estimated Divergence Times
Monosiga brevicolis	Protist, Choanoflagellate	1400 mya
Caenorhabditis elegans	Invertebrate, Protostome, Nematoda	950 mya
Drosophila melanogaster	Invertebrate, Protostome, Arthropoda	950 mya
Ciona intestinalis (sea squirt)	Invertebrate, deuterostome, Tunicata	750 mya
Branchiostoma floridae (lancelet)	Invertebrate, deuterostome, Cephalochordata	650 mya
Salmo salar (Atlantic salmon)	Vertebrate, Actinopterygii	450 mya
Danio rerio (zebrafish)	Vertebrate, Actinopterygii	450 mya
Xenopus tropicalis (Western clawed frog)	Vertebrate, Amphibia	360 mya
Gallus gallus (chicken)	Vertebrate, Aves, Phasianidae	310 mya
Sus scrofa (domestic pig)	Vertebrate, Mammalia, Suidae	66 mya
Mus musculus (domestic house mouse)	Vertebrate, Mammalia, Muridae	66 mya
Homo sapiens	Vertebrate, Mammalia, Hominidae	0

CHAPTER 3

Arabidopsis ACTIN-DEPOLYMERIZING FACTOR 6 (ADF6) represents a non-essential member of the ACTIN-DEPOLYMERIZING FACTOR gene family[1]

[1]Eileen M. Roy-Zokan, Muthugapatti K. Kandasamy, Elizabeth C. McKinney, Brunilis Burgos-Rivera, Lori M. King, Daniel R. Ruzicka, and Richard B. Meagher. To be submitted to the Journal of Plant Physiology.

Abstract:

The members of the ACTIN-DEPOLYMERIZING FACTOR (ADF) gene family represent a group of proteins that are important modulators of actin filament dynamics. Aside from cytoskeletal dynamics, these versatile proteins have also been implicated in such cellular processes as lipid and membrane metabolism, mitochondrial dependent apoptosis, chemotaxis, and the cytonuclear trafficking of actin into the nucleus. The plant ADF gene family phylogenetically partitions into four ancient subclasses that are differentially regulated. While great strides have been made in assessing the functional role of the 10 ADF proteins comprising subclasses I, II, and III, little is known about subclass IV. Within Arabidopsis there is a single subclass IV ADF variant, ADF6, and it is unique in that it is the only ubiquitously expressed ADF. Here we report that ADF6 expression, while ubiquitous, is temporally and spatially regulated through seedling and reproductive development. Compared to wild-type, plants defective in ADF6 developed normally with no cost on fitness over two generations and no apparent morphological phenotype. The *adf6-1* null mutation suppressed the early flowering phenotype of the *adf9-1* defective mutant suggesting ADF6 has an antagonistic role with ADF9 in regulating flowering time. However, this role is entirely dependent on ADF9 as the *adf6-1* mutant flowers normally. The results from this study provide initial evidence that ADF6 may be a non-essential gene in Arabidopsis, in spite the evidence that subclass IV ADFs have been preserved through 200 my of angiosperm evolution.

Introduction

The ability to adapt and handle environmental stress is of vital importance to all organisms, but especially so for plants. As sessile organisms, plants do not have the luxury of escaping their stress by simply moving to a new location. As such, plants have evolved a repertoire of biological tools that allow them to deal with changing environmental conditions. One toolkit that enhances their ability to cope with environmental stresses is the dynamic nature of the actin cytoskeleton. The actin cytoskeleton is involved in a number of important cellular processes such as determining cell polarity, apical growth, vesicle transport and fusion, organelle movement, and chromatin remodeling, to name a few. The multitude of interactions between the different actin protein variants within plants along with their highly diverse binding partners contributes to the organism's ability to respond to both mechanical and environmental stresses (Meagher et al., 1999a). The importance of the cytoskeleton in eliciting stress responses and its role in signaling pathways has been well documented (e.g. Machesky and Insall, 1999; Staiger, 2000). There are a number of essential classes of actin binding proteins that contribute to these cellular responses, and one of those is encoded by the actin-depolymerizing factor (ADF) gene family.

Actin-depolymerizing factors and very closely related animal cofilins (ADF/CFL) are a group of small proteins ranging from 13 - 19 kDa in size that are characterized by the presence of five β-strands that lie internal to four or more α-helices (Bamburg, 1999; Poukkula et al., 2011). These proteins are best known for their role in actin filament turnover, where they bind and sever actin monomers from actin filaments in a concentration dependent manner (Andrianantoandro and Pollard, 2006; Bamburg, 1999).

ADF/CFLs bind to at least two sites on G-actin and F-actin. At low cellular concentrations, ADF/CFL binds to the pointed end of an actin filament (F-actin), causing an increase in filament bending between bound and unbound regions that ultimately results in the bound actin being severed from the filament. At intermediate cellular concentrations, ADF/CFL will bind along the entire filament causing stabilization. At even higher cellular concentrations ADF/CFL proteins actually promote nucleation of filaments (Bernstein and Bamburg, 2010; Tholl et al., 2011). As more research has been conducted on these proteins, especially focusing on mammalian ADF/CFL, it has become quite evident that the cellular role of ADF/CFL extends far beyond actin filament turnover. The ADF/CFL proteins have been shown to be an essential component in cell motility and intracellular trafficking (Chan et al., 2009; Salvarezza et al., 2009; Okreglak and Drubin, 2007). These proteins have also been implicated in a variety of other cellular processes such as lipid and membrane metabolism, mitochondrial dependent apoptosis, chemotaxis, vesicle transport, and the cytonuclear trafficking of actin into the nucleus (Ohta et al., 1989; Chua et al., 2003; Bettinger et al., 2004; Lehman et al., 2006; Han et al., 2007; Bernstein and Bamburg, 2010).

Although the majority of studies on the ADF/CFL gene family have focused primarily on vertebrates, recent research into the function of plant ADF proteins suggest that their role within a plant cell is just as complex. The members of the plant ADF gene family in angiosperms are phylogenetically partitioned into four ancient subclasses (Ruzicka et al., 2007; Feng et al., 2006) with each subclass exhibiting differential expression patterns (Ruzicka et al., 2007). Plant ADF protein variants do function in the traditional sense as an actin filament depolymerizing protein as demonstrated in studies

focusing on Arabidopsis ADF1, ADF4, and ADF9 (Tian et al., 2004; Henty et al, 2011; Tholl et al., 2011); however, studies have also shown that they are equally important in developmental processes and signaling pathways as their vertebrate homologs. Some of the strongest evidence of this has come from work focusing on members of the Arabidopsis ADF gene family. Subclass I ADF variants seem to have a particularly important role in immunity, with both beneficial and harmful consequences to the plant. For instance, Arabidopsis ADF4 has been shown to be absolutely vital in defense against a particular strain of *Pseudomonas* as *adf4-1* mutant plants showed increased susceptibility to *Pseudomonas* infection (Tian et al., 2004). This response was specific for ADF4 and no other Arabidopsis ADF variant. Another subclass I variant, ADF2, was also shown to be important for pathogen response but instead of providing defense like ADF4, it actually facilitates nematode infection. Knocking down ADF2 within Arabidopsis using RNAi impedes the proper formation of large feeding cells upon nematode infection and resulted in very few nematodes reaching maturity (Clement et al., 2009). This indicates that ADF2 is important for the *success* in nematode infection rather than implementing a defense against infection. When we switch our focus to subclass II and subclass III Arabidopsis variants, we find that these ADF proteins are fundamental for developmental processes. Subcellular localization experiments focusing on pollen specific subclass II ADF variants highlight their role in pollen tube formation and growth (Daher et al., 2011). The subclass III ADF variant, ADF9, has been shown to be an important component of the flowering time pathway as knocking down ADF9 results in early flowering (Burgos-Rivera et al., 2009). Screening different regulators of flowering time revealed a significant decrease in transcript levels of FLOWERING LOCUS C

(FLC), a MADS-box transcription factor that acts as a master repressor of flowering, in *adf9-1* mutants (Burgos-Rivera et al., 2009). Additionally, transcript levels of downstream flowering activators showed consistent expression profiles that correlates with what is expected with the down regulation of FLC (Burgos-Rivera et al., 2009).

While great strides have been made in assessing the functional role of ADF proteins within subclasses I, II, and III, virtually nothing is known about subclass IV ADF variants. This subclass appears to predate the major split early in angiosperm evolution 200 to 250 million years ago (mya), because both monocots and dicots have ADF that are easily identifiable as subclass IV members based on protein sequence. Within Arabidopsis there is only a single subclass IV ADF variant, ADF6, and it is unique in that it is the only ubiquitously expressed ADF. It is moderately expressed in all tissues, including mature pollen, but it is most strongly expressed in root vascular tissue, root tip, and in cotyledons (Ruzicka et al., 2007). Aside from its general expression pattern, what is known about ADF6 is that it is 146 amino acids in length and contains three exons and two introns (Dong et al., 2001). The first intron of ADF6 is considerably larger than the second intron, being 248 base pairs in length versus 89 base pairs (Dong et al., 2001). To date, no studies have attempted to determine the functional role of ADF6 within Arabidopsis. In this investigation, we show that ADF6 expression is spatially and temporally regulated throughout development, both in young seedlings as well as in reproductive tissue. Despite this, knocking down ADF6 resulted in mutant plants that are morphologically identical to wild-type plants with no cost on fitness. These results suggest that ADF6 may represent a redundant and non-essential protein variant within the Arabidopsis gene family.

Materials and Methods

Plant growth conditions

All plant lines and mutant strains used for this study were of the Columbia ecotype genetic background. For plate-based assays, seeds were surface sterilized and plated onto ½ MS growth media consisting of 1% sucrose and 0.8% phytoagar (Caisson). Seeds were stratified for 48 hours at 4 ^{o}C before transferring to an incubator with 16 hours light 8 hour dark light cycle at 22 ^{o}C. For soil-based assays, seeds were sown directly onto 70% organic soil (3B) and again stratified for 48 hours at 4 ^{o}C. After stratification, flats were transferred to growth rooms where plants were grown at 22 ^{o}C under a 16 hour/8 hour light/dark cycle. Transgenic lines were established using the *Agrobacterium* floral dip method, as described by Clough et al. (2005). Positive transformants were selected on ½ MS growth media supplemented with 50 mg/L hygromycin and 300 mg/L tementin.

Establishment of Arabidopsis mutant lines

A T-DNA insertion line for the ADF6 gene (At2g31200) was obtained from TAIR (SAIL-648-A03), hereafter referred to as *adf6-1*. Other T-DNA insertion lines were available for the ADF6 gene; however, these lines all had the T-DNA insertion located at the very distal end of the 3'-UTR. As such, it was felt that they would not be good candidates for screening. The *adf6-1* line examined in this study contains a single insertion within the first intron of ADF6, which has been shown to be important in gene expression regulation in ADFs from other plant species (Mun et al., 2000, 2002; Jeong et al., 2007). To confirm the insertion location, DNA was extracted from individual plants

from the *adf6-1* line and genotyped for the insertion using the Sigma Plant DNA extraction kit. The wild-type ADF6 allele was amplified using two primers flanking the insertions site: ADF6-RP 5' – TATTGATAGATACGCGGCGAC – 3' and ADF6-LP 5' – GAATCCCTTGAAGCTCCCTAC – 3'. T-DNA insertion sites was verified using the ADF6-RP sense primer that is upstream from the insertion site and a T-DNA border primer SAIL LB-Distal (5' – TAGCATCTGAATTTCATAACCAATCTCGATACAC – 3') as the antisense primer. The insertion site was confirmed by sequencing putative T-DNA mutants with both the ADF6-RP and SAIL LB-Distal primers. Once the insertion site was confirmed and verified as a single insertion, the *adf6-1* line was backcrossed twice to Columbia wild-type and homozygous backcrossed lines were propagated through selfing. Double mutants were generated by crossing homozygous *adf6-1* plants with homozygous *adf5-1* or *adf9-1* plants while triple mutants were generated by crossing a homozygous adf5/9 double mutant already established in the lab with homozygous *adf6-1* plants. Heterozygous plants were identified through genotyping and were selfed to generate homozygous double and triple mutants, respectively.

Establishment of adf6-1 complementation and fimbrin-GFP transgenic lines

The ADF6 coding sequence was cloned in frame into the ACT2 promoter terminator pCAMBIA binary vector as previously described by Kandasamy et al. (2002). Briefly, an NcoI and BamHI restriction site was introduced to the beginning and end of the ADF6 cDNA sequence from a genomic cDNA library through modified primers. These restriction sites were used to clone the modified ADF6 sequence into a BlueScript vector and then the construct was shuttled into the ACT2 promoter-terminator pCAMBIA

binary vector already established in the laboratory. The A2pt::ADF6 expression vector was transformed into the *adf6-1* mutant and overexpression was quantified in T2 lines using qRT-PCR. The actin cytoskeleton of *adf6-1* was examined using the GFP-fABD2 reporter developed by Wang et al. (2004). Actin filament organization was screened in trichomes of 10 – 15 day old seedlings grown vertically on ½ MS plates using a Leica fluorescent stereo microscope (Leica Microsystems).

Histochemical analysis of ADF6 expression through development

Transgenic ADF6pi::GUS promoter fusion plant lines generated and described by Ruzicka et al. (2007) were used in this study. To examine the expression of ADF6 through development, T3 seeds from the ADF6pi::GUS transgenic line were sown out on ½ MS plates as described above. After stratification, plates were transferred to growth chambers where seedlings were grown vertically. The temporal and spatial expression of ADF6 expression during early seedling development was assayed every 12 hours for 15 days starting at 36 hours after stratification. Harvested seedlings were incubated at 37^{o}C in GUS staining solution (Jefferson et al., 1987) for 12 hours. To screen ADF6 expression in reproductive tissue, ADF6pi::GUS T3 seeds were sown directly onto soil; flowers were collected between stages 10 – 13 of flower development (staging based on Smyth et al., 1990) and stained in GUS staining solution overnight at 37^{o}C. GUS expression was examined using a Leica stereo microscope fitted with a digital color camera (Leica Microsystems).

Quantitative Real Time PCR

RNA expression levels were quantified on either 10-day old seedlings or mature rosette leaf tissue using quantitative real-time PCR (qRT-PCR). Harvested tissue was immediately frozen in liquid nitrogen and RNA was extracted using Spectrum Plant Total RNA kit (Sigma) per manufacturer's instructions. RNA samples were DNase treated and cDNA was synthesized with the Super Script III kit (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions. Quantitative real-time PCR was performed in triplicate with 25 ul reactions consisting of SYBR GREEN PCR Master Mix (Applied Biosystems, Foster City, CA, USA), 0.4 uM of each primer, and a 1:25 dilution of cDNA. Primer sets used for qRT-PCR analyses were as follows: ADF1 (ADF1reti2-s: 5' – ACTGGTACTTGATTATGTTTTTACATTGTG – 3', ADF1reti2-a: 5' – ACAAGACCGAAACACCGATAGAA – 3'); ADF2 (ADF2reti-1S: 5' – GCCAAAGTGAGAGACAAGATGATTT – 3', ADF2reti-1A: 5' – GAATCCCATCTAGTTCTCTCTTGAACCT – 3'); ADF3 (ADF3-3UTR-RT-S: 5' – TCGGTTGAATCAAACTTTTTCGT – 3', ADF3-3UTR-RT-N: 5' – GGTACCGTCACAGCAAACTTTAGG – 3'); ADF4 (ADF4-RT-3'UTR-S: 5' – GGTCTGTTCTCTGTGTCTATGTTACCTT – 3', ADF4-RT-3'UTR-N: 5' – GCAAACACAGCACAAACCTTGT – 3'); ADF5 (ADF5exon2RT: 5' – CGCCGGTGAAAGCTACCA – 3', ADF5exon2RT-a: 5' – AGACAGCGTAGCGACAATCATC – 3'); ADF6 (ADF6-3Sense: 5' – AGTGTTACGCGAACGAGCGA – 3', ADF6-3Anti: 5' – TCGCTCGTTCGCGTAACACT – 3'); ADF9 primer set #1 (ADF9reti-S: 5' – ATATAACGAAAGAACAAGAAGACA – 3', ADF9reti-A: 5' –

CACTCGTCGCCGTCTTCAA – 3'); ADF9 primer set #2 (ADF9x3Reti-S: 5' – ACGTCGAAGAGCGGACTGA – 3', ADF9x3Reti-A: 5' – CGGTGGCTTGAAGCTCGTA – 3'); ADF11 (ADF11reti3-S: 5' – GGTACGCAATAATCATCCCAATAGA – 3', ADF11reti3-A: 5' – GTTGCTTGAAATTGCCAGCTT – 3'); FLC (FLCrt2S: 5' – TCTTCCGGTGACTCTCCCA – 3', FLCrt2A: 5' – ATATCTTTTGGATTTTGATTTCAA – 3'); FT (FT-RT-S: 5' – GGCGCCAGAACTTCAACACT – 3', FT-RT-A: CGGGAAGGCCGAGATTG – 3'). UBIQUITIN 10 was used as an endogenous control for all qRT-PCR analyses, with the following primers: UBIQ10-real-S (5' – AGAACTTCATATGTTTCGTTTCATGTAA – 3') and UBI10-real-A (5' – GAACGGAAACATAGTAGAACACTTATTCA – 3'). UBIQUITIN 10 mRNA levels were set to 1.0 and target gene expression was then expressed as a Relative Quantity (RQ) that was computed using the 2^{-ddCT} method (Livak and Schmittgen, 2001).

Multigeneration study of fitness

To assess the fitness cost of *adf6-1* mutant plants, a heterozygous F_0 population was sown out onto soil and the wild-type and mutant allele frequencies were quantified across multiple generations (Gilliland et al., 1998; Assmusen et al., 1998). DNA was extracted from 100 random individuals from the F_1 and F_2 populations; seeds from these individuals (100 individuals per generation) were pooled together and used for analysis of the subsequent generation. Allele frequencies were assessed in the F_2 generation only using the ADF6 genotyping primers listed above. Deviations from expected frequencies

of the wild-type, heterozygous, and homozygous T-DNA mutant genotypes were calculated using a chi-squared (X^2) test.

Assaying the role of stress on fitness

In order to examine the response of ADF6 to stress, the Genevestigator Expression Data online database was used to identify potential biotic and abiotic stress elements that cause an up- or down-regulation in ADF6 expression. From these results, four stress factors were chosen: heat stress, exposure to high glucose concentration, nitrogen deprivation, and exposure to abscisic acid (ABA). Columbia wild-type seeds were sown out onto germination media, stratified for 48 hours, and transferred to a growth chamber where they were grown vertically. Eight days after vernalization, seedlings were transferred to their respective experimental condition. For glucose assay, seedlings were transferred to ½ MS media containing 3% glucose in place of sucrose. To test nitrogen deprivation, 8-day-old seedlings were transferred to ½ MS plates that completely lacked nitrogen; response to ABA was tested by transferring seedlings to ½ MS plates supplemented with 10 μM of ABA. For all experimental conditions, seedlings were exposed to their respective stressors for 48 hours. Root and shoot tissue were harvested separately for each experiment and total RNA was extracted using the Spectrum Plant Total RNA kit (Sigma). ADF6 expression levels were assayed through qRT-PCR. Control samples were also transferred onto fresh plates at the same time for each experimental condition to normalize for experimental design.

Results

ADF6 expression is temporally and spatially regulated through development

ADF6 expression pattern through development was screened using a transgenic plant line that was originally developed by Ruzicka et al. (2007). The GUS reporter gene construct was comprised of the promoter and first intron of ADF6 fused with the coding sequence of the β-glucoronidase gene (ADF6pi::GUS). Initial studies only examined ADF6pi::GUS expression at a few fixed time points. Herein, longitudinal studies were performed examining GUS expression at numerous time points during seedling and plant development. Seedlings began germinating approximately 36 hours post-stratification. Prior to germination, ADF6 levels were completely undetectable in seeds, but began to initiate in root tissue as the roots begin to emerge from the seed coat (Figure 3.1a). Once the seedlings had completely emerged from the seed coat, 72 hours after stratification, ADF6pi::GUS was strongly expressed throughout the entire seedling (Figure 3.1b). Five days post-stratification, ADF6 expression was localized to the root cambium and in the two cotyledon leaves (Figure 3.1c). As seedlings continued to develop, ADF6 expression was even more refined, with expression remaining strong in root tissue but being restricted to the outer fringes of developing rosette leaves and away from the meristem (Figure 3.1d-f).

This spatial and temporal regulation of ADF6 expression was also seen in developing reproductive tissue. There was no detectable ADF6 expression within early flower buds, around stage 10 (Figure 3.2a). As flowers enter into stage 11, ADF6 became strongly expressed in the stigma and correlated with the appearance of the stigmatic papillae (Figure 3.2b-c). However, in both stage 10 and stage 11, ADF6

expression was completely lacking in male reproductive organs. ADF6 was not upregulated in male reproductive tissue until stage 13, where we saw the emergence of strong expression within the stamens (Figure 3.2d-f). ADF6 was also expressed at moderate levels in mature pollen grains (Figure 3.2d, e, f), but not at earlier stages during microspore development (Figure 3.2a, b, c).

ADF6 defective plants are morphologically identical to wild-type through development

A T-DNA mutant containing an insertion 44 bp within the first intron of ADF6 was obtained from TAIR (SAIL-648-A03) and backcrossed twice to Columbia wild-type. ADF6 expression within the *adf6-1* mutant was assayed by measuring mRNA levels in 10-day-old seedlings using qRT-PCR. There were no detectable levels of ADF6 RNA within the *adf6-1* seedlings (Figure 3.3a); this was also validated for mature rosette leaf tissue (data not shown). The levels of ADF6 RNA were so low *adf6-1* should be classified as a null mutant. Despite this drastic knock-down of ADF6 expression, *adf6-1* developed normally and was morphologically indistinguishable from Columbia wild-type plants throughout all developmental stages (Figure 3.4). An overexpression transgenic line was generated by transforming the *adf6-1* mutant with a binary vector containing the coding sequence of ADF6 driven by the ACTIN2 promoter terminator (Figure 3.3C). Overexpressing ADF6 had no aberrant effect as these plants also developed normally and were indistinguishable from wild-type plants (Figure 3.4c,f,i,l). Previous work using a GFP tagged ADF6 protein has suggested that the ADF6 protein binds along actin filaments (Dong et al., 2001). Examining the actin cytoskeleton in *adf6-1*, however, shows that there are no detectable cytoskeletal defects in these mutants (Figure 3.5).

Before any conclusions could be made about the lack of essentiality or apparent redundancy of this gene, we needed to assess the cost on fitness for carrying a non-functional allele of ADF6. The cost on fitness was estimated by measuring allele frequencies across multiple generations, starting with a heterozygous F_0 population of plants. The progeny from this F_0 generation (the F_1 generation) were sown out and seeds were harvested and pooled together from 100 randomly selected F_1 plants. In the subsequent F_2 generation, 92 random individuals were genotyped to determine the frequency of wildtype allele to mutant T-DNA ADF6 allele. In a selfing plant, as long as there is no selective pressure on the allele in question, then the ratio of wild-type to heterozygous to homozygous mutant should be 3:2:3 at the F_2 generation (Gilliland et al., 1998; Asmussen et al., 1998). Using the designation A_1 to represent the wild-type ADF6 allele and A_2 to represent the mutant ADF6 allele, in the F_2 generation we would expect 37.5% of individuals to be A_1A1, 25% of individuals to be A_1A_2, and 37.5% of individuals to be A_2A_2. What we find is that the number of A_1A_1, A_1A_2, and A_2A_2 individuals in the F_2 generation is exactly as expected ($p = 0.5247$; Table 3.1). As such, we can conclude that there is no strong detectable cost on fitness to be a homozygous *adf6-1* mutant and that this gene may represent a non-essential gene. To rule out the possibility that another ADF protein variant may be up-regulated in *adf6-1* to compensate for the loss of ADF6, we screened expression of 10 vegetative ADF protein variants in 10-day-old seedlings using qRT-PCR. This is not an unreasonable expectation considering that actin *ACT7* mRNA is up-regulated several fold in plants lacking two other actins ACT2 and ACT8 (Kandasamy et al., 2012). Expression levels for the other 8 vegetative ADFs were identical in the *adf6-1* mutant as in wild-type, ruling out the

possibility of that increased expression of another ADF compensated for the loss of ADF6 (Figure 3.6). Additionally, overexpressing ADF6 did not effect the expression of any of the other 8 ADF mRNAs (Figure 3.6).

ADF6 responds to environmental stress but plays an unknown role in stress response

While *adf6-1* resembles a wild-type plant when grown under normal conditions, if it is a protein that is mainly used for stress response, there may be differential ability to handle certain stress conditions between wild-type and *adf6-1* plants. To explore whether ADF6 could be a component of any stress response pathways, we used the online database Genevestigator to identify abiotic and biotic stress factors that cause ADF6 expression to be either up or down regulated. The search revealed that there were only a few stresses and phytohormones that resulted in a drastic up- or down-regulation of ADF6. From these, we decided to test six of the identified conditions: high glucose concentrations, heat stress, abscisic acid (ABA), nitrogen starvation, UV exposure, and osmotic stress (Table 3.2). ADF6 response to only four of these conditions (nitrogen starvation, high glucose concentration, heat stress, and exposure to ABA) was quantified by exposing the plants to each condition for 48 hours. After 48 hours, tissue was harvested, RNA extracted, and *ADF6* mRNA levels measured through qRT-PCR. *ADF6* was not up or down regulated in either the glucose stress or ABA stress experiments using our conditions, contrary to the array data at Genevestigator. However, ADF6 was significantly up regulated in shoot tissue in a heat stress experiment where plants were grown for 48 hours at 37^{o}C (Figure 3.7a-b). Additionally, ADF6 was slightly downregulated during nitrogen starvation in both root and shoot tissue (Figure 3.7a-b).

However, for all stress conditions tested, we found that the growth of *adf6-1* mutant plants was equally sensitive or insensitive to the particular stress condition as wild-type plants (Table 3.2). So while ADF6 expression does respond to a few specific stress conditions, loss of ADF6 does not result in a morphological change in growth after stress.

Knocking down ADF6 expression, but not overexpression, suppresses early flowering in adf9-1

Double mutants were generated by crossing *adf6-1* with mutants defective in expression of its two closest paralogs in subclass III, ADF5 and ADF9. The *adf5/6* double mutant was morphologically identical to the *adf5-1* mutant, which is characterized by a marked increase in the number of rosette leaves (Figure 3.8b). By contrast, we found that a double mutant between *adf9-1* and *adf6-1* completely suppressed the early flowering phenotype of *adf9-1* (Figure 3.8a). In other words, in *adf9-1* plants in which ADF6 was also knocked down, the plants had a wild-type phenotype instead of the *adf9-1* phenotype. The genotype of these double mutants plants was confirmed repeatedly to ensure that the wild type phenotype was not due to the presence of a wild type *ADF9* allele. This result is both interesting and complex, for it appears that ADF6 expression is necessary to reveal the *adf9-1* early flowering phenotype. The *adf9-1* mutant has other distinct developmental phenotypes besides early flowering, such as longer leaf petioles, and early excessive branching of the lateral influorescences (Burgos-Riviera et al., 2009). For all three phenotypes, the *adf9/6* double mutant was also morphologically indistinguishable from the *adf6-1* mutant or wild-type (Figure 3.8). Adding to the complexity of this result is the fact that the overexpression of ADF6 in *adf9-1* neither

suppressed the *adf9-1* phenotype nor exacerbated the early flowering phenotype (Figure 3.8a).

The molecular basis of the *adf9-1* phenotype is well characterized and due in part to the lowered expression of flowering time regulators (Burgos-Rivera et al., 2000). The expression levels of two of these important regulators of flowering time, *FLC* and *FT*, were quantified on 10-day old shoot tissue using qRT-PCR (Figure 3.9). FLC is a master repressor of flowering while FT is an important activator of flowering as it integrates signals from multiple pathways to initiate flowering (Amasino, 2010). The expression levels of these three genes in *adf9-1* correlated with what has been previously published, with *FLC* being down-regulated and *FT* being up-regulated. Screening *FLC* and *FT* levels in *adf6-1* revealed that these genes were expressed at the same levels as wild-type (Figure 3.9). In the *adf9/6* double mutant, we find that both FLC and FT are all returned to wild-type levels (Figure 3.9). Hence, the molecular phenotype agrees with the normal flowering phenotype in the *adf9/6* double mutant.

Discussion

The higher plant ADF gene family is comprised of four classes of highly diverse actin-binding proteins. Previous work has demonstrated the specialized roles of family members within different subclasses (e.g., subclass I, II, and III). The goal of this study was to dissect and characterize the functional role of the single subclass IV Arabidopsis ADF variant, ADF6. ADF6 is the sole member of an ADF subclass that has been conserved in angiosperms for an estimated 250 million years. Furthermore, it is the only

ubiquitously expressed ADF variant within the gene family. Based on these findings, we hypothesized that ADF6 represents an essential gene within Arabidopsis.

As just mentioned, at the onset of this study it was suggested that ADF6 expression was ubiquitous and the first questions we wanted to address was whether expression remained constant throughout development. One of the most exciting findings was that ADF6 expression was found to be both spatially and temporally regulated through development. This further bolstered support of the hypothesis that ADF6 represents an essential gene as it is clearly being turned on and off during specific points in development. Given this regulated expression pattern, it came as a surprise that the *adf6-1* mutants exhibited no developmental or morphological defects. Additionally, there were no detectable cytoskeletal defects in the *adf6-1* mutant. By fluorescently tagging ADF6, Dong et al. (2001) had previously demonstrated that ADF6 binds along actin filaments and possibly localizes to the nucleus. As such, it was expected that, at the minimum, that there would be some impact on the actin cytoskeleton when knocking down ADF6. However, our limited results suggest that the actin cytoskeleton may be completely normal in the *adf6-1* mutant with no obvious differences detected as compared to wild-type.

Given ADF/CFL's role in stress response coupled with the finding that *adf6-1* developed like wild-type when grown under normal conditions, it was then hypothesized that ADF6 may be a protein that is used in stress response pathways rather than developmental processes. The results from the Genevestigator search revealed that few stress conditions resulted in a 1 or 2 fold up or down regulation of ADF6 (compared to wild-type levels) and most conditions only resulted in a weak response from ADF6.

Despite this, six conditions were chosen to test the hypothesis that ADF6 is a stress response protein. Out of these six conditions, five were described as causing an upregulation of ADF6 and one (ABA) was described as causing a 1-fold decrease in ADF6 expression. For all conditions tested, growth of *adf6-1* mutant plants responded exactly as a wild-type plants.

Due to these results, we decided to quantify the degree of ADF6 response through qRT-PCR for four of these conditions in order to verify the data on Genevestigator. Exposing wild-type plants to high glucose concentrations and to ABA had no effect on ADF6 expression leading us to conclude that ADF6 is not involved in response to these two stress conditions. This is further supported by the initial report by Dong et al. (2001) that demonstrated the lack of ADF6 response to ABA using histochemical assays. Two of the conditions, however, did show a response in ADF6 expression: heat stress and nitrogen starvation. According to Genevestigator, ADF6 is upregulated 1-fold when exposed to nitrogen limiting conditions. Interestingly, we found the exact opposite as ADF6 is actually down-regulated when wild-type plants are switched to nitrogen deplete media for 48 hours. The results from the heat stress experiment did confirm the Genevestigator results as exposing wild-type plants to 37 ^{o}C for 48 hours caused nearly a 3-fold increase in ADF6 expression. Despite these findings, as mentioned above, *adf6-1* still responded as a wild-type plant to nitrogen starvation and heat stress. Therefore, although these two conditions elicit a response in ADF6 expression, ADF6 does not appear to play a vital role in the response pathway to these two stressors.

As an independent assay of whether ADF6 is a non-essential gene within Arabidopsis, we performed a multigenerational allele frequency experiment to determine

the cost on fitness for carrying a defective ADF6 allele. Previous studies conducted on actin mutants in Arabidopsis yielded results similar to those documented here for ADF6 where single actin mutants such as *act4-1* and *act2-1* did not exhibit any gross developmental or morphological phenotypes (Gilliland et al., 1998). As with the ADF gene family, actin also belongs to a large gene family within Arabidopsis and is comprised 8 family members that partition into two ancient classes that have vegetative and reproductive specific expression patterns (Meagher et al., 1999). Actin protein variants are highly conserved across kingdoms, and therefore it was expected that knocking out just one variant would have a large impact on development. Although there were no gross phenotypes, by quantifying the allele frequency of a mutant allele of actin across multiple generations, it was found that there is a significant cost of fitness to have a defective allele of actin. In other words, after inbreeding a heterozygous parent and passing at least 100 plants in each successive generation, we did not observe the expected 3:2:3 by the second generation. There were statistically significant reduced numbers of the *act4-1* and *act2-1* mutant alleles in the F2 generation. Looking for parallel evidence of the deleterious impact of the *adf6-1* allele on plant survival over multiple generations, we performed the same experiment on a heterozygous *adf6-1* line (F_0 generation). After two generations, it was found that the number of wild-type, heterozygote, and mutant genotypes were exactly as expected for a selfing organism indicating that there is no detectable cost on fitness for carrying a defective allele of ADF6. All of these results combined lead us to conclude that ADF6 is likely a non-essential member of the Arabidopsis ADF gene family.

Despite its being a non-essential protein, to further attempt to dissect the functional role of ADF6 within Arabidopsis, we decided to generate double mutants in the hopes that this would unmask the functional role of ADF6. Since there are no other subclass IV ADF variants, *adf6-1* was crossed with mutants from subclass III, which contains the two closest paralogs to ADF6. Both *adf5-1* and *adf9-1* exhibit striking developmental phenotypes with *adf5-1* showing altered rosette leaf morphology and *adf9-1* flowering early. Knocking down both ADF5 and ADF6 shows that these proteins are in completely separate pathways as the double mutant exhibited the same developmental phenotypes as the single *adf5-1* mutant. One of the most fascinating results came about from crossing *adf6-1* and *adf9-1* as knocking down ADF6 completely suppressed the early flowering phenotype and other phenotypes of *adf9-1*. Screening two key flowering time genes, FLC and FT, shows that all of these important regulators of flowering time are returned to wild-type levels in the *adf6/9* double mutant. The fact that gene expression is returned to wild-type levels in the *adf9/6* double mutant strongly suggests that ADF6 must act antagonistically with ADF9 in regulating flowering time gene expression. One way that this can occur is through differential regulation of cytoskeletal dynamics.

A model to explain the genetic interaction phenotype of adf6-1/adf9-1

In animal cells, cytoskeletal dynamics have been shown to be important in the regulation of downstream targets of serum response factor (SRF), a MADS box transcription factor like FLC. SRF remains bound to its target genes and requires the activation by its cofactor MAL, which normally binds to G-actin in the cytoplasm, in order to activate its bound targets (Guettler *et al.* 2008). A signal cascade that is

triggered by serum causes an increase in actin polymerization, which decreases the G-actin pool and subsequently releases MAL. Upon release, MAL translocates into the nucleus, binding and activating SRF (Posern and Treisman 2006, Zheng *et al.* 2009). King et al. (in prep) provide strong evidence that FLC may be regulated in a similar manner. In this study, the authors showed that increasing the G-actin pool within *adf9-1*, either by overexpression ACT2 or by the use of cytochalaisin D, resulted in suppression of early flowering (King et al., in prep). The hypothesis that we propose here is that the antagonistic activity of ADF9 and ADF6 on flowering time is most likely due to these two protein variants having different effects on actin filament turnover. It is possible that, much like ADF and CFL1 in mammalian cells (Bernstein and Bamburg, 2010), ADF6 is more efficient at nucleating filaments whereas ADF9 is a more potent filament severer. Additionally, we propose that much like MAL, there is a repressor of FLC or a transcription factor that targets a repressor of FLC that remains bound to G-actin in the cytoplasm (Figure 3.10). Using a version of this model to explain the wild type state, ADF9 and ADF6 work together in actin filament turnover and the G-actin pool remains at normal levels, keeping the repressor or transcription factor bound to G-actin (Figure 3.10A). Reducing ADF9 levels in *adf9-1* results in a reduction of filament severing and prevents the replenishment of the G-actin pool. When this is coupled with the nucleating activity of the still active ADF6, this ultimately results in a reduction of the G-actin pool and releases the bound repressor or transcription factor (Figure 3.10B). However, when ADF6 is also knocked out along with ADF9, the G-actin pool remains unaffected. In other words, the loss of filament severing by ADF9 is compensated by a reduction in filament nucleation due to the loss of ADF6. These two outcomes result in a G-actin

pool that largely remains the same and therefore the repressor or transcription factor remains bound to G-actin as it would in a wild-type plant (Figure 3.10D). This can also explain why there is no detectable flowering phenotype in an *adf6-1* mutant as knocking down ADF6 would have little effect on the G-actin pool (since it would remain large) and the repressor of transcription factor would remain bound to G-actin (Figure 3.10C).

The goal at the onset of this study was to detail the functional role of the single subclass IV Arabidopsis protein variant, ADF6. Here, we have shown that ADF6 expression is dynamic through development as it is both temporally and spatially regulated. Our analyses also provide strong evidence that ADF6 may play a non-essential role in development and stress response. Crossing *adf6-1* with *adf9-1* did uncover a hidden clue that ADF6 acts antagonistically with ADF9 to indirectly regulate flowering time. However, this role is entirely dependent on ADF9 as the *adf6-1* mutant flowers normally. Additionally, knocking out ADF6 had no cost on fitness, leading to the final conclusion that ADF6 is a non-essential member of the ADF gene family. However, such a conclusion still begs the question as to why ADF6 has been preserved in the Arabidopsis lineage if it has no essential function. Perhaps ADF6 is under latent selection, and we have not yet tried the relevant stress condition. Alternatively, the Arabidopsis lineage may have accrued other genetic changes that have made ADF6 non-essential.

References Cited

Amasino, R. (2010) Seasonal and developmental timing of flowering. The Plant Journal. 61:1001–1013.

Andrianantoandro, E., and Pollard, T.D. (2006) Mechanism of actin filament turnover by severing and nucleation at different concentrations of ADF/Cofilin. Mol. Cell. 24: 13–23.

Asmussen, M.A., Gilliland, L.U., and Meagher, R.B. (1998) Detection of deleterious genotypes in multigenerational studies. II. Theoretical and experimental dynamics with selfing and selection. Genetics. 149:727-737.

Bamburg, J R. (1999) Proteins of the ADF/Cofilin Family: Essential Regulators of Actin Dynamics. Annual Review of Cellular Developmental Biology. 15:185-230.

Bernstein, B.W., and Bamburg, J.R. (2010) ADF/Cofilin: a functional node in cell biology. Trends in Cell Biology. 20(4): 187-195.

Bettinger, B.T., Gilbert, D.M., and Amberg, D.C. (2004) Actin up in the nucleus. Nature Reviews Molecular Cell Biology. 5: 410-415.

Burgos-Rivera, B., Ruzicka, D.R., Deal, R.B., McKinney, E.C., King-Ried, L., and Meagher, R.B. (2008) Actin depolymerizing factor9 controls development and gene expression in *Arabidopsis*. Plant Mol Biol. 68(6):619-632.

Chan, C., Beltzner, C.C., and Pollard, T.D. (2009) Cofilin dissociates Arp2/3 complex and branches from actin filaments. Current Biology. 19(7): 537-545.

Chua, B.T., Volbracht, C., Tan, K.O., Li, R., Yu, V.C., and Li, P. (2003) Mitochondrial translocation of cofilin is an early step in apoptosis induction. Nat. Cell Biol. 5:1083–1089.

Clement, M., Ketelaar, T., Rodiuc, N., Banora, M.Y., Smertenko, A., Engler, G., Abad, P., Hussey, P., and Engler, J.D. 2009. Actin-Depolymerizing Factor2-Mediated Actin Dynamics Are Essential for Root-Knot Nematode Infection of *Arabidopsis*. Plant Cell. 21(9), 2963-2979.

Clough, S J, and Bent, A F. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. The Plant Journal. 16, 735-743.

Daher, F.B., van Oostende, C., and Geitmann, A. (2011) Spatial and temporal expression of actin depolymerizing factors ADF7 and ADF10 during male gametophyte development in *Arabidopsis thaliana*. Plant Cell Physiol. 52(7): 1177–1192.

Dong, C.H., Kost, B., Xia, G., and Chua, M.H. (2001) Molecular identification and characterization of the *Arabidopsis AtADF1*, *AtADF5* and *AtADF6* genes. Plant Molecular Biology. 45:517–527.

Feng, Y., Liu, Q., and Xue, Q. (2006) Comparative study of rice and Arabidopsis Actin-depolymerizing factors gene families. Journal of Plant Physiology 163:69-79

Guettler, S., Vartiainen, M.K., Miralles, F., Larijani, B. and Treisman, R. (2008) RPEL Motifs Link the Serum Response Factor Cofactor MAL but not Myocardin to Rho Signaling via Actin Binding. *Molecular and Cellular Biology*, **28**, 732-742.

Gilliland, L.U., McKinney, E.C., Asmussen, M.A., and Meagher, R.B. (1998) Detection of deleterious genotypes in multigenerational studies. I. Disruptions in individual Arabidopsis actin genes. Genetics. 149:717-725.

Han, L., Stope, M.B., Lopez de Jesus, M., Oude Weernink, P.A., Urban, M., Wieland, T., Rosskopf, D., Mizuno, K., Jakobs, K.H., and Schmidt, M. (2007) Direct stimulation of receptor-controlled phospholipase D1 by phospho-cofilin. EMBO J. 26:4189–4202.

Henty, J.L., Bledsoe, S. W., Khurana, P., Meagher, R.B., Day, B., Blanchoin, L., and Staiger, C.J. (2011) Arabidopsis Actin Depolymerizing Factor4 Modulates the Stochastic Dynamic Behavior of Actin Filaments in the Cortical Array of Epidermal Cells. The Plant Cell. Vol. 23:3711–3726.

Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W. (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. 6:3901–3907.

Jeong, Y.M., Mun, J.H., Kim, H., Lee, S.Y., and Kim, S.G. 2007. An upstream region in the first intron of petunia actin-depolymerizing factor 1 affects tissue-specific expression in transgenic Arabidopsis (*Arabidopsis thaliana*). Plant J. 50, 230-239.

Kandasamy, M.K., McKinney, E.C., and Meagher, R.B. (2002) Functional nonequivalency of actin isovariants in Arabidopsis. Mol. Biol. Cell. 13: 251-261.

Lehman, N., Di Fulvio, M., McCray, N., Campos, I., Tabatabaian, F., and Gomez-Cambronero, J. (2006) Phagocyte cell migration is mediated by phospholipases PLD1 and PLD2. Blood. 108:3564–3572.

Livak, K.J., and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 25:402-408.

Machesky, L.M., and Insall, R.H. (1999) Signaling to actin dynamics. J. Cell Biol. 146:267–72

Meagher, R.B., McKinney, E.C., and Kandasamy, M.K. (1999a) Isovariant dynamics expand and buffer the responses of complex systems: The diverse plant actin gene family. The Plant Cell. 11:995-1005.

Meagher, R.B., McKinney, E.C., and Vitale, A.V. (1999b) The evolution of new structures: Clues from plant cytoskeletal genes. Trends Genet. 15: 278–284.

Mun, J.H., Yu, H.J., Lee, H.S., Kwon, Y.M., Lee, J.S., Lee, I., and Kim, S.G. (2000) Two closely related cDNAs encoding actin-depolymerizing factors of petunia are mainly expressed in vegetative tissues. Gene. 257:167-176.

Mun, J.H., Lee, S.Y., Yu, H.J., Jeong, Y.M., Shin, M.Y., Kim, H., Lee, I., and Kim, S.G. (2002) Petunia actin-depolymerizing factor is mainly accumulated in vascular tissue and its gene expression is enhanced by the first intron. Gene. 292:233-243.

Ohta, Y., Nishida, E., Sakai, H., and Miyamoto, E. (1989) Dephosphorylation of cofilin accompanies heat shock-induced nuclear accumulation of cofilin. The Journal of Biological Chemistry. 264:16143-16148.

Okreglak, V., and Drubin, D.G. (2007) Cofilin recruitment and function during actin-mediated endocytosis dictated by actin nucleotide state. The Journal of Cell Biology. 178(7):1251-1264

Posern, G. and Treisman, R. (2006) Actin' together: serum response factor, its cofactors and the link to signal transduction. *Trends in Cell Biology*, **16**, 588-596.

Poukkula, M., Kremneva, E., Serlachius, M., and Lappalainen, P. (2011) Actin-Depolymerizing Factor Homology Domain: A conserved fold performing diverse roles in cytoskeletal dynamics. Cytoskeleton. 68:471-490.

Ruzicka, D.R., Kandasamy, M.K., McKinney, E.C., Burgos-Rivera, B., and Meagher, R B. (2007) The ancient subclasses of Arabidopsis ACTIN DEPOLYMERIZING FACTOR genes exhibit novel and differential expression. The Plant Journal. 52: 460-472.

Salvarezza, S.B., Deborde, S., Schreiner, R., Campagne, F., Kessels, M.M., Qualmann, B., Caceres, A., Kreitzer, G., and Rodriguez-Boulan E. (2009) LIM Kinase 1 and Cofilin Regulate Actin Filament Population Required for Dynamin-dependent Apical Carrier Fission from the Trans-Golgi Network. Molecular Biology of the Cell. 20(1): 438-451.

Staiger, C.J. (2000) Signaling to the actin cytoskeleton in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 51:257–288.

Tian, M.Y., Chaudhry, F., Ruzicka, D.R., Meagher, R.B., Staiger, C.J., and Day, B. (2009) *Arabidopsis* Actin-Depolymerizing Factor AtADF4 Mediates Defense

Signal Transduction Triggered by the *Pseudomonas syringae* Effector AvrPphB. Plant Physiology. 150(2): 815-824.

Tholl, S., Moreau, F., Hoffmann, C., Arumugam, K., Dieterle, M., Moes, D., Neumann, K., Steinmetz, A., and Thomas, C. (2011) Arabidopsis actin-depolymerizing factors (ADFs) 1 and 9 display antagonist activities. FEBS Lett. 585:1821–1827.

Wang, Y.S., Motes, C.M., Mohamalawari, D.R. and Blancaflor, E.B. (2004) Green fluorescent protein fusions to Arabidopsis fimbrin 1 for spatio-temporal imaging of F- actin dynamics in roots. Cell Motil Cytoskel. 59:79-93.

Zheng, B., Han, M., Bernier, M. and Wen, J.-k. (2009) Nuclear actin and actin-binding proteins in the regulation of transcription and gene expression. *FEBS Journal*, **276**, 2669- 2685.

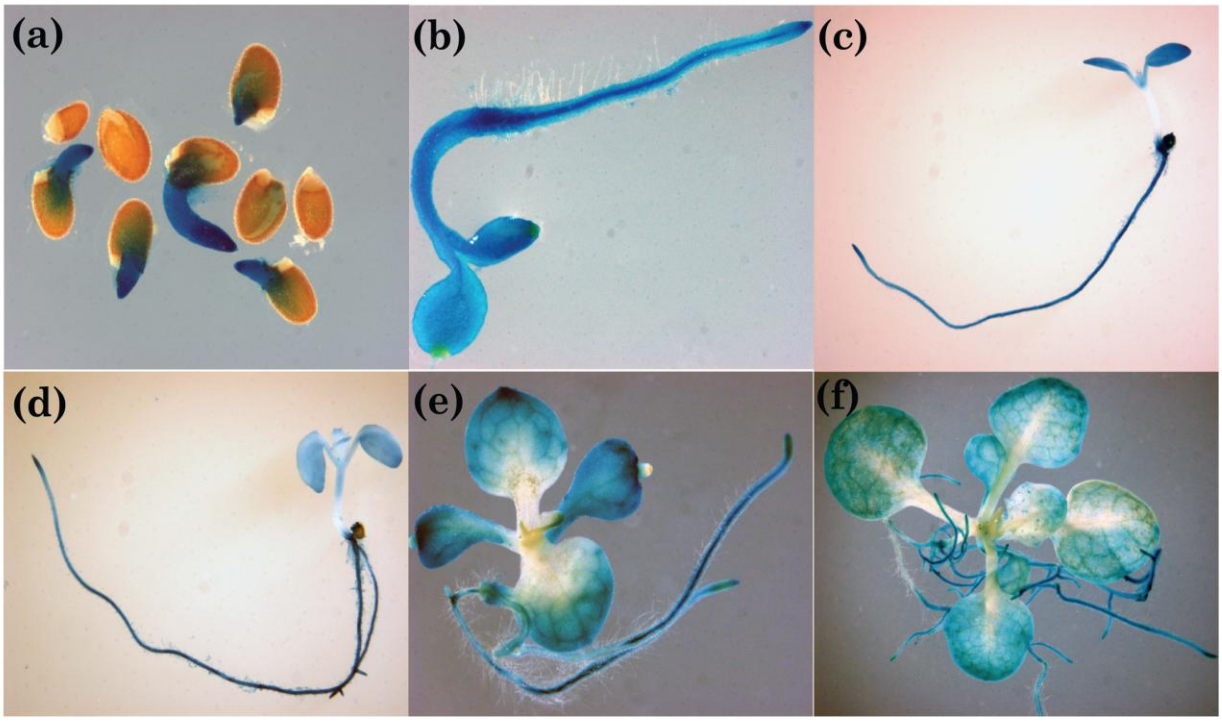


Figure 3.1: ADF6 expression is temporally and spatially regulated during early seedling development. (a – f) Histochemical staining of a translational fusion line consisting of the GUS reporter gene fused with the ADF6 promoter and first intron. (a) Germinating seedlings 36-hours post stratification and transfer to 22^{o}C growth chamber. (b) Seedling 72 hours post stratification. (c) 5 day old seedling. (d) 7 day old seedling (e) 10 day old seedling (f) 13 day old seedling.

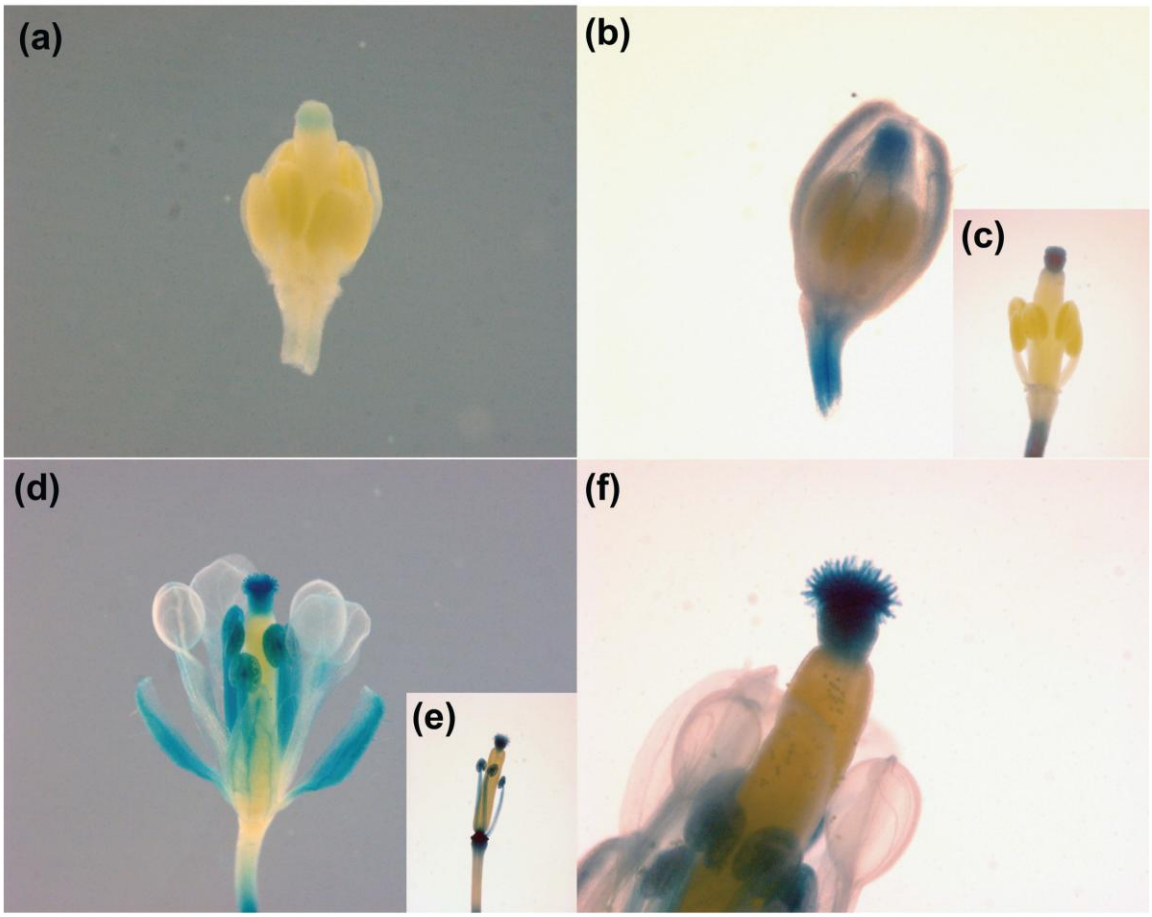


Figure 3.2: ADF6 expression through reproductive development. (a) Flower bud stage 10. (b – c) Stage 11 flowers. (d – f) Stage 13 flowers. (c, e) Flower petals were removed for easier viewing of anthers and stigma. (f) Magnified (20X) view of stage 13 flower to detail ADF6 expression in pollen. Developmental staging of flowers based on Smyth et al., 1990.

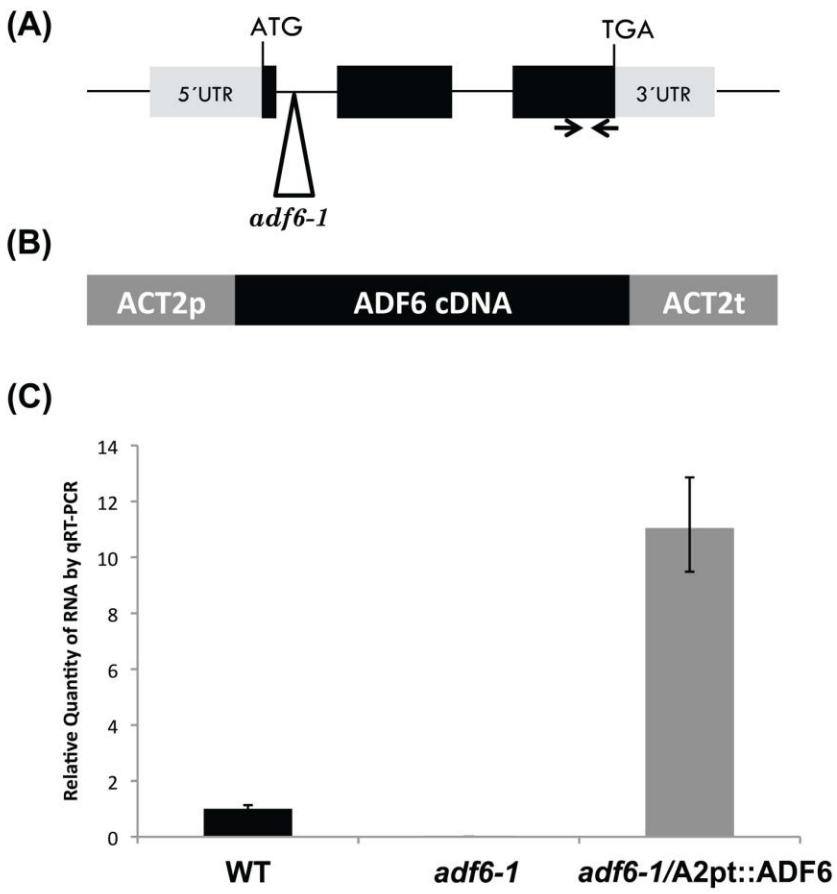


Figure 3.3: Insertion site and ADF6 expression levels in *adf6-1* homozygous mutant. (A) The T-DNA insertion for *adf6-1* (SAIL-648-A03) is located in the first intron of ADF6. Primers used for qRT-PCR analyses targeted the distal end of the third exon, as indicated by arrows. (B) Overexpression of ADF6 coding sequence within *adf6-1* was achieved by cloning ADF cDNA into an ACTIN2 promoter-terminator cassette. (c) Relative quantity of ADF6 RNA was measured on WT, *adf6-1*, and *adf6-1*/A2pt::ADF6 10-day old whole seedlings by qRT-PCR. ADF6 RNA was undetectable in *adf6-1* and was up-regulated approximately 11-fold in *adf6-1*/A2pt::ADF6.

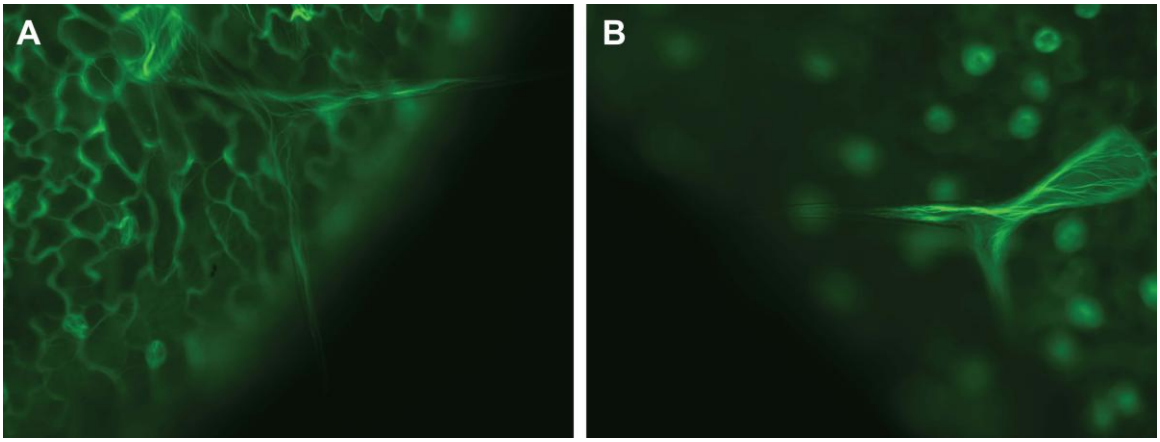


Figure 3.4: Actin cytoskeleton appears normal in the *adf6-1* mutant. Actin filament structure was examined using the F-actin reporter fABD:GFP as described by Wang et al., 2004. (A) Actin filaments in Columbia wild-type trichomes at 20X magnification. (B) Actin filaments in *adf6-1* trichomes at 20X magnification.

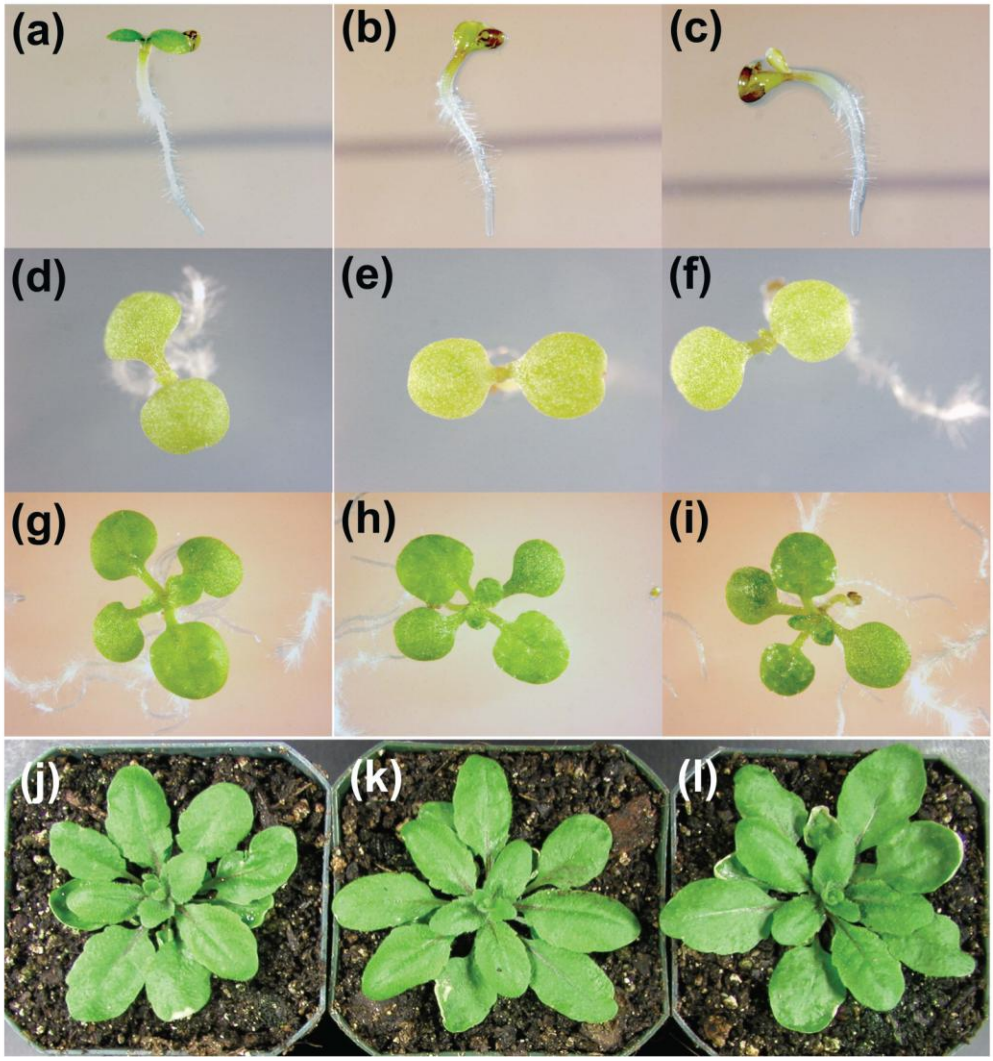


Figure 3.5: Early seedling and mature plant development is normal in *adf6-1*.

Plants vertically pictured at 4 days, 7 days, 13 days, and 31 days post-stratification. (a, d, g, j) Columbia wild-type plants. (b, e, h, k) *adf6-1* plants. (c, f, I, l) *adf6-1*/A2pt::ADF6 transgenic plants.

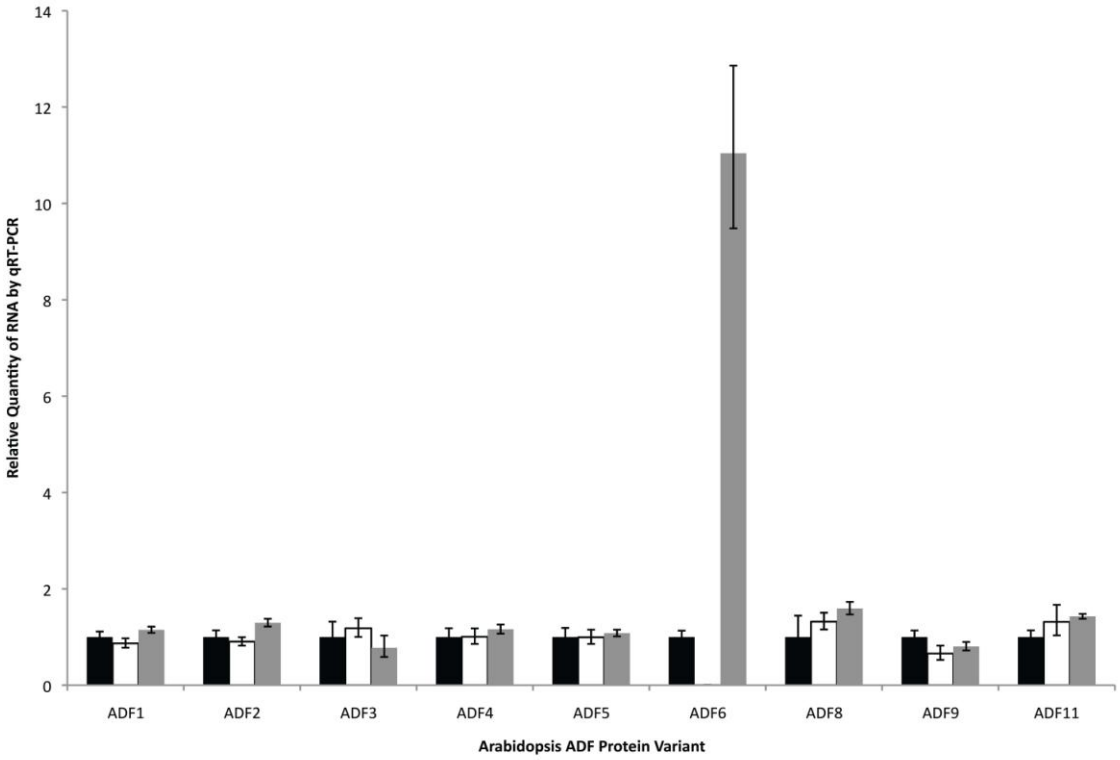


Figure 3.6: Expression of vegetative ADF gene family members unaffected in *adf6-1* mutant. The relative quantity of RNA for eight vegetative ADF gene family members were quantified by qRT-PCR: ADF1, ADF2, ADF3, ADF4, ADF5, ADF8, ADF9, and ADF11. Analyses were performed on 10-day old whole seedling tissue from Columbia wild-type (black bars), *adf6-1* (white bars), and the overexpression line *adf6-1*/A2pt::ADF6 (gray bars). Knocking down or overexpressing ADF6 had no effect on the expression of any of the ADF members screened.

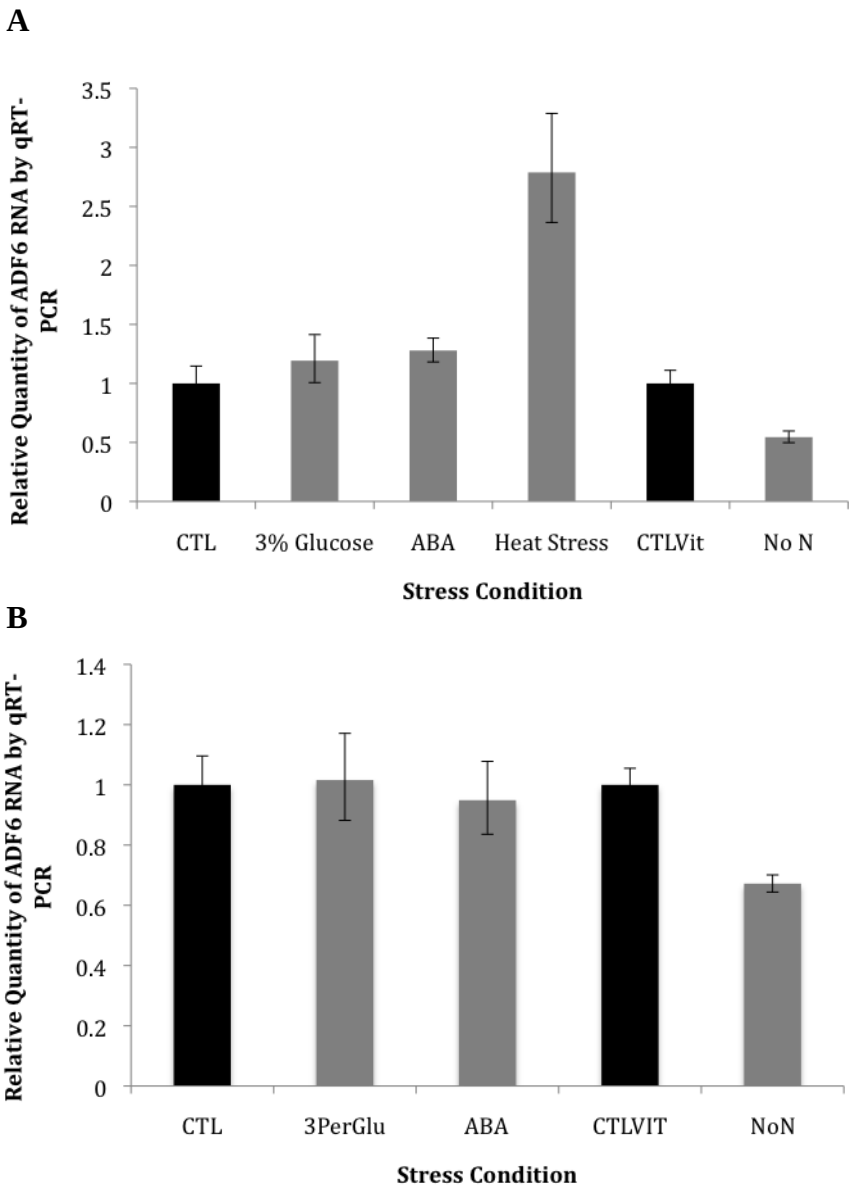


Figure 3.7: ADF6 is upregulated under heat stress conditions and down-regulated during nitrogen starvation. ADF6 response to four stress conditions was assessed on Columbia wild-type seedlings. Eight-day-old seedlings were transferred to stress conditions and maintained for 48 hours. (A) ADF6 expression in shoot tissue. (B) ADF6 expression in root tissue, only. Quantification of ADF6 for heat stress on roots is absent due to low RNA yield after extraction.

Figure 3.8: *adf5/6* and *adf5/9/6* are morphologically indistinguishable from *adf5-1* homozygous mutant. *adf6-1* homozygous mutant line was crossed with *adf5-1* homozygous mutant and *adf5/9* homozygous double mutant. Plants defective in ADF5 exhibit an increase in the number of rosette leaves with shorter petioles. The *adf5/6* double mutant and the *adf5/9/6* triple mutant exhibit the same altered rosette leaf morphology as *adf5-1* single mutant, suggesting that ADF6 and ADF5 are involved in different developmental pathways. Plants were photographed at 27 days post stratification.

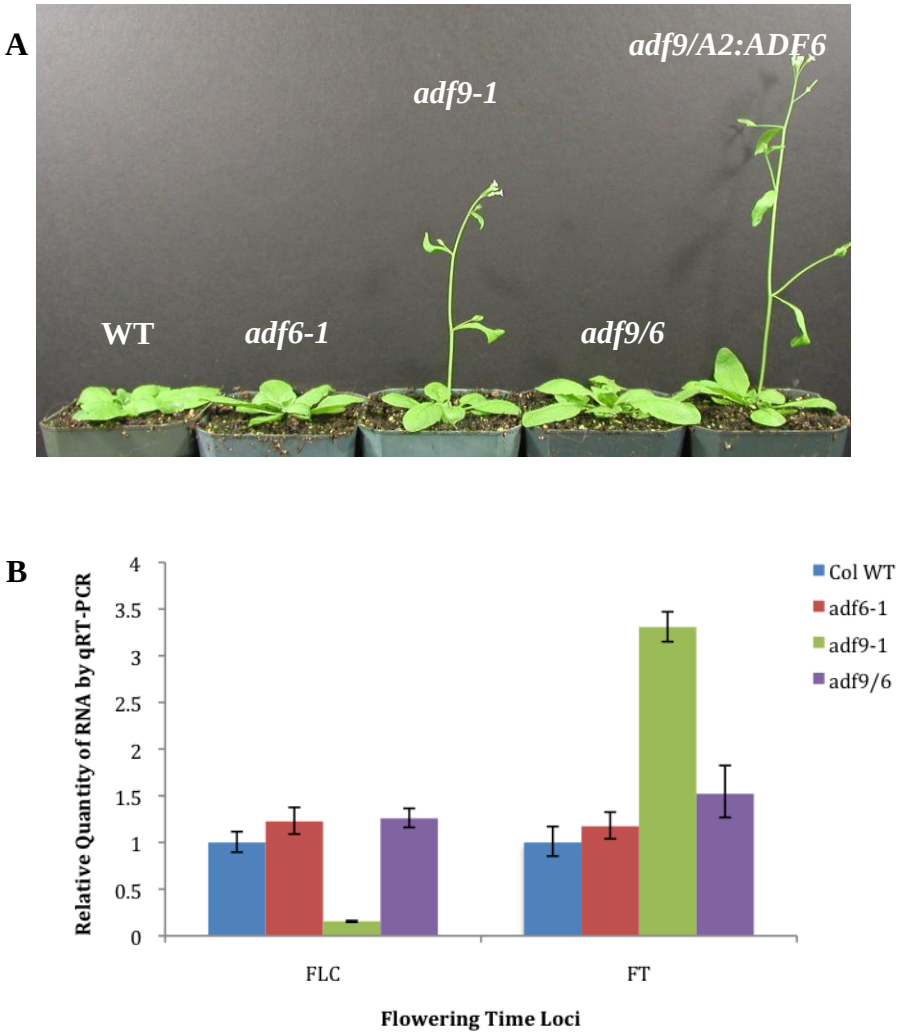
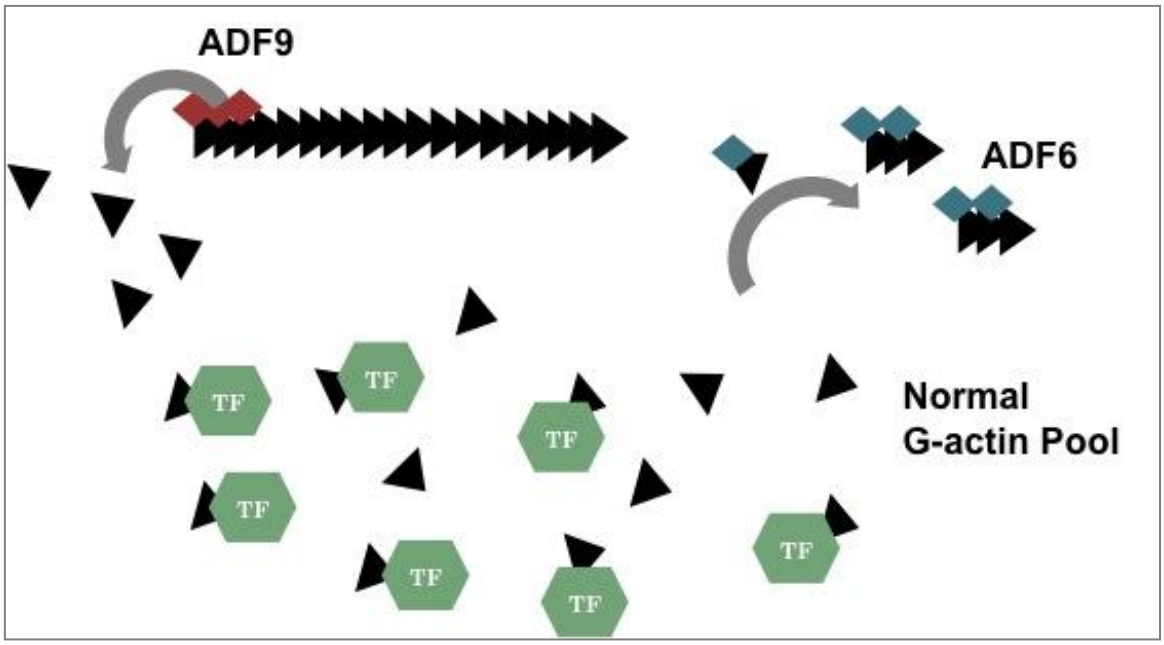


Figure 3.9: Knocking down ADF6 in *adf9-1* restores both the morphological and molecular early flowering phenotypes to wild-type levels. (A) *adf9-1* flowers significantly earlier than Columbia wild-type, *adf6-1*, and *adf9/6*. Overexpressing ADF6 within adf9-1 does not suppress the early flowering phenotype. Plants pictured at 27 days post stratification. (B) Expression levels of the master repressor of flowering, FLC, and an activator of flowering, FT, were quantified in wild-type (blue), *adf6-1* (pink), *adf9-1* (green), and *adf9/6* (purple). FLC and FT transcript levels are returned to wild-type levels in the *adf9/6* double mutant.

A



B

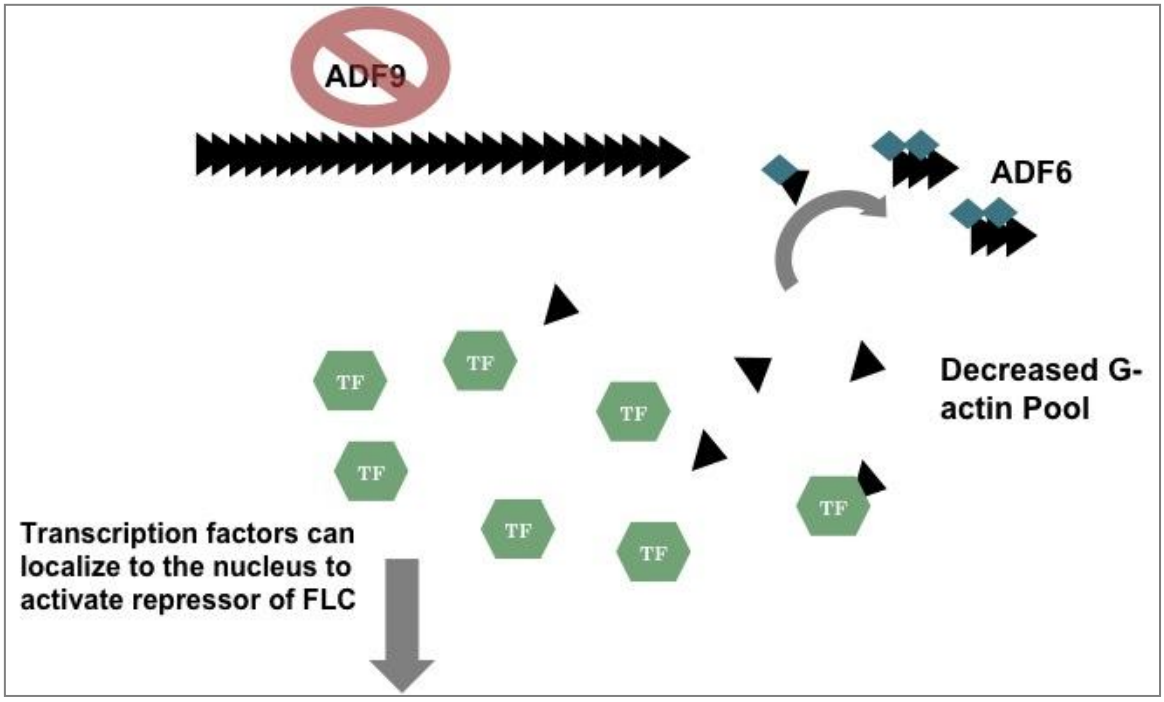
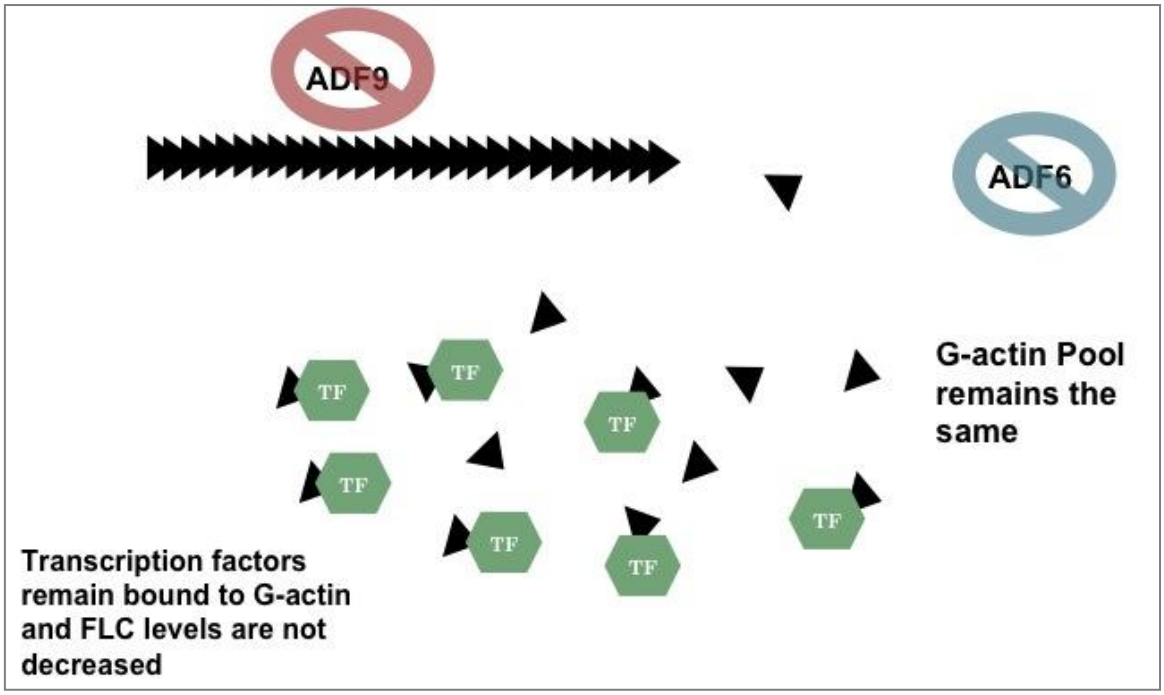


Figure 3.10A-B: See below for text

C



D

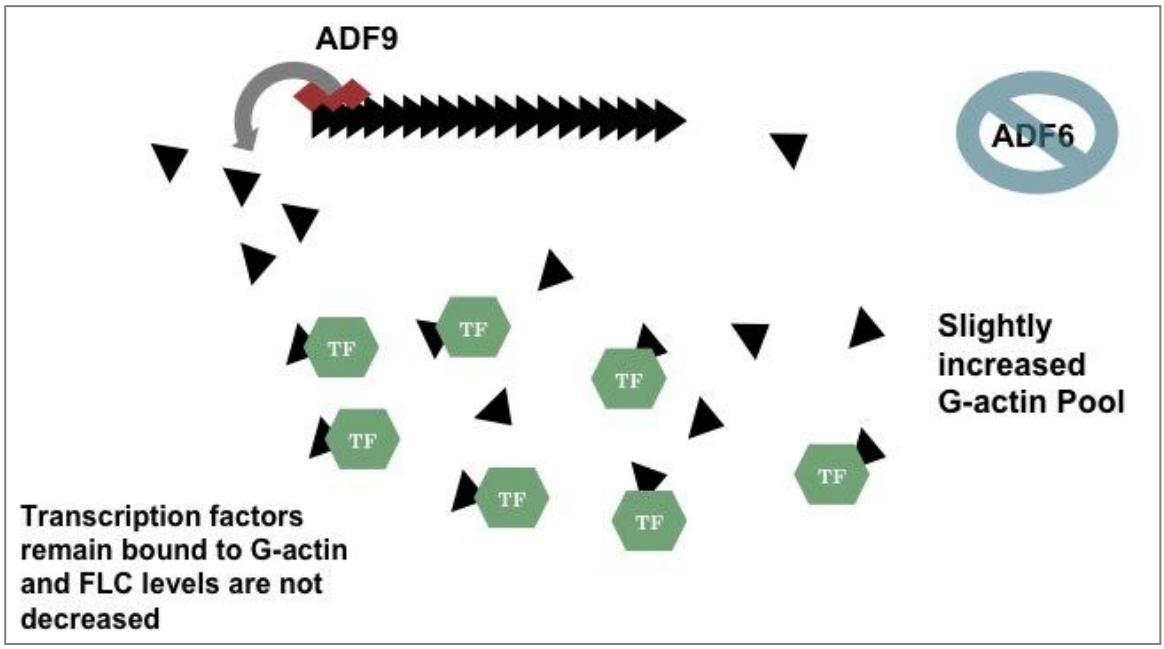


Figure 3.10C-D: See below for text

Figure 3.10: Model of the antagonistic behavior of ADF6 and ADF9 in the regulation of flowering time. In this model, a repressor of FLC or a transcription factor for a repressor of FLC remains bound to G-actin in the cytoplasm. Transcription factor/repressor (TF) is represented by green hexagons. Black triangles are actin monomers, pink diamonds represent ADF9, and ADF6 is represented as teal diamonds. (A) Hypothesized wild-type conditions. ADF9 functions more as a filament severing protein whereas ADF6 initiates nucleation of new filaments through the sequestration of G-actin. As such, the G-actin pool remains normal and the transcription factor/repressor remains bound to G-actin in the cytoplasm. (B) Knocking down ADF9 results in a lack of filament severing, resulting in the non-replenishment of the G-actin pool. The still existent nucleation process of ADF6 further depletes the G-actin pool and results in the TF translocating into the nucleus. (C) Hypothesized condition when both ADF6 and ADF9 are knocked down. Loss of both ADF6 and ADF9 results in a cessation of filament turnover; therefore, the G-actin pool is similar to what is found in wild-type conditions. (D) Knocking out only ADF6 has minimal impact on FLC expression as the G-actin pool only slightly increases.

Table 3.1: No strong cost on fitness for homozygous adf6-1 (-/-) mutant

Genotype	Observed # of plants (%)	Expected # of plants (%)	$(O - E)^2/E$
A_1A_1	34 (37)	34.5 (37.5)	0.007246377
A_1A_2	19 (20.6)	23 (25)	0.695652174
A_2A_2	39 (42.4)	34.5 (37.5)	0.586956522
	Total = 92 (100)	Total = 92 (100)	$X^2 = 1.29$
			$p = 0.5247$

Table 3.2: Abiotic and biotic stress assay

Stress Condition	Fold Change	Duration	Response to stress	Compared to WT
3% Glucose	1 fold up	48 hours	Delayed germination and development	Identical to WT
Heat stress (37°C)	2 fold up	48 hours	Yellowing of leaves	Identical to WT
ABA	1 fold down	48 hours	Reduced root length	Identical to WT
Nitrogen starvation	1 fold up	48 hours	Reduced root length	Identical to WT
UV Exposure	1 fold up	15 min.	Yellowing of leaves with later recovery	Identical to WT
Osmotic stress (high mannitol)	0.5 fold up	10 days	Delayed germination and development	Identical to WT

Six stress conditions tested based on Genevestigator search. Fold change column indicates the fold change in ADF6 expression in response to the particular condition and duration indicates the length of time a plant was exposed to a particular condition.

CHAPTER 4

A walk through evolutionary history: Reconstructing ancestral states of the *Arabidopsis thaliana* ACTIN-DEPOLYMERIZING FACTOR (ADF) Gene Family[1]

[1]Eileen M. Roy-Zokan, Elizabeth C. McKinney, Brunilis Burgos-Rivera, Daniel R. Ruzicka, and Richard B. Meagher. To be submitted to the Journal of Molecular Evolution.

Abstract

The actin cytoskeleton is involved in a variety of important cellular processes such as organelle movement, cell division, and chromatin remodeling. One important modifier of actin's activity is the actin-depolymerizing factor (ADF) family of proteins. These small proteins contribute to actin dynamics by binding to and severing monomers from F-actin filaments, regulating actin dynamics, and participating in nuclear actin function. The members of the *Arabidopsis thaliana* ADF gene family are comprised of eleven distinct variants that fall into four ancient subclasses. They provide for an interesting case study for gene family evolution due to potential functional divergence. The tissue-specific regulation and great diversity of protein sequences suggests both genes and protein products may have undergone subfunctionalization. Furthermore, recent investigations into Arabidopsis lines that were mutant for either ADF5 or ADF9 indicate that the various ADFs may each play a recently evolved and distinct role in the nuclear regulation of gene expression, suggesting neofunctionalization. Using ancestral state reconstruction we show that the Arabidopsis ADF variants have rapidly diverged in function after recent duplication events. The ancestral protein state of ADF5 and ADF9 failed to suppress the novel phenotypes of the *adf5-1* and *adf9-1* mutants. Additionally, the protein sequences from more ancestral nodes within the phylogeny failed to suppress the phenotypes from subclass II and subclass III mutants. The one exception was the root length phenotype in the subclass I mutant, *adf4-1*, suggesting that ADF4's role in root development may represent an ancestral function shared between ADF proteins.

Introduction

Genome and gene duplication events have had a profound impact in the evolution of angiosperm species. Numerous studies have focused their efforts on demonstrating the importance of duplication events in the evolution of transcription factors and kinases, but the role of duplications on the evolution of proteins that are involved in multiple processes crucial for both development and survival, such as actin and actin-binding proteins have received comparatively little study. However, it is known that the duplication of plant actin genes has had a major influence on the evolution of plant organ tissue (Meagher et al., 1999; Kandasamy et al., 2009). Therefore, understanding the evolution of the actin and various actin-binding proteins that contribute to actin dynamics can further enhance investigations into the evolution of plant organ systems. The actin cytoskeleton is an essential component of plant cells and is involved in a variety of important cellular processes such as organelle movement, cell division, and chromatin remodeling. One important modifier of the actin cytoskeleton is the actin-depolymerizing factor (ADF) family of proteins, which belong to a group of proteins that contain a conserved protein motif known as the actin-depolymerizing factor homology (ADF-H) domain. These small proteins contribute to actin dynamics by effecting actin filament turnover and have been recently implicated in such various cellular processes as intracellular trafficking, lipid and membrane metabolism, mitochondrial dependent apoptosis, chemotaxis, and the cytonuclear trafficking of actin into the nucleus (Ohta et al., 1989; Chua et al., 2003; Bettinger et al., 2004; Lehman et al., 2006; Han et al., 2007; Okreglak and Drubin, 2007; Chan et al., 2009; Bernstein and Bamburg, 2010).

The ADF gene family within higher plants phylogenetically partition into four ancient subclasses (Feng et al., 2006; Ruzicka et al., 2007) that are differentially regulated (Ruzicka et al., 2007). The number of ADF proteins within a genome varies between angiosperm species; however, each species contains at least one ADF representative in each subclass. An updated phylogenetic analysis of plant ADF variants that included the single *Physcomitrella patens* ADF and two ADF protein variants from the lycopod *Selaginella moellendorffii* suggests that there was an ancient divergence between ADF variants approximately 450 million years ago (MYA) that predates the diversification of the four subclasses found in angiosperms (Figure 4.1) (Roy-Zokan et al., in prep). The members of the *Arabidopsis thaliana* ADF gene family (11 variants in total) provide for an interesting case study for gene family evolution due to the potential of functional divergence. The tissue-specific regulation and great diversity of protein sequences suggests both genes and protein products have undergone subfunctionalization. However, recent investigations into the functions of ADF proteins to suggest that there is extensive neofunctionalization between family members. For instance, subclass I ADF variants within Arabidopsis have been shown to have specialized roles in plant immunity (Tian et al., 2004; Clement et al., 2009). One of the strongest evidence for neofunctionalization comes from investigations on the two Arabidopsis subclass III ADF variants, ADF5 and ADF9. Plants that were defective for ADF5 had significantly altered rosette leaf morphology and also showed reduced levels of leaf death and electrolyte leakage under freezing conditions. In contrast, Arabidopsis plants mutant in ADF9 exhibited various aberrant phenotypes, with the most notable phenotype being early flowering during long-day light cycles but not during short day light cycles (Burgos-

Rivera, 2008). Overexpressing its closest paralog, ADF5, did not suppress this early flowering phenotype, suggesting that ADF9's role in flowering time is a highly evolved (unpublished data). Likewise, overexpressing ADF9 in *adf5-1* did not suppress the altered rosette leaf morphology (unpublished data). This indicates that there must be key differences between these variants, whether it is solely structural and biochemical, or whether it is due to different binding partners cannot be concluded at this point.

Dissecting the structural and biochemical components that played major roles in protein evolution is the central core of the study of molecular evolution (Harms and Thornton, 2010). In this study, we take a close look at the molecular evolution of a family of proteins that have been classified as a "functional node in cell biology" as they regulate so many aspects of cell and organismal responses (Bernstein and Bamberg, 2010). Understanding the evolution of these important modulators of actin dynamics, which are so crucial for mediating stress responses as well as proper plant development, will help further our understanding on how plants are capable of adapting to a rapidly changing environment. Our goal is to investigate how these ADF variants differ structurally and mechanistically in order to determine how these proteins have functionally diverged over their evolutionary course. To do this, we will be employing the technique of ancestral state reconstruction in order to identify the mutational changes that have resulted in functional divergence or new interactions between binding partners. Ancestral state reconstruction has been used in a number of studies to answer a variety of different questions in molecular evolution. For instance, this method was used to resurrect the rhodopsin gene of the extinct archosaur to better understand the evolution of vision in vertebrates (Chang, 2003). Additionally, Harms and Thornton (2010) used this

technique to investigate the functional divergence between the human glucocorticoid and mineralocorticoid receptors.

For the ADF gene family, the simplest hypothesis might be that the Arabidopsis gene family members have simply partitioned their functions between each other through subfunctionalization. Therefore, we tested whether the ancestral function has been retained across the diverse Arabidopsis ADF protein variants through a series of genetic suppression studies in Arabidopsis ADF mutants using the extant *P. patens* and *S. moellendorffii* ADF protein sequences. *Physcomitrella patens* contains only a single ADF and is unique in that it completely lacks the presence of introns (Augustine et al., 2008). This is in stark contrast to other plant lineages, where most ADFs contain two introns that have been implicated as being important to the tissue specific expression pattern (Jeong et al., 2007). This attribute combined with its basal placement within the phylogeny (i.e., it is not contained in any of the subclasses) gives a strong indication that this ADF may be the closest representative of the ancestral function of the ADF variants within plant lineages. The *S. moellendorffii* genome possesses two ADF variants: one which groups phylogenetically with the clade containing subclass I and subclass II while the other ADF variant groups with the clade containing subclass III and subclass IV. This is especially interesting as it indicates that this gene family exhibited a major split prior to the further diversification seen in the angiosperm lineage. Here our goal was to determine the degree of functional conservation of these two lycopod ADFs across the four ancient subclasses. To test this hypothesis, we performed a series of suppression tests with Arabidopsis ADF mutants from three of the four ancient subclasses.

However, there are a number of reasons that the *P. patens* and *S. moellendorffii* proteins may not suppress the phenotypes of Arabidopsis ADF mutants in our study. It could be due to the high degree of divergence between Arabidopsis family members or it could be simply due to the accumulation of lineage specific mutations hindering the suppression studies. In anticipation of this possibility, we reconstructed the ancestral proteins along three key points of the plant ADF phylogeny and used these reconstructed proteins in the same series of suppression studies described above. First, we reconstructed the ancestral protein sequence between the two closely related subclass III ADF variants, ADF5 and ADF9, in order to identify the specific mutations that have facilitated their highly specialized role within Arabidopsis. Then, we reconstructed the ancestral proteins at the nodes leading to the four subclasses to address the more ancient divergences seen prior to the diversification of the four ADF subclasses (Figure 4.1). The results from our suppression studies highlight the high degree of divergence between ADF family members and provide strong evidence that family members have been rapidly evolving post-duplication.

Materials and Methods

Ancestral state reconstruction

The ancestral protein sequence for three particular nodes in the ADF phylogeny (Figure 4.1) was reconstructed using the program PAML (Yang, 2007). As with performing a phylogenetic analysis, there are three different sequence data formats that are available for estimating the ancestral states along a phylogeny: nucleotide sequence data, codon-based data, and amino acid sequence data. For each of these three sequence formats,

there are multiple substitution models that can be used to analyze input data. To determine the best-fit model for sequence prediction, a series of nested models were performed and tested using a log-likelihood ratio test (LRT) for each of the three methods. Four nucleotide substitution models were tested both with and without a gamma distribution parameter: JC69 (Jukes and Cantor, 1969), F81 (Felsenstein, 1981), HKY85 (Hasegawa et al., 1985), and GTR (Yang, 1994; Zharkikh, 1994). For the codon-based approach, two different codon substitution models were tested: the one-ratio model (M0) and the discrete model (M3) of codon evolution (Yang, 2007). For each of the two models, four different equilibrium codon frequency parameters were also tested. For the amino acid based approach, three well-known models of protein evolution were tested and compared: Dayoff (Dayoff et al., 1978), Jones (Jones et al., 1992), and the Wag (Whelan and Goldman, 2001) model of protein evolution.

Again, for each of the three approaches (nucleotide, codon, and amino acid) the best-fit model was determined through a series of nested log-likelihood ratio tests under a χ^2 distribution. Once the best-fit model was determined, the sequences from the nodes of interest from each method were compared and a consensus sequence was built. Any discrepancies between sequences from the three approaches were resolved by choosing the most probable amino acid using either a majority rule method (i.e., two out of the three methods agree on a sequence) or by choosing the amino acid with the highest posterior probability. Two consensus sequences were developed for the two most ancestral nodes while only one sequence was tested for the ancestral node between ADF5 and ADF9.

Ancestral protein synthesis and design

A *P. patens* ADF cDNA clone was generously provided by the laboratory of Dr. Magdalena Bezanilla (University of Massachusetts, Biology Department) and was cloned into a pCAMBIA binary vector under the ACT2 promoter-terminator as described by Kandasamy et al. (2002). Briefly, an NcoI and BamHI restriction site was introduced to the beginning and end of the cDNA sequence, respectively, through modified primers. These restriction sites were used to clone the sequence into a BlueScript vector and then the construct was shuttled into the ACT2 promoter-terminator pCAMBIA binary vector already established in the laboratory. The cDNA for the two *S. moellendorffii* ADF sequences and the five ancestral proteins were synthesized by GenScript and cloned into the ACT2 promoter-terminator pCAMBIA binary vector as describe above.

Transgenic lines and mutant analysis

Each construct was transformed into one mutant from each subclass, except for subclass IV. In Arabidopsis, there is only one subclass IV protein, ADF6, and plants defective in ADF6 have no observable or testable phenotype (Roy-Zokan et al., in prep). Additionally, due to the high degree of functional divergence in the two subclass III members, ADF5 and ADF9, both mutants were used for analysis. As mentioned in the introduction, *adf9-1* has an early flowering phenotype while *adf5-1* displays altered rosette leaf morphology (Burgos-Rivera et al., 2009; King et al., in prep). For subclass I, we chose to use an *adf4-1* mutant line that is already established in our laboratory. Plants defective in ADF4 have decreased resistance to *Pseudomonas* infection (Tian et al., 2004) and also have increased root lengths (Henty et al., 2011). Since our laboratory is

not equipped to analyze immunity, we screened root length as a test of suppression. Our laboratory also has an RNAi line that targets the silencing of the subclass II ADF protein variant, ADF11, which was also used in our suppression analyses. Our preliminary analyses show that knocking down ADF11 results in plants with stunted root hairs, a phenotype that can be rapidly screened (unpublished data). Each construct was transformed into the subclass-specific mutants using the *Agrobacterium* floral dip method, as described by Clough et al. (2005). Positive transformants were selected on ½ MS growth media supplemented with 50 mg/L hygromycin and 300 mg/L tementin. The suppression of mutant phenotypes was screened on either T3 or T4 plant lines.

Plant Growth Conditions

All mutant strains used for this study were of the Columbia ecotype genetic background. For plate-based assays, seeds were surface sterilized and plated onto ½ MS growth media consisting of 1% sucrose and 0.8% phytoagar (Caisson). Seeds were stratified for 48 hours at 4 ^{o}C before transferring to an incubator with 16 hours light 8 hour dark light cycle at 22 ^{o}C. For soil-based assays, seeds were sown directly onto 70% organic soil (3B) and again stratified for 48 hours at 4 ^{o}C. After stratification, flats were transferred to growth rooms where plants were grown at 22 ^{o}C under a 16 hour/8 hour light/dark cycle.

Results

Ancestral state reconstruction

The ancestral protein sequences at three key nodes in the plant ADF phylogeny (Figure 4.1) were estimated by using nucleotide, codon, and protein based models of evolution. Overall, the three different methods of modeling substitution patterns (nucleotide-based versus codon-based versus amino acid-based models of evolution) yielded fairly congruent results (Figure 4.2A-C). The ancestral protein sequence at the node leading to the divergence between subclass I and subclass II, hereafter referred to as IvII ADF, differed at only six sites between the three sequences generated (Figure4. 2A). For the first consensus sequence, most of these sites were resolved by choosing the amino acid that two out of the three methods agreed upon. The two exceptions were amino acid position 51 and amino acid position 149. For amino acid position 51, we chose to go with valine instead of glutamine even though two of the three methods predicted glutamine for this position. This decision was based on the fact that no subclass I sequences had a glutamine in this position; they all had a valine in this position so therefore it was felt that based on the extant protein sequences we should use valine in this position over glutamine. The same reasoning was used for choosing serine over glycine for amino acid position 149. Most of the extant sequences in subclass I and subclass II had a serine in position 149, so therefore we chose serine over glycine. For the second IvII ADF consensus sequence, only positions 21 and 90 were changed. For these positions, the second predicted amino acid was chosen over the most predicted amino acid (Figure 4.2A).

There was a greater number of variable positions in the ancestral sequence at the node leading to subclass III and subclass IV, referred to as IIIvIV ADF (Figure 4.2B). In total, the models differed in nine amino acid positions. Incongruencies were resolved by choosing the amino acid that two out of the three models predicted with the exception of two sites: 90 and 107. For amino acid position 90, two of the models predicted aspartic acid for this position while the amino acid based model predicted glutamic acid. Most extant proteins had either valine, serine, or glycine in this position but the *S. moellendorffii* ADF that is basal to these two subclasses contains glutamic acid in this position. Going with our hypothesis that the *S. moellendorffii* ADFs represent ancestral states, we decided to use glutamic acid for this position over aspartic acid. For amino acid position 107, all models predicted different amino acids for this position. Looking across extant proteins, threonine was the most common amino acid in this position, so we chose threonine over the other predicted amino acids. For the second consensus sequence, a valine was chosen for this position (Figure 4.2B). The only other site that was changed in the second consensus sequence was position 21, where a serine was used in place of threonine. The predicted sequence for the ancestral state of ADF5 and ADF9 was completely congruent between the three methods (Figure 4.2C). Therefore, only this one sequence was synthesized and tested.

Genetic suppression analyses

To test our hypothesis that the *P. patens* and *S. moellendorffii* extant ADF protein sequences represent the ancestral state of the four subclasses of ADF proteins, a series of suppression studies were performed in Arabidopsis mutants. For the subclass I ADF

variant tested, *adf4-1*, only the *S. moellendorffii* protein variant that groups with subclass I and subclass II suppressed the long root phenotype (Figure 4.3). Interestingly, the same degree of suppression was not seen in *adf4-1* with the *P. patens* ADF protein (Figure 4.3). Neither the *P. patens* ADF protein nor either of the two *S. moellendorffii* protein variants suppressed the stunted root hairs of the ADF11Ri mutant (Figure 4.4), the altered rosette leaf morphology of *adf5-1* (Figure 4.5), or the early flowering phenotype of *adf9-1* (Figure 4.6).

Similar results were seen between the three ancestral protein sequences synthesized and tested. The first ancestral sequence synthesized for the node leading to the divergence between subclass I and subclass II failed to suppress the root phenotype of *adf4-1* but the second consensus sequence did rescue the root length phenotype (Figure 4.3). However, neither consensus ancestral sequence was able to suppress the root hair phenotype of the ADF11 RNAi line (Figure 4.4). Likewise, the two sequences synthesized for the ancestral state between subclass III and IV did not suppress the rosette leaf phenotype of *adf5-1* (Figure 4.5) or the early flowering phenotype of *adf9-1* (Figure 4.6). To further examine the divergence between ADF5 and ADF9, the ancestral state between these two closely related subclass III paralogs was synthesized and tested within both mutants. This hypothetical ancestral protein was unable to suppress either of the developmental phenotypes of *adf5-1* or *adf9-1* (Figure 4.5 and Figure 4.6).

Discussion

The plant ADF gene family is comprised of a group of diverse and versatile proteins that provide an excellent model for studying functional diversification after gene

duplication. Again, the simplest hypothesis is that these protein variants have subfuctionalized, that is, they partitioned ancestral functions between closely related paralogs. This is supported by the phylogenetic relationships between extant ADF protein sequences. If the subfunctionalization hypothesis is correct, then we would expect that *P. patens* ADF would be able to suppress the phenotypes from mutants in all three of the subclasses examined. Likewise, we would expect *S. moellendorffii* 146459 to suppress the phenotypes of mutants in subclass I and II and *S. moellendorffii* 230142 to suppress the phenotypes from mutants in subclasses III and IV. What we found in this study is that the Arabidopsis ADF variants have diverged significantly from these ancient moss ADFs. The one exception was that the Sm 146459 variant was able to suppress the root length phenotype of *adf4-1*. It would be interesting to see if this *Selaginella* protein variant is also capable of suppressing the specialized role of ADF4 in *Pseudomonas* immunity (Tian et al., 2004). Perhaps the root phenotype represents an ancient function conserved across ADF variants whereas pathogen defense may represent a derived function; therefore, it would be interesting to see if the lycopod ADF variant was capable of functioning in this highly specialized role.

Since the *P. patens* ADF was not able to suppress the root phenotype of *adf4-1* yet a *Selaginella* ADF was able to restore roots to wild-type lengths, this supported the idea that lineage specific mutations may be hampering our analyses. When we compare between the Arabidopsis ADF variants and the *P. patens* ADF, we are looking across an estimated 600 million years of evolution. Likewise, comparing between the *Selaginella* ADF proteins with the Arabidopsis ADF variants has us assessing changes that have occurred across an estimated 450 million years of evolution. These mutations may have

little to do with actual functional differences and may reflect the large degree of evolutionary divergence between these species (Harms and Thornton, 2010). To address this possible complication, we reconstructed the hypothetical ancestral proteins at three key nodes of the ADF gene phylogeny. Since our main interest was to investigate the mutational changes that facilitated the deep divergences between the four ancient ADF subclasses, we reconstructed the ancestral states at the node leading to the divergence between subclass I and II and also at the node between subclasses III and IV (Figure 4.1). Surprisingly, despite the high divergence times between these ancestral nodes and the extant Arabidopsis ADF variants, the three different substitution methods examined during the reconstruction yielded highly congruent results. Even with the strong support of predicted amino acid sequence between methods, these ancestral proteins failed to suppress the phenotypes of the Arabidopsis mutants examined with the exception of *adf4-1*. The second consensus sequence of IvII suppressed the root length phenotype of *adf4-1*. This sequence only differed from the first consensus sequence by two amino acid positions. This suggests that very minor changes in the protein sequence can have profound impact in the function of the ADF proteins.

These results also give a strong indication that the Arabidopsis ADF variants have rapidly diverged post duplication. This rapid divergence is especially evident when examining the divergence between ADF5 and ADF9. These two Arabidopsis ADF variants are their own closest paralogs within subclass III; i.e., there are no ADF protein variants from other species that is more closely related to ADF5 than ADF9 and vice versa. It seems reasonable to hypothesize that one of these two protein variants retained the ancestral function of the two while the other variant diverged quite rapidly. We had

expected, at the minimum, that the ancestral protein between ADF5 and ADF9 would at least suppress the phenotype of one of the paralogs. All three models compared yielded the same amino acid sequence for this particular protein and each site had a posterior probability of 0.98 or above with the exception of one site. Due to this, it was believed that the predicted protein sequence was the best approximation to the true sequence. Therefore, it is quite surprising that this ancestral protein was incapable of even partially rescuing either the altered rosette leaf morphology of *adf5-1* or the early flowering phenotype of *adf9-1*. This is highly suggestive that both proteins have diverged significantly, in a relatively short amount of evolutionary time, from the original ancestral function of subclass III ADF proteins.

Our null hypothesis at the onset of this study was that Arabidopsis ADF protein variants have evolved through subfunctionalization. Our results provide strong evidence that this null hypothesis can be rejected. The Arabidopsis ADF protein variants have diverged significantly in function compared to the ancestral states at the point of subclass divergences. We even find that these proteins have diverged significantly after recent duplication events (e.g., ADF5 and ADF9). However, a series of fundamental tests need to be performed before this can be offered as a final conclusion. First, it is imperative that the function of these hypothetical proteins be tested. We know from our qRT-PCR analyses that the proteins are indeed expressed in our transgenic lines; however, we cannot determine whether these proteins are actually functional. To do this, purified proteins would be tested in an in vitro system described by Chaudry et al., 2007. Briefly, the assay quantifies the rate of nucleotide exchange between ATP-G-actin and ADP-G-actin by measuring the increase in fluorescence with the incorporation of ε-ATP

(Invitrogen) (Chaudry et al., 2007). The second point that needs to be addressed is the number of phylogenetic nodes that were tested in this study. With the exception of ADF5 and ADF9, only the two very ancient nodes prior to subclass divergences were tested for functional conservation across the different subclasses. It would be interesting to see if the ancestral protein sequence of nodes closer to ADF4 and ADF11 could possibly suppress the mutant phenotypes. The high degree of divergence between ADF5 and ADF9 may just be specific to these two subclass III paralogs. We would need to test whether more closely related ancestral proteins to ADF4 and ADF11 could suppress their mutant phenotypes before any solid conclusions about their functional divergence can be asserted.

References Cited

Augustine, R.C., Vidali, L., Kleinman, K.P., and Bezanilla, M. 2008. Actin depolymerizing factor is essential for viability in plants, and its phosphoregulation is important for tip growth. The Plant Journal. 54, 863-875.

Bernstein, B.W., and Bamburg, J.R. (2010) ADF/Cofilin: a functional node in cell biology. Trends in Cell Biology. 20(4): 187-195.

Bettinger, B.T., Gilbert, D.M., and Amberg, D.C. (2004) Actin up in the nucleus. Nature Reviews Molecular Cell Biology. 5: 410-415.

Burgos-Rivera, B., Ruzicka, D.R., Deal, R.B., McKinney, E.C., King-Ried, L., and Meagher, R.B. (2008) Actin depolymerizing factor9 controls development and gene expression in *Arabidopsis*. Plant Mol Biol. 68(6):619-632.

Chan, C., Beltzner, C.C., and Pollard, T.D. (2009) Cofilin dissociates Arp2/3 complex and branches from actin filaments. Current Biology. 19(7): 537-545.

Chang, B.S.W. (2003) Ancestral gene reconstruction and synthesis of ancient rhodopsin in the laboratory. Integrative and Comparative Biology. 43:500-507.

Chua, B.T., Volbracht, C., Tan, K.O., Li, R., Yu, V.C., and Li, P. (2003) Mitochondrial translocation of cofilin is an early step in apoptosis induction. Nat. Cell Biol. 5:1083–1089.

Clement, M., Ketelaar, T., Rodiuc, N., Banora, M.Y., Smertenko, A., Engler, G., Abad, P., Hussey, P., and Engler, J.D. (2009) Actin-Depolymerizing Factor2-Mediated Actin Dynamics Are Essential for Root-Knot Nematode Infection of *Arabidopsis*. Plant Cell. 21(9):2963-2979.

Clough, S J, and Bent, A F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. The Plant Journal. 16, 735-743.

Dayhoff, M. O., R. M. Schwartz, and B. C. Orcutt. (1978) A model of evolutionary change in proteins. Pp. 345- 352. Atlas of protein sequence and structure, Vol 5, Suppl. 3. National Biomedical Research Foundation, Washington D. C.

Felsenstein, J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution 17:368-376.

Feng, Y., Liu, Q., and Xue, Q. (2006) Comparative study of rice and Arabidopsis Actin-depolymerizing factors gene families. Journal of Plant Physiology 163:69-79.

Han, L., Stope, M.B., Lopez de Jesus, M., Oude Weernink, P.A., Urban, M., Wieland, T., Rosskopf, D., Mizuno, K., Jakobs, K.H., and Schmidt, M. (2007) Direct stimulation of receptor-controlled phospholipase D1 by phospho-cofilin. EMBO J. 26:4189–4202.

Harms, M.J., and Thornton, J.W. (2010) Analyzing protein structure and function using ancestral gene reconstruction. Current Opinions in Structural Biology. 20, 360-366.

Hasegawa, M., Kishino, H., and Yano, T. (1985) Dating the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution. 22:160-174.

Henty, J.L, Bledsoe, S.W., Khurana, P., Meagher, R.B., Day, B., Blanchoin, L. and Staiger, C.J. (2011) Arabidopsis Actin Depolymerizing Factor4 Modulates the Stochastic Dynamic Behavior of Actin Filaments in the Cortical Array of Epidermal Cells. The Plant Cell. 23: 3711–3726.

Jeong, Y.M., Mun, J.H., Kim, H., Lee, S.Y., and Kim, S.G. 2007. An upstream region in the first intron of petunia actin-depolymerizing factor 1 affects tissue-specific expression in transgenic Arabidopsis (*Arabidopsis thaliana*). Plant Journal 50:230-239.

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992) The rapid generation of mutation data matrices from protein sequences. CABIOS 8:275-282.

Jukes, T. H., and Cantor, C. R. (1969) Evolution of protein molecules. Pp. 21-123 *in* H. N. Munro, ed. Mammalian Protein Metabolism. Academic Press, New York.

Kandasamy, M.K., McKinney, E.C., and Meagher, R.B. (2002) Functional nonequivalency of actin isovariants in Arabidopsis. Mol. Biol. Cell. 13: 251-261.

Kandasamy, M.K., McKinney, E.C., Meagher, R.B. (2009) A Single Vegetative Actin Isovariant Overexpressed under the Control of Multiple Regulatory Sequences Is Sufficient for Normal Arabidopsis Development. Plant Cell. 21(3):701-718.

Lehman, N., Di Fulvio, M., McCray, N., Campos, I., Tabatabaian, F., and Gomez-Cambronero, J. (2006) Phagocyte cell migration is mediated by phospholipases PLD1 and PLD2. Blood. 108:3564–3572.

Ohta, Y., Nishida, E., Sakai, H., and Miyamoto, E. (1989) Dephosphorylation of cofilin accompanies heat shock-induced nuclear accumulation of cofilin. The Journal of Biological Chemistry. 264:16143-16148.

Okreglak, V., and Drubin, D.G. (2007) Cofilin recruitment and function during actin-mediated endocytosis dictated by actin nucleotide state. The Journal of Cell Biology. 178(7):1251-1264

Meagher, R.B., McKinney, E.C., and Vitale, A.V. 1999 The evolution of new structures—clues from plant cytoskeletal genes. Trends in Genetics. 15 (7), 278-284.

Ruzicka, D.R., Kandasamy, M.K., McKinney, E.C., Burgos-Rivera, B., and Meagher, R B. (2007) The ancient subclasses of Arabidopsis ACTIN DEPOLYMERIZING FACTOR genes exhibit novel and differential expression. The Plant Journal. 52: 460-472.

Tian, M.Y., Chaudhry, F., Ruzicka, D.R., Meagher, R.B., Staiger, C.J., and Day, B. (2009) *Arabidopsis* Actin-Depolymerizing Factor AtADF4 Mediates Defense Signal Transduction Triggered by the *Pseudomonas syringae* Effector AvrPphB. Plant Physiology. 150(2): 815-824.

Whelan, S., and N. Goldman. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum likelihood approach. Molecular Biology and Evolution. 18:691-699.

Yang, Z. (1994) Estimating the pattern of nucleotide substitution. Journal of Molecular Evolution. 39:105-111.

Yang, Z. (2007) PAML4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution. 24(8), 1586-1591.

Zharkikh, A. (1994) Estimation of evolutionary distances between nucleotide sequences. Journal of Molecular Evolution. 39:315-329.

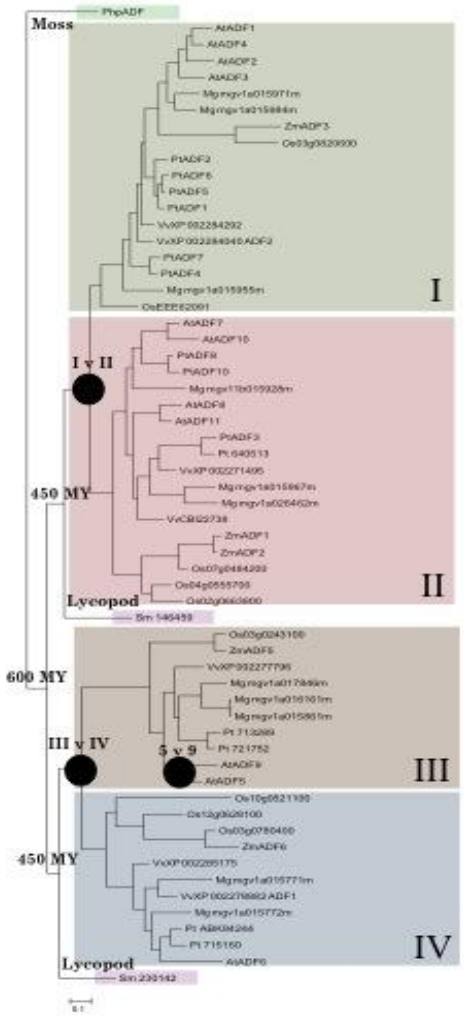


Figure 4.1: Bayesian inference phylogenetic analysis of the plant ADF gene family. *Physcomitrella patens* ADF is highlighted in green and *S. moellendorffii* variants are highlighted in purple. Nodes at which ancestral sequences were predicted and synthesized are indicated by black dots. Species designations are as follows: At, *Arabidopsis thaliana*; Pt, *Populus trichocarpa*; Mg, *Mimulus guttatus*; Vv, *Vitis vinifera*; Os, *Oryza sativa*; Zm, *Zea mays*; Sm, *Selaginella moellendorffii*; and Php, *Physcomitrella patens*. Dates for divergence times in millions of years (MY) interpolated from those in Yoon et al., 2004; Heckman et al., 2001; Moore et al., 2007; Zimmer et al., 2007; Meagher et al., 1989.

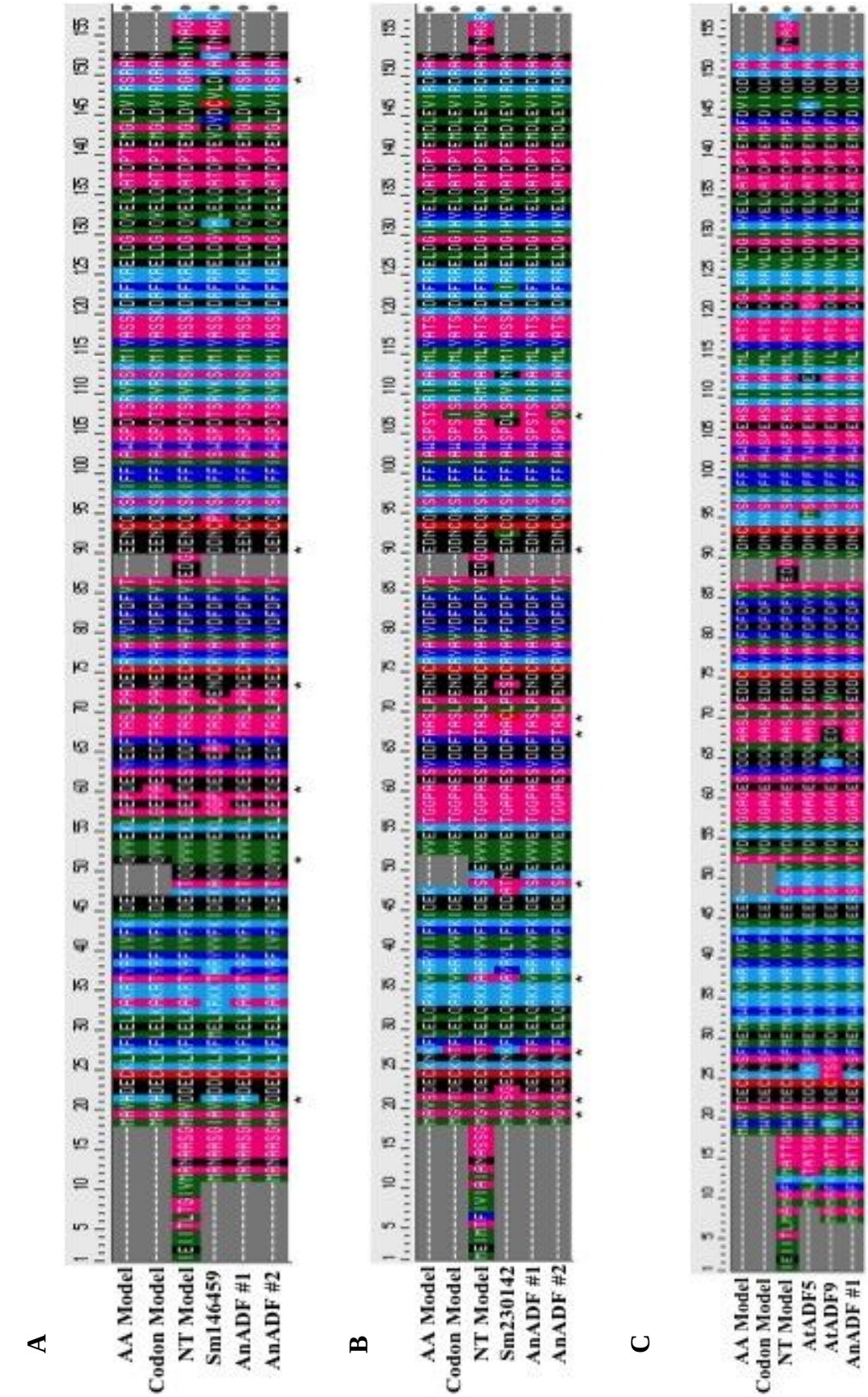
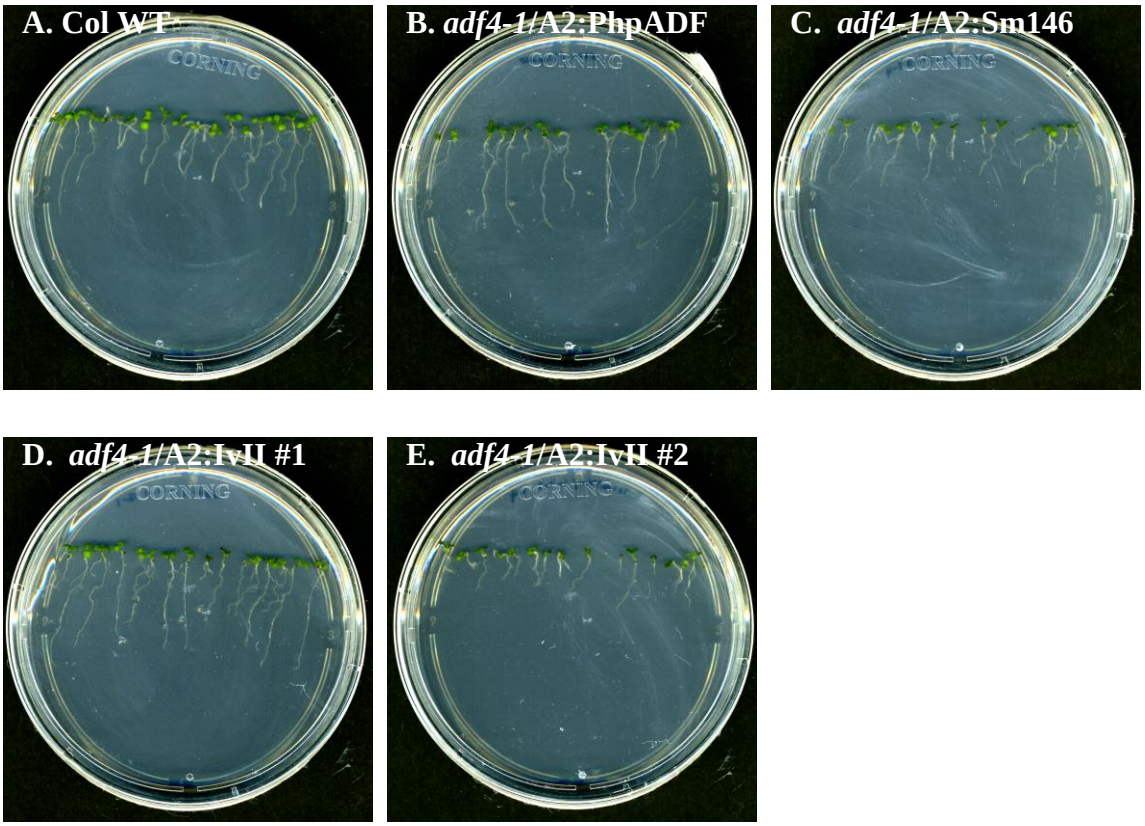

A
AA Model
Codon Model
NT Model
Sm146459
AnADF #1
AnADF #2
B
AA Model
Codon Model
NT Model
Sm230142
AnADF #1
AnADF #2
C
AA Model
Codon Model
NT Model
AtADF5
AtADF9
AnADF #1

Figure 4.2: Predicted sequences of ancestral proteins along the ADF phylogeny from ancestral state reconstruction. The predicted sequence from the three different methods of modeling substitution patterns are shown. Model designations are as follows: AA Model = Dayoff model of protein evolution; Codon Model = M3 model of codon evolution; and NT Model = the General Time Reversal (GTR) model with gamma distribution. Consensus sequences are represented as AnADF #1 and AnADF #2. Incongruencies between modeling methods are denoted with an asterisk. (A) Alignment of the predicted ancestral protein sequence at the phylogenetic node between subclass I and subclass II (IvII). Included in the alignment is the *S. moellendorffii* ADF variant that groups basal to subclass I and subclass II (Sm 146459). (B) Alignment of the predicted ancestral protein sequence at the phylogenetic node between subclass III and subclass IV (IIIvIV). Included in the alignment is the *S. moellendorffii* ADF variant that groups basal to subclass III and subclass IV (Sm 230142). (C) Alignment of the predicted ancestral protein sequence of the two subclass III ADF variants, ADF5 and ADF9. Amino acids colored by chemical similarity.



F.

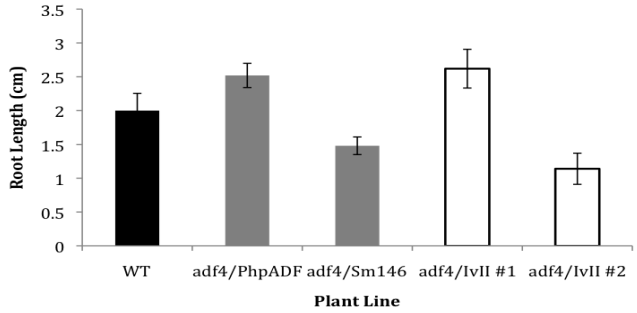


Figure 4.3: The role of ADF4 in root development represents an ancient function conserved across evolution. (A-E) Nine day old seedlings of Columbia wild-type and the following *adf4-1* transgenic lines: (B) P. patens ADF (PhpADF), (C) *S. moellendorffii* ADF 146459 (Sm146), (D) consensus ancestral protein sequence #1 (IvII #1), and (E) the second consensus sequence for the ancestral protein sequence (IvII #2). (F)Quantification of root length measurements in centimeters for all lines pictured averaged across five individuals with the longest roots.

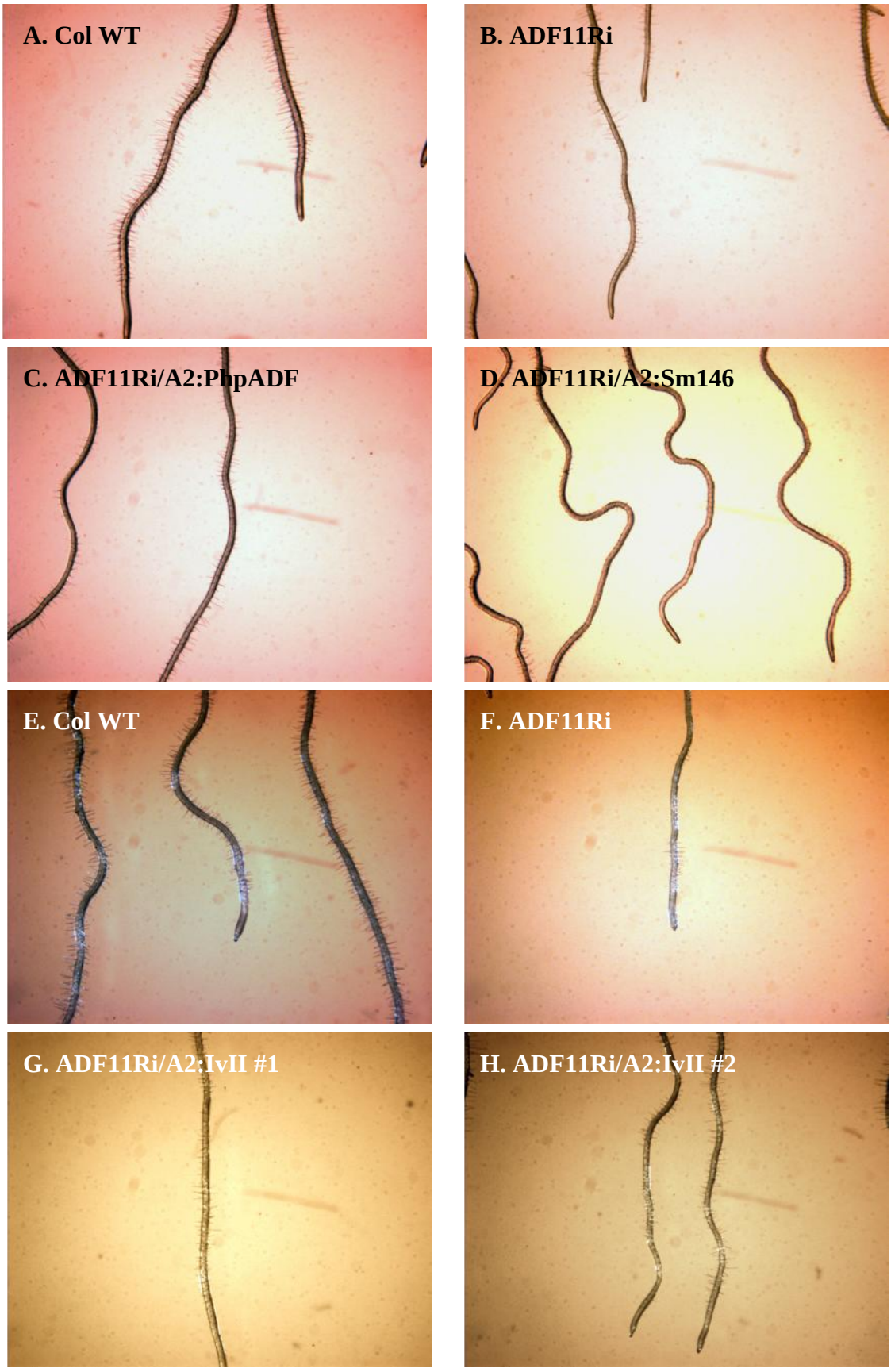
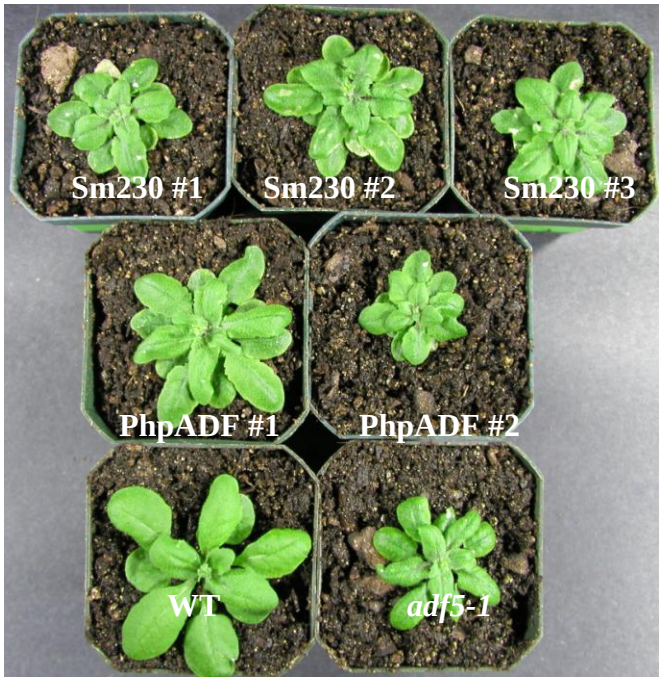

A. Col WT
B. ADF11Ri
C. ADF11Ri/A2:PhpADF
D. ADF11Ri/A2:Sm146
E. Col WT
F. ADF11Ri
G. ADF11Ri/A2:IvII #1
H. ADF11Ri/A2:IvII #2

Figure 4.4: Ancestral ADF proteins fail to suppress the root hair phenotype of the ADF11 RNAi line or produce dominant negative phenotypes. (A-H) Suppression analyses for root hair development in Columbia wild-type (Col WT) and ADF11Ri transgenic lines expressing the *P. patens* ADF (PhpADF), the *S. moellendorffii* 146459 ADF (Sm146), and the two ancestral proteins (IvII #1 and IvII #2). A-D. 22 day old plants. E-H. Nine day old seedlings.

A


Sm230 #1
Sm230 #2
Sm230 #3
PhpADF #1
PhpADF #2
WT
adf5-1

B


WT
adf5-1
IvII #1
5v9

C

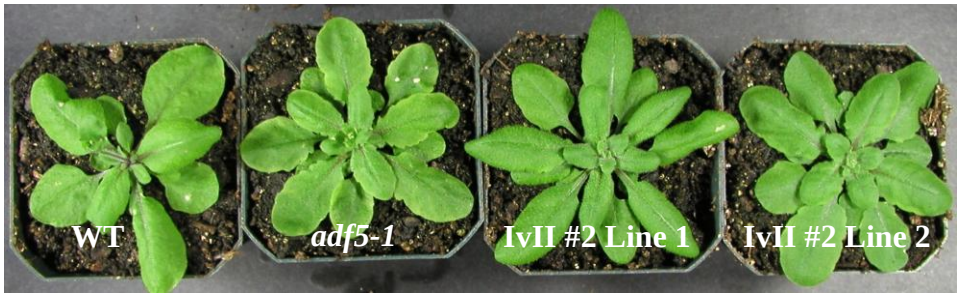

WT
adf5-1
IvII #2 Line 1
IvII #2 Line 2

Figure 4.5: Ancestral ADF proteins fail to suppress the increased number of rosette leaves in *adf5-1*. (A) Suppression analyses for rosette leaf morphology in Columbia wild-type (WT), *adf5-1*, and *adf5-1* transgenic lines expressing the *P. patens* ADF (PhpADF) and the *S. moellendorffii* 230142 ADF (Sm230). 29 day old plants pictured. (B) Suppression analyses for rosette leaf morphology in Columbia wild-type (WT), *adf5-1*, and *adf5-1* transgenic lines expressing the first consensus ancestral sequence (IIIvIV #1) and the ADF5 and ADF9 ancestral sequence (5v9). (C) Suppression analyses for rosette leaf morphology in Columbia wild-type (WT), *adf5-1*, and two *adf5-1* transgenic lines expressing the second consensus ancestral sequence (IIIvIV #2). (B-C) Plants pictured at day 31.

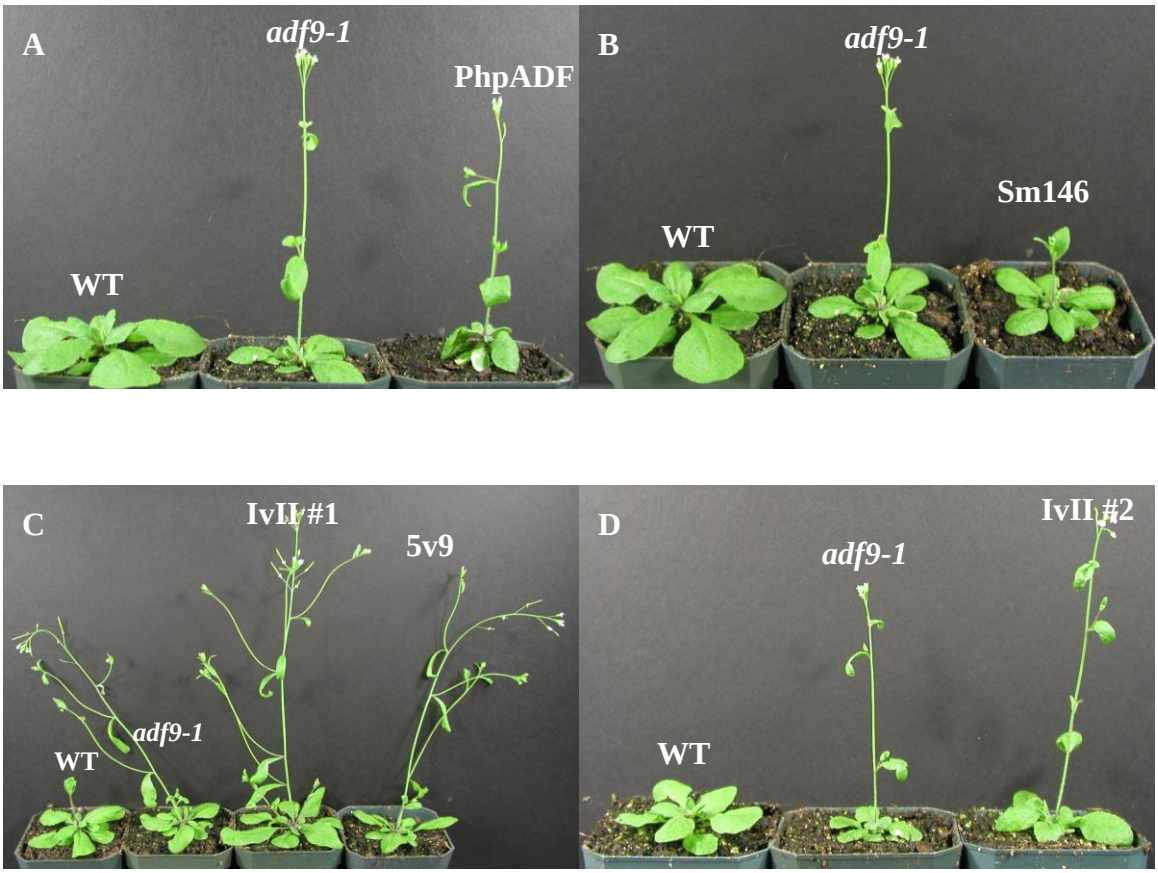


Figure 4.6: Ancestral ADF proteins fail to suppress the early flowering time phenotype of *adf9-1*. (A) Suppression analyses of flowering time in Columbia wild-type (WT), *adf9-1*, and *adf9-1* transgenic line expressing the *P. patens* ADF (PhpADF). (B) Suppression analyses of flowering time in Columbia wild-type (WT), *adf9-1*, and *adf9-1* transgenic lines expressing the *S. moellendorffii* 230142 ADF (Sm230). (C) Suppression analyses of flowering time in Columbia wild-type (WT), *adf9-1*, and *adf9-1* transgenic line expressing the first consensus ancestral sequence (IIIvIV #1) and the ADF5 and ADF9 ancestral sequence (5v9). (D) Suppression analyses of flowering time in Columbia wild-type (WT), *adf9-1*, and *adf9-1* transgenic line expressing the second consensus ancestral sequence (IIIvIV #2).

Table 4.1: LRT analyses of nucleotide and codon substitution models

	Model	ln(Likelihood)	Comparison	LRT Statistic	df	p-value
Nucleotide Substitution Models						
	JC69 + G	-13636.9497	JC69+G v HKY85+G	684.926	4	p < 0.0001
			JC69+G v GTR+G	755.37	8	p < 0.0001
	F81 + G	-13667.223	F81+G v HKY85+G	745.472	1	p < 0.0001
			F81+G v GTR+G	815.916	5	p < 0.0001
	HKY85	-13667.4493				
	HKY85 + G	-13294.4867	HKY85 v HKY85+G	745.924		p < 0.0001
			HKY85+G v GTR+G	70.444	4	p < 0.0001
	GTR	-13624.9744				
	GTR + G	-13259.2646	GTR v GTR+G	731.418		p < 0.0001
			HKY85+G v GTR+G	**70.444**	**4**	**p < 0.0001**
Codon Substitution Models	M0 CF0	-8756.73712				
	M0 CF1	-8809.64236	CF0 v CF1	0	3	p > 0.05
	M0 CF2	-8765.12325	CF1 v CF2	0	6	p > 0.05
	M0 CF3	-8697.51366	CF2 v CF3	135.219	5	p < 0.001
	M3 CF0	-8595.59947				
	M3 CF1	-8650.10271	CF0 v CF1	0	3	p > 0.05
	M3 CF2	-8601.32529	CF1 v CF2	0	6	p > 0.05
	M3 CF3	-8546.60873	CF2 v CF3	109.433	5	p < 0.001
	M0 CF3	-8697.51366				
	M3 CF3	-8546.60873	**M0 v M3**	**301.81**	**4**	**p < 0.001**

CF = Codon frequency model. CF0 (1/61) assumes the frequency of nucleotide usage at each codon position is equal (i.e., unbiased codon usage). CF1 uses the average nucleotide frequency calculated for the entire sequence while CF2 uses the average nucleotide frequencies calculated at each of the three codon positions. CF3 allows codon frequency parameter to vary freely without any assumptions (Yang, 2007).

CHAPTER 5

Conclusions and Perspectives

The actin cytoskeleton is an essential component of eukaryotic cells and has had a profound impact on the evolution of multicellular organisms (Meagher et al., 1999). Actin is involved in a number of fundamental processes such as organelle movement, cell division and polarity, and chromatin remodeling to name a few (Bettinger et al., 2004; Miralles and Visa, 2006). There are approximately 20 classes of actin-binding proteins that contribute to the dynamic nature of the actin cytoskeleton, and one such group of proteins are encoded by the Actin-Depolymerizing Factor/Cofilin (ADF/CFL) gene family. As with actin, ADF/CFL are also highly diverse in their cellular role. ADF/CFLs have been implicated in a variety of cellular processes such as lipid and membrane metabolism, mitochondrial dependent apoptosis, chemotaxis, and the cytonuclear trafficking of actin into the nucleus (Ohta et al., 1989; Chua et al., 2003; Bettinger et al., 2004; Lehman et al., 2006; Han et al., 2007; Bernstein and Bamburg, 2010). While recent studies have begun to detail how ADF/CFLs are able to perform their various roles within the cell, there is still much that remains to be learned. What drew my interest to the ADF/CFL gene family, is that although these proteins were relatively small compared to other actin-binding proteins, their impact on various cellular processes was far reaching (Bernstein and Bamburg, 2010). As just stated, these proteins are implicated as regulators of very intricate cellular processes, from cellular mobility to the regulation of

gene expression. Yet they are only ~150 amino acids in length. How could these protein variants be evolving so many different functions?

In both plant and animal literature, we see that the ADF/CFL variants are differentially regulated, even on the cellular level. A prime example of this differential regulation within a single cell comes from work done on the Arabidopsis reproductive specific subclass II ADF variants, ADF7 and ADF10 (Daher et al., 2011). These two protein variants are expressed at different stages of pollen grain development and have vastly different localization patterns. For example, ADF7 localizes to the nucleus and mainly associates with actin filaments in the elongating pollen tube only whereas ADF10 does not exhibit nuclear localization and is associated with filamentous actin throughout development of the male gametophyte (Daher et a., 2011). From studies like these, we know that these proteins have at least subfunctionalized; that is, they have partitioned their ancestral functions between each other. But what about neofunctionalization? How functionally distinct are gene family members? These were some of the basic questions I set out to answer with the three projects detailed in this dissertation. More specifically, my goals were to decipher the degree of divergence between family members, identify the mutational changes across the gene family's evolutionary history that facilitated this divergence, and to understand the evolutionary processes that have shaped the evolution of this actin-binding protein gene family.

As reported in previous studies, plant ADF variants partition into four phylogenetic subclasses (Chapter 2, Figure 2.2; Feng et al., 2006; Ruzicka et al., 2007) in which each subclass is differentially regulated across development and tissue type (Ruzicka et al., 2007). The same is seen in animal ADF/CFL protein variants. After

looking at previous phylogenetic studies of plant and animal ADF/CFL variants, I became interested in understanding the evolutionary processes that have driven the divergence between family members. More specifically, I wanted to test whether particular protein variants, or binding domains, have experienced accelerated rates of evolution. The selection analyses presented in Chapter 2 demonstrated that there was variation in selective pressure across both the plant and animal ADF/CFL gene families. Interestingly, there were differing patterns of codon evolution specific to subclasses of plant ADFs as well as to the different classes of animal ADF/CFL variants. More importantly, the sites that exhibited subclass- and class-specific patterns of accelerated evolution were located in exterior protein regions that are potentially new binding sites or may cause conformational changes to the ADF/CFL protein structure or alter its allosteric potential, any one of which could seriously impact function. These sites provide a good target for future molecular studies to investigate the functional differences between these protein variants.

One particular Arabidopsis ADF variant drew my attention when I initiated my studies of the Arabidopsis ADF gene family: the single subclass IV Arabidopsis ADF variant, ADF6. Characterizing the functional role of ADF6 was a crucial component to understanding the evolution of these protein variants. While great strides had been made in understanding the functional role of ADF variants within the other three subclasses, virtually nothing was known about ADF6. The subclass IV ADFs predate the divergence of monocots and dicots and hence is quite ancient. However, the selection studies in Chapter 2 did indicate that there was relaxed selection occurring in the dicot subclass IV ADF variants. Although ADF6 itself did not contain a signature of accelerated rate of

evolution, the phylogenetic lineages leading up to ADF6 did have this signature. This finding, coupled with its ubiquitous expression pattern, led me to hypothesize that the ADF6 protein variant must have a crucial role within Arabidopsis. The revelation of the spatial and temporal regulation of ADF6 expression pattern further supported this hypothesis. However, it came as a surprise to find that knocking down ADF6 had absolutely no aberrant effect on development and that there was absolutely no cost of fitness for carrying a defective ADF6 allele.

Generating double mutants with the two subclass III ADF variants, ADF5 and ADF9, provided the most exciting results from this study. By crossing *adf6-1* with *adf9-1*, we see that ADF6 plays an antagonistic role with ADF9 in regulating flowering time as the *adf9/6* mutant flowers at the same time as wild-type. More importantly, we see the expression levels of key regulatory genes in flowering time restored to wild-type levels in the *adf9/6* double mutant. It must be stressed that this role of ADF6 in flowering time is completely dependent on ADF9, as *adf6-1* flowers normally and shows no alteration in expression of any of the key flowering time genes. Given what we know from animal literature on cofilin's role in gene expression regulation, coupled with recent findings on how ADF9 regulates Flowering Locus C (FLC) expression, the best model that explains these results is one that relies on cytoskeletal dynamics. In my model (Chapter 3, Figure 3.10), I hypothesize that there is a repressor or transcription factor controlling a repressor of FLC that is bound to G-actin in the cytoplasm. The antagonistic activity of ADF9 and ADF6 on flowering time is most likely due to these two protein variants having different effects on actin filament turnover. As seen with ADF/Destrin and CFL1 within mammalian cells, ADF6 and ADF9 could have fundamental biochemical differences that

allow one protein to function more as a nucleating binding protein while the other is more efficient at severing actin filaments. In this model, I hypothesize that ADF6 is more efficient at nucleating filaments whereas ADF9 is more potent at filament severing. Under normal conditions, ADF9 and ADF6 work in conjunction with each other to regulate actin filament turnover; G-actin levels within the cytoplasm remains at normal levels, keeping the repressor or transcription factor bound to G-actin. Reducing ADF9 levels results in a reduction of filament severing and prevents the replenishment of the G-actin pool. When this is coupled with the nucleating activity of the still active ADF6, this ultimately results in a reduction of the G-actin pool and releases the bound repressor or transcription factor. The repressor can then translocate into the nucleus, repressing the expression of FLC and ultimately leading to early flowering. There are some fundamental experiments that need to be performed in order to test this hypothesis. First, we would need to test whether ADF6 truly acts more as a nucleating protein versus a filament severing protein. Second, the exact role of ADF9 in the regulation of FLC expression needs to be defined. Once we know exactly how ADF9 impacts gene expression, and the other proteins that are involved in this process, we will be better equipped to determine how ADF6 fits in to the flowering time pathway.

These results from ADF6 further fueled my curiosity about the divergences between subclasses. Here we have a single protein variant in a subclass that has been conserved for an estimated 250 million years, and yet, this protein is non-essential. This is in stark contrast with what we find in the other three subclasses. Subclass I ADF variants seem to have vital roles in immunity (Clement et al., 2002; Tian et al., 2004) while subclass II ADFs are important in tip growth (Daher et al., 2011; Ruzicka et al.,

unpublished data). When we move into subclass III, there are only two paralogs that are weakly expressed and yet have a significant impact in gene expression regulation (Burgos-Rivera et al., 2009; King et al., in prep). So this begs the question - what is driving this functional diversity. We've seen that protein variants within the plant ADF phylogeny have experienced different rates of evolution, but how can we check for mutational changes that played a significant role in divergence. For this I employed the technique of ancestral state reconstruction and synthesized the hypothetical ancestor of three key nodes in the plant ADF gene phylogeny. These hypothetical ancestral proteins, along with the ADF protein variants from *P. patens* and *S. moellendorffii*, were used in a series of suppression studies to test the null hypothesis that the Arabidopsis ADF variants have diverged solely by subfunctionalization. The results presented in Chapter 4, coupled with the selection analyses performed in Chapter 2, clearly show that the Arabidopsis ADF variants have been rapidly diverging from one another post-duplication. This is clearly demonstrated by the fact that the ancestral state between ADF5 and ADF9 failed to suppress the phenotype from either mutant. Although our suppression analyses only provided further support for the dynamic evolution of the ADF/CFL gene family instead of detailing how these proteins have diverged from one another, this method still has potential to reach our initial goal. To complete this project, it is imperative that the ancestral states along more nodes within the phylogeny be constructed and tested. If we could identify specific protein regions that seem to have changed repeatedly along lineages leading to an extant ADF protein, then we could target our efforts to these regions and attempt a series of point mutations. This approach can also be taken for the ancestral protein for ADF5 and ADF9. By comparing between the

protein sequences of ADF5, ADF9, and the ancestral protein sequences, we may be able to identify particular regions that are implicated in the divergence between these two closely related paralogs. Once we identify these regions, we will be able to make more conclusive assertions about the functional diversification of the ADF/CFL gene family.

References Cited

Bernstein, B.W., and Bamburg, J.R. (2010) ADF/Cofilin: a functional node in cell biology. Trends in Cell Biology. 20(4): 187-195.

Bettinger, B.T., Gilbert, D.M., and Amberg, D.C. (2004) Actin up in the nucleus. Nature Reviews Molecular Cell Biology. 5: 410-415.

Burgos-Rivera, B., Ruzicka, D.R., Deal, R.B., McKinney, E.C., King-Ried, L., and Meagher, R.B. (2008) Actin depolymerizing factor9 controls development and gene expression in *Arabidopsis*. Plant Mol Biol. 68(6):619-632.

Chua, B.T., Volbracht, C., Tan, K.O., Li, R., Yu, V.C., and Li, P. (2003) Mitochondrial translocation of cofilin is an early step in apoptosis induction. Nat. Cell Biol. 5:1083–1089.

Clement, M., Ketelaar, T., Rodiuc, N., Banora, M.Y., Smertenko, A., Engler, G., Abad, P., Hussey, P., and Engler, J.D. 2009. Actin-Depolymerizing Factor2-Mediated Actin Dynamics Are Essential for Root-Knot Nematode Infection of *Arabidopsis*. Plant Cell. 21(9), 2963-2979.

Daher, F.B., van Oostende, C., and Geitmann, A. (2011) Spatial and temporal expression of actin depolymerizing factors ADF7 and ADF10 during male gametophyte development in *Arabidopsis thaliana*. Plant Cell Physiol. 52(7): 1177–1192.

Feng, Y., Liu, Q., and Xue, Q. (2006) Comparative study of rice and Arabidopsis Actin-depolymerizing factors gene families. Journal of Plant Physiology 163:69-79

Lehman, N., Di Fulvio, M., McCray, N., Campos, I., Tabatabaian, F., and Gomez-Cambronero, J. (2006) Phagocyte cell migration is mediated by phospholipases PLD1 and PLD2. Blood. 108:3564–3572.

Meagher, R.B., McKinney, E.C., and Vitale, A.V. (1999b) The evolution of new structures: Clues from plant cytoskeletal genes. Trends Genet. 15: 278–284.

Miralles, F., and Visa, N. (2006) Actin in transcription and transcription regulation. Curr. Opin. Cell Biol. 18:261-266.

Ohta, Y., Nishida, E., Sakai, H., and Miyamoto, E. (1989) Dephosphorylation of cofilin accompanies heat shock-induced nuclear accumulation of cofilin. The Journal of Biological Chemistry. 264:16143-16148.

Ruzicka, D.R., Kandasamy, M.K., McKinney, E.C., Burgos-Rivera, B., and Meagher, R B. (2007) The ancient subclasses of Arabidopsis ACTIN DEPOLYMERIZING FACTOR genes exhibit novel and differential expression. The Plant Journal. 52: 460-472.

Tian, M.Y., Chaudhry, F., Ruzicka, D.R., Meagher, R.B., Staiger, C.J., and Day, B. (2009) *Arabidopsis* Actin-Depolymerizing Factor AtADF4 Mediates Defense Signal Transduction Triggered by the *Pseudomonas syringae* Effector AvrPphB. Plant Physiology. 150(2): 815-824.