Cytokinin response factors (CRFs) are important transcription factors that form a side branch of the cytokinin signaling pathway and have been linked to cytokinin-regulated processes during development. CRF proteins are defined as belonging to a specific transcription factor family by the presence of an AP2/ERF DNA-binding domain and can be distinguished within this family by a group-specific CRF domain involved in protein–protein interactions. Here we further delimit CRFs into five distinct clades (I–V) represented across all major angiosperm lineages. Protein sequences within each clade contain a clade-specific C-terminal region distinct from other CRFs, suggesting ancient evolutionary divergence and specialization within this gene family. Conserved patterns of transcriptional regulation support these clade divisions. Despite these important differences, CRFs appear to show preferential localization or targeting to vascular tissue in quantitative real-time PCR and reporter line analyses of Arabidopsis thaliana and Solanum lycopersicum (tomato). Phloem tissue expression within the vasculature often appears the strongest in CRF reporter lines, and an analysis of CRF promoter sequences revealed conservation and significant enrichment of phloem targeting cis-elements, suggesting a potential role for CRFs in this tissue. An examination of CRF loss-of-function mutants from cytokinin-regulated clades revealed alterations in higher order vein patterning. This supports both the general link of CRFs to vascular tissue and clade-specific differences between CRFs, since alterations in vascular patterning appear to be clade specific. Together these findings indicate that CRFs are potential regulators of developmental processes associated with vascular tissues.

**Keywords:** Arabidopsis • CRF • Cytokinin response factor • Genomic analysis • Phloem • Tomato.

**Abbreviations:** AHK, Arabidopsis hybrid histidine kinase receptor; AP2, apetala 2; ARR, Arabidopsis response regulator; CRF, cytokinin response factor; ERF, ethylene response factor; GUS, β-glucuronidase; MAPK, mitogen-activated protein kinase; qPCR, quantitative real-time PCR.

**Introduction**

The cytokinin response factors (CRFs) are a subset of the AP2/ERF family of transcription factor proteins found in all land plants. Within the ethylene response factor (ERF) subfamily, CRFs are defined by the presence of a group-specific domain, known as the CRF domain, several of which were initially identified in Arabidopsis microarray experiments as induced by cytokinin (Rashotte et al. 2003, Rashotte et al. 2006, Rashotte and Goertzen 2010). More recent studies have expanded the number of CRFs genes to 12 in Arabidopsis (AtCRFs) and identified similar numbers in other plant genomes (Rashotte and Goertzen 2010, Cutcliffe et al. 2011, Shi et al. 2012). Studies to date on CRF genes have primarily focused on their interactions with cytokinin and the cytokinin signaling pathway. A broader analysis of CRF gene family diversity and tissue specificity that might suggest potential functional roles beyond cytokinin linkages has not been performed.

One way to improve understanding of a group of highly related genes is through a genomic approach utilizing the rapidly expanding number of available genome sequences, expressed sequence tag (EST) libraries and other genomic data resources from across the plant kingdom. A similar approach used previously for CRFs helped identify and characterize the conserved CRF domain that now defines this group of proteins as distinct from other ERFs (Rashotte and Goertzen 2010). Patterns of evolutionary change within a gene family may also imply functional differentiation, and the clarification of evolutionary relationships should help in interpretation of the results of functional studies. One goal of this study was to identify further conserved sequence motifs and particular amino acid residues within the understudied C-terminal ends of CRFs potentially to understand the evolution and function of these key transcription factors.

In addition to functional and genomic studies, much can be learned about a group of related genes by studying key similarities and differences in expression patterns. Often genes encoding proteins of highly similar structure and function are differentially expressed either spatially or temporally.
Expression can also be regulated by myriad different signals including endogenous developmental signals such as hormones, as well as external stress responses. While it has been previously shown that some, but not all CRFs are transcriptionally regulated by cytokinin (Rashotte et al. 2003, Kiba et al. 2005, Rashotte et al. 2006, Shi et al. 2012), this study reveals that several CRFs from at least two species show a high degree of expression in vascular tissue, which appears to be a unifying characteristic of this group of genes. Future investigation of CRF function can now be directed at specific processes related to vasculature development and function.

Results

CRF proteins have distinct lineages within flowering plants

Although CRF genes have been identified across all land plant groups, a significant expansion and radiation of CRFs is evident in the angiosperms. We identified multiple, distinct groups of CRF genes in all major flowering plant lineages including Amborella, a representative of the earliest branching or most 'basal' flowering plants.

Phylogenetic analyses of >400 (403) expressed CRF proteins across land plant species were performed using only the conserved CRF and AP2 domain sequences due to a lack of homology in C-terminal regions. These analyses repeatedly indicate that CRF genes may have diversified into five distinct groups in flowering plants. Nearly every available angiosperm genome contained at least one CRF gene from each group, whereas gymnosperm and other non-flowering plant genomes contain CRFs of a single type, themselves distinct from any in angiosperms. Despite consistent identification of this pattern, the limited resolution afforded by phylogenetic analyses of short, highly conserved CRF and AP2 domains yields little support for distinct nodes corresponding to these divergences. However, independent analyses of individual groups led to the identification of novel highly similar motifs in the previously CRF-wide unalignable C-terminal regions (Fig. 1; Supplementary Figs. S1–S3). Importantly, these group-specific C-terminal regions that range in size from 70 to 200 amino acids cannot be aligned with other groups and therefore could not be included in phylogenetic analyses. However, these sequences map out as nearly perfect clade-specific characters on the phylogenetic tree, lending strong support to relationships otherwise based solely on conserved domains.

Phylogenetic analyses of the five angiosperm CRF groups show they are collectively a derived, monophyletic group with respect to all other plant CRF domain-containing proteins, which we have named CRF clades (I–V) (Fig. 1; Supplementary Fig. S1). Each CRF clade contains 1–2 genes per species, except for clade V which is represented by 2–4 genes in several taxa (e.g. four in Arabidopsis AtCRF9–AtCRF12 and three in tomato SlCRF9–SlCRF11). Clade V is sister to all other clades, as was previously found in a study where these genes were collectively referred to as 'B-clade' CRF-related sequences (Rashotte and Goertzen 2010). Clade V CRFs show little to no transcriptional induction by cytokinin (Winter et al. 2007, Shi et al. 2012).

Clades I, II and III appear to be closely related based on loose similarities in the C-terminal and extreme N-terminal regions of these genes, coupled with a low overall sequence divergence. The resolution of clade III as a monophyletic group sister to clades I and II is not always captured in phylogenetic analyses, but is strongly suggested by the presence of conserved C-terminal elements (Fig. 1; Supplementary Figs. S1–S3). Clade III includes members that are transcriptionally induced by cytokinin: AtCRF5, AtCRF6 and SlCRF5 (Table 1). Similarly, clade I members are induced by cytokinin: AtCRF2 and SlCRF2 (Table 1). Additionally, our quantitative real-time PCR (qPCR) analyses of the other Arabidopsis clade I member, AtCRF1, also revealed transcriptional induction to 1.93 ± 0.16-fold the levels in 14-day-old seedlings after cytokinin treatment of 1 μM BA for 2 h. Clade II members, in contrast, are not induced by cytokinin: AtCRF3, AtCRF4, SlCRF4 and SlCRF6 (Table 1; Rashotte et al. 2006, Shi et al. 2012).

Although the exact relationship between clade IV proteins and those of clades I–III is not clearly resolved by phylogenetic analysis, the overall lack of similarity in C-terminal sequence suggests that they could be sister to the remaining clades. At least one member of this group (SlCRF1) has been shown to be transcriptionally induced by cytokinin (Table 1). Interestingly, clade IV sequences are conspicuously absent from Brassicaceae genomes including Arabidopsis, but are well represented by other rosid families and even other Brassicales such as papaya. This appears to be a major exception to the rule of each clade being represented in every genome.

Although some clade-unique features can be identified within the CRF and AP2 domains, it is the C-terminal sequences that show the greatest degree of clade specificity. The beginning of each clade-specific region is marked by a putative mitogen-activated protein kinase (MAPK) phosphorylation site recognizable by the SP[T/S]SVL motif (Fig. 1). Even this motif found in all CRFs has some clade-based variation to it: the variable [T/S] residue of this motif in clade V is always a S, while it is primarily a T in the other clades (Fig. 1). Despite differences in C-terminal size and sequence composition, some clades do contain shared amino acid motifs, including a strongly conserved FQDI motif in both clades I and II (Fig. 1). Additionally, clades I–III all have a PX[D/E]XF[F] motif, as well as a SGY[D/E]S motif.

Clade IV proteins are defined by a short but broadly conserved C-terminus with an abundance of acidic residues of little similarity to other CRF clades. Clade V proteins have the least C-terminal homology to the other clades; however, a PX[D/E]XF[F] motif can be seen similar to that found in clades I, II and III, although positioned further from the MAPK site.

We further attempted to find support for these CRF clade divisions using a hierarchical cluster analysis of microarray transcriptome experiments, since differences in sequence among clades might also be linked to functional expression differences.
Fig. 1 Five distinct clades of CRF proteins. Alignment of C-terminal sequences of CRF protein clades. Partial sequence alignment from representative members of each CRF protein clade I–V is shown. Conserved motifs are indicated by boxes and are labeled with conserved residue patterns. A model of a CRF protein shows the orientation of conserved domains along with lines expanded to show the relative position of aligned sequences.
Two species that have numerous publicly available microarray experiments as well as distinct CRF clade sequences were analyzed: Arabidopsis and Populus. Analysis of the seven Arabidopsis clade-associated CRFs on ATH1 microarrays from all experiments (1,529 arrays) showed clustering of the genes similar to the sequence-based arrangement of clades (Supplementary Fig. S4). Identical results were found from analysis of more specific groups of experiments as well, including perturbation (1,473 arrays), and anatomical/developmental data sets (74 arrays). A similar analysis of 11 CRFs in Populus from all five clades over 384 microarray experiments also yielded clustering that mirrors the clade divisions (Supplementary Fig. S4). Together these findings further suggest that the conserved sequence diversity among CRF clades is probably correlated to functional differentiation.

CRFs are expressed in the vasculature

Stably transformed homozygous CRF promoter::GUS (β-glucuronidase) reporter plant lines for clades I–IV (AtCRF1, 2, 4, 5 and 6, and SlCRF1) were analyzed for spatial and temporal expression patterns in Arabidopsis and tomato. For all genes examined, in multiple independent lines, GUS expression was observed in vascular tissues across the plant, with the strongest expression in the phloem and procambium (Figs. 2, 3), different from other gene promoters expressed using the same vector (Siefers et al. 2009). Clade I genes (AtCRF1 and 2) showed strong expression in the vasculature of leaves, cotyledons, hypocotyls and roots (Figs. 2, 3). Expression in the first pair of leaves at very early stages of development appears to be strong and widespread; whereas these leaves when older, as well as subsequent leaves, have somewhat attenuated yet highly vascular-specific expression (Fig. 2). In roots, both clade I genes were expressed in the stele; AtCRF1 was restricted to the basal-most region of the primary root closest to the hypocotyl, whereas AtCRF2 could be found throughout the mature vasculature of primary and lateral roots and the tip in the area of the columella (Fig. 3). Clade I expression in the shoot apex was unique compared with genes of other clades in being present in young leaf primordia of the first pair of leaves (Fig. 2E).

The clade II gene AtCRF4 was found broadly expressed in the vasculature across the plant in primary and lateral roots, hypocotyls, cotyledons, rosette and cauline leaves, inflorescence stems, sepals and petals (Figs. 2, 3). Staining was also observed for AtCRF4 in the root tip in a pattern similar to AtCRF2 (Fig. 3), but no expression was observed in the shoot apex.

Clade III CRFs (AtCRF5 and 6) were also found to be highly expressed in vascular tissues. Both AtCRF5 and AtCRF6 promoters directed expression in veins of mature expanded leaves, while younger non-expanded leaves showed patchy or discontinuous expression (Fig. 2A). Leaves undergoing expansion began to show vascular-specific expression at the distal end first, possibly coinciding with maturation of higher order veins or the transition of the leaf to a source state. These genes also have tissue expression patterns distinct from each other. Similar to clade I gene expression, AtCRF6 expression was found in the shoot apex, although it is absent in leaf primordia (Fig. 2E). AtCRF6 expression in the roots was strongest in the stele, yet it could also be detected in root tip columella and in pre-cortical cells (Fig. 3). AtCRF5 was conspicuously absent from the root vascular column just above the zone of maturation, while found abundantly in root tips around the quiescent center expanding into the columella and developing vascular column (Fig. 3). Expression is strong in the majority of the mature stele, with the exception of the region noted above, and can be observed in the phloem (Fig. 3B). In lateral root tips, proAtCRF5::GUS expression could be detected shortly after the initial periclinal division in the pericycle (stage III or IV) (Fig. 3).

Because Arabidopsis lacks a clade IV gene, GUS expression analysis of this clade was carried out in tomato using SlCRF1. SlCRF1 was expressed in the vasculature of roots, stems, leaves and fruit (Fig. 4). In leaves, vascular expression was limited to the mid-vein along with the first- and second-order laterals. Examination of secondary growth in both the root and stem also indicated that secondary vascular tissues highly express SlCRF1 (Figs. 2F, 4). No expression was observed in root or shoot apices, or in any floral tissues. Interestingly, the pattern of the clade IV expression showed the least tissue specificity in our analysis. While expression was generally strongest in the vasculature, it was also found in other tissues including epidermal cells, the mesophyll of younger non-expanded leaves and the pericarp of unripe fruits (Fig. 4).

CRF promoters contain clade-specific vascular-related cis-elements

To investigate possible regulatory mechanisms responsible for CRF vascular expression, 1 kb of upstream CRF promoter sequences was examined for conserved motifs using MEME analysis (Bailey and Elkan 1994). A highly conserved TC-rich 29-mer (Motif 1) was found to be significantly enriched in promoters across all CRFs, and multiple copies occur in nearly every CRF in

| Table 1 Cytokinin induction of specific CRF clades |
|-------------|-----------------|-------------|
| CRF clade   | Cytokinin induction | References   |
| I           | +                | AtCRF1<sup>a</sup>, AtCRF2<sup>b</sup>, SlCRF<sup>c</sup> |
| II          | –                |             |
| III         | +                | AtCRF5<sup>d</sup>, AtCRF6<sup>e</sup>, SlCRF<sup>f</sup> |
| IV          | +                | SlCRF1<sup>g</sup> |
| V           | –                |             |

CRF clades are indicated as transcriptionally induced by cytokinin (+) or not (–) along with citations of specific CRF genes that have been shown to be induced.

<sup>a</sup> This study.
<sup>b</sup> Rashotte et al. (2006).
<sup>c</sup> Brenner et al. (2012).
<sup>d</sup> Shi et al. (2012).
<sup>e</sup> Rashotte and Goertzen (2010).
Arabidopsis and tomato (Fig. 5A, B). Several shorter TC-rich motifs were also identified in the promoters of individual CRF clades (from a wide range of species). These clade-specific motifs were even more conserved, in agreement with the evolutionary relationships inferred from the sequence alignments (Fig. 5A). Each of the motifs presented was the highest scoring and most abundant motif identified in the specific analysis indicated (i.e. of all CRFs or of a specific clade).

TC repeat motifs similar to those which were identified have been shown to direct expression to the phloem (Ruiz-Medrano et al. 2011). GUS expression was examined more closely to determine if CRF expression could be found in the phloem, PC, procambium.

Fig. 2 Promoter-driven GUS reporter gene expression for AtCRF1, 2, 4, 5 and 6 and SlCRF1 is seen in vascular tissues of various aerial organs. (A) Leaves and cotyledons of young Arabidopsis rosettes [10 days after germination (DAG)]. (B) Fifth or sixth leaves from 18 DAG. (C) Reproductive organs: AtCRF4 in sepal and petals, AtCRF5 in anthers, AtCRF6 in the transmitting tissue of an immature silique. (D) Close-up of minor veins in leaves expressing GUS under the regulation of AtCRF4 and AtCRF6 promoters (E) Shoot apices. (F) Free-hand (SlCRF1) and microtome sections from stems (SlCRF1), inflorescence stems (AtCRF4) and leaves (AtCRF1 and AtCRF6). Scale bars are equal in each grouping. X, xylem; P, phloem; PC, procambium.
possibly due to these conserved putative cis-elements. Cross-sections cut from roots, stems and leaves from proCRF:GUS lines revealed expression of CRFs in the phloem, as well as other vascular-associated cells (Figs. 2F, 4A). The phloem was (often along with others) the tissue with the strongest GUS expression in the samples examined, supporting the prediction based on promoter motif analysis. Additionally, using Nomarski optics, GUS expression in the veins of cleared intact leaves was found to be associated with the phloem, as well as with pro-cambium (Fig. 2D).

We hypothesized that the expanded expression domain of SlCRF1 beyond the vascular tissue could be due to other factors interacting with vascular targeting to modulate its expression. The fact that Arabidopsis lacks a clade IV CRF presented a unique way to examine this possibility, since it may also have lost or lack certain non-vascular-specific targeting enhancers that would bind to clade IV promoters. As such, a clade IV CRF promoter in Arabidopsis should only have the TC-rich repeat motif acted upon to regulate and target expression. This was tested by transforming Arabidopsis with the tomato SlCRF1-promoter::GUS construct. Examination of SlCRF1 promoter::GUS expression in Arabidopsis revealed a highly vascular-specific pattern, lacking expression in the additional non-vascular tissues seen in tomato (Fig. 4); this was taken as partial support of the hypothesis.

Examination of CRF vascular transcript expression for all clades in Arabidopsis and tomato leaves was conducted by comparing vascular-enriched or vascular-poor tissues. To accomplish this, RNA was isolated from vascular-enriched tissue (the mid-veins of leaves or leaflets), and relatively vascular-poor tissues (the remaining leaf laminae), and the relative transcript abundance between these two tissue samples collected from a pool of plants was compared using qPCR (Fig. 5C). AtCRF1–AtCRF6 showed between 55 and 110% greater levels of expression in vascular-enriched tissue. Additionally, tomato clade I–III members, SlCRF2, 4, 5 and 6, showed the same trend, and in some cases nearly 200% greater expression. Analysis of clade V members in both Arabidopsis and tomato, AtCRF9–AtCRF12 and SlCRF9–SlCRF11, also revealed a similar vascular-enriched expression pattern, although at slightly lower differentials (Fig. 5C). Statistical analysis of the 23 different CRF genes from Arabidopsis and tomato revealed that there was a significant increase in expression in vascular-enriched tissue for 19 genes (at P < 0.05) and an increase in all 23 (at a relaxed P < 0.10). This indicates that most CRFs are expressed at greater levels in vascular tissue, with the rest of the CRFs showing a strong trend towards that same finding. Overall, these findings are in agreement with GUS expression analyses and suggest that CRFs are transcriptionally targeted to the vasculature.
CRFs are required for normal leaf vasculature patterning

Visual examination of leaves from Arabidopsis crf mutants (crf1, crf2, crf3, crf5, crf6 and crf9) suggested that vein patterning in clades I and III was altered with respect to the wild type, particularly for secondary and higher order venation. While examination of primary and first-order lateral veins for mutants in all clades showed them to be similar to the wild type, mutants from clades I and III showed alterations in higher order veins and areoles. This was further investigated using LEAF GUI software specifically designed to examine these features (Price et al. 2011). Again analyses of whole leaves revealed only minor vein patterning changes in crf mutants due to similarities in primary and first-order vein measurements, as initially noted (Supplementary Table S1). However, examinations of higher order veins, from areas bounded by the mid-vein and two first-order lateral veins, revealed significant differences in crf mutants (Fig. 6).

Analyses of crf1 and crf2 (clade I) as well as crf5 and crf6 (clade III) leaves revealed distinct clade-specific patterns of higher order veins: the number of areoles and the density of free-ending veins (Supplementary Table S1). In contrast, no differences were seen in the clade II mutant (crf3) that was examined (Fig. 6; Supplementary Table S1). Because initial analyses revealed highly similar results for within-clade members, a single representative from each clade (crf2, crf3, crf5 and crf9) was used for extensive analysis of additional parameters: areole area and vein network density (Fig. 6). crf2 was found to have fewer areoles that were larger in area (Fig. 6). This is in direct contrast to crf5 with an increased number of smaller areoles. Neither difference was due to changes in overall leaf area. Both mutants also had increases in the density of vein networks and the density of free-ending veins (Fig. 6). It should be noted that AtCRF2 and AtCRF5 (along with their respective clade members) have nearly opposing expression patterns in young developing leaves before transitions into a vascular-specific pattern: ubiquitous for clade I vs. almost absent for clade III. Examination of the clade V mutant (crf9) did show differences in areole numbers in initial analyses; however, this was due to a larger overall leaf size rather than alterations in patterning (Supplementary Table S1).

Discussion

There are five distinct clades of CRF proteins in flowering plants

We have shown that the monophyletic group of CRF proteins underwent considerable diversification prior to the divergence
Fig. 5 (A) Over-represented motifs found in angiosperm CRF promoters by MEME analysis. Motif 1, consensus from all clades; motif 2, clade I consensus; motif 3, the TC-rich sequence from clade IV; and motif 4, the CA-rich motif from clade IV. Motifs shown are those with the lowest E (Expectation) values for each group, with the exception of motif 4 which is the second highest to motif 3. (B) Block plot showing the position of motif 1 in Arabidopsis and tomato gene promoters (1 kb upstream of ATG). (C) CRF transcripts in vascular-enriched tissue samples. Expression of CRF genes as indicated was measured in tissue from both the mid-veins and leaf blades or leaflets using qPCR. Values presented are the percentage increase in normalized expression from blade to mid-vein. Data represent the average ± SE of two biological replicates comprised of pooled RNA from at least six individual plants. Each run was performed with at least four technical replicates.

Fig. 6 Aberrant vascular patterns in crf mutant leaves. (A) Representative examples of wild-type and mutant leaves used for analysis from each clade with magnification of the regions examined showing alterations in the patterns formed by minor veins. (B) Table of mean values of data collected for genotypes representative of mutants for each clade ± SE. An asterisk indicates a significant change from the wild type (Student’s t-test P < 0.05).
of flowering plants, giving rise to five distinct CRF clades (I–V). Each of these clades is best defined by unique and highly conserved (within clade) C-terminal sequences, and is supported by clade-specific signature sequences in promoter, CRF and AP2 domains as revealed by sequence-based analyses. While CRFs are present in all land plants, it is only within angiosperms that this remarkable diversification can be observed. Even earlier branching angiosperm lineages such as Amborella or Nuphar have CRF genes from the more derived CRF clades. Almost every fully sequenced angiosperm genome contains at least one representative gene per clade. Interestingly, there are consistently low numbers of CRFs found in each clade: 1–2 in clades I–IV and 1–4 in clade V. From this pattern it may be inferred that a degree of selective pressure has maintained these clades and their low gene numbers, indicating that each group is of functional importance. Phylogenetic analyses using only the conserved CRF and AP2 domains define these clades roughly with moderate to low support, but their definition by conserved C-terminal sequence motifs is nearly perfect. Additionally, hierarchical clustering analyses of CRFs based on large sets of transcriptomic data from both Arabidopsis and Populus indicate that expression patterns reflect the clade-specific evolutionary relationships.

A rapidly growing pool of newly available sequences as well as the ability to identify additional CRFs based solely on C-terminal characters has greatly increased our ability to resolve CRFs via genomic analysis. The inclusion of many more CRF sequences with greater diversity has allowed us to find several previously unidentified well-conserved blocks in C-terminal regions (Fig. 1). Foremost is the identification of a conserved MAPK phosphorylation motif in all CRF proteins, which was only weakly detectable previously (Rashotte and Goertzen 2010). In addition, the existence of another novel conserved amino acid motifs became evident in specific clades. The FQDI of clades I and II, the SGY[D/E]S of I, II and III, and the PX[D/E]XF[F] element, recognizable in all except clade IV proteins, are all likely to be of significance in the function or potential regulation of these proteins. Aside from these short intraclade conserved sequences, most of the CRF C-terminal regions are uniquely CRF clade specific and indicate divergence from one another potentially during the emergence of angiosperms as a dominant plant group.

There are some additional CRF proteins, including AtCRF7, AtCRF8, SiCRF3, SiCRF7 and SiCRF8, that are not easily placed into any clearly defined clade, primarily because they lack C-terminal sequences downstream of the AP2 domain including the conserved MAPK motif. Phylogenetic analyses of these CRFs with just the AP2 and CRF domains does not clearly resolve their clade membership, although detailed examination of Arabidopsis and tomato genomes, including large-scale syntenic determination around these loci, suggests that each may have been the result of independent duplications of sequences from different clades (Supplementary Fig. S1; data not shown). Our inability to place such sequences clearly into a specific CRF clade emphasizes the significance of C-terminal sequences in these analyses. There are also a group of unplaced monocot sequences, probably belonging to either clade I, II or III (based on the presence of the FQDI motif), that in Neighbor–Joining and maximum likelihood trees group somewhat on their own near the base of clades I, II and III (Fig. 1). The inability to place this set of sequences is likely to be the result of the relatively poor availability of monocot sequences other than fast evolving grass sequences.

Sequence divergence among clades probably indicates a similar divergence or specification in gene function, which we found was supported by hierarchical clustering of transcriptome analysis experiments in Arabidopsis and Populus. An example of this specification is immediately evident from differences in transcript induction by cytokinin that occur in some clades (I, III and IV) but not others (II and V) (Table 1). There are likely to be additional characteristics pertaining to both expression and function that further delimit the proteins from each clade, some of which are probably related to the clade-specific C-termini. While there has been little study of CRF functionality, two genes have been examined that are now known to be clade IV CRFs: Pti6/SiCRF1 from tomato and Tsi1 from tobacco, both linked to pathogen resistance and immune responses (Zhou et al. 1997, Park et al., 2001, Gu et al. 2002). Our study would suggest that similar functions might exist for other clade IV members; however, it is unclear what this potentially means for Brassicaceae species, such as Arabidopsis, that lack clade IV CRFs. Interestingly, the final 15 amino acids at the C-terminal end of Tsi1 were shown to be necessary for activation of this transcription factor. Amino acids in this region are well conserved in clade IV proteins; however, a homologous motif is not obvious in other CRF clades. Preliminary data from expression and physiological analyses of other CRFs suggest there may also be clade-specific functionality related to different phloem-based processes; however, additional work is required to determine these connections further and the overall level of functional specification among CRFs.

**CRF genes show vascular expression**

Despite the observed clade-based sequence divergences of CRF genes, we have shown in Arabidopsis and tomato that CRFs show vascular expression patterns with the strongest expression in phloem tissues based on our in situ and in silico examinations. One possible mechanism for CRF phloem-specific expression is the abundance of phloem targeting cis-elements occurring in CRF promoters. Schneidererit et al. (2008) showed that a small region containing a TC-rich motif was necessary for phloem-targeted expression of the Arabidopsis Sucrose Transporter 2 (AtSUC2) gene. Additionally, it has recently been shown that similar TC sequence repeats along with a minimal promoter are sufficient to target expression of their cognate genes to phloem cells (Ruiz-Medrano et al. 2011). The expression of CRFs in phloem cells makes them good candidates for mediators of a number of developmental and environmental response processes known to be cytokinin regulated.
such as sink/source regulation, senescence and responses to biotic and abiotic stresses (Mok and Mok 2001, Werner and Schmulling 2009). Additionally, an AC-rich motif was found to be highly enriched in clade IV promoters (Fig. 5A). Highly similar motifs have been implicated in specifying xylem expression in Phaseolus vulgaris (Hatton et al., 1995) and Pinus taeda (Patzlaff et al. 2003). Together, the identification of these vascular-related promoter motifs supports the vascular-specific expression found in the reporter gene analysis and provides a potential mechanism for such specification.

Broad expression patterns of clade I CRF genes in young leaves that transition to highly vascular-specific expression as leaves expand is juxtaposed to that of clade III CRF genes that lack expression in young leaves, yet are expressed in a vascular-specific manner as they mature. Interestingly, these contrasting expression patterns parallel the findings from CRF mutant examinations indicating that these clades have opposing defects in leaf vein patterning, while other clades do not appear to follow any similar patterns. Together these findings suggest that while clade I and III CRFs may regulate vascular expression in developing leaves, they would appear to do so in differing if not opposing manners.

Root expression could be seen for nearly all genes examined, and at least one member from each clade was expressed in the root vasculature (Figs. 3, 4). Although AtCRF5 lacked vascular expression just above the zone of maturation in the root, it was the only CRF examined showing expression in pre-vascular meristematic cells. This may imply a role for AtCRF5 in actively dividing cells, also supported by its expression in emerging lateral roots. Other non-vascular regions of the root tip such as the quiescent center, columella and pre-cortical cells could also be detected in CRF reporter lines, indicating that expression of these genes is not limited to vascular tissue in the root. SICRF1 expression appeared to be confined to the vasculature in the root and is absent in the root tip; however, more extensive examination will be necessary to determine if this lack of expression is a clade-defining characteristic.

High levels of expression could be seen in the meristematic tip of shoot apices for clade I (AtCRF1 and 2) along with AtCRF6 from clade III (Fig. 2E). Each of these genes is regulated by cytokinin and could be involved in shoot apical meristem maintenance, known to be dependent upon cytokinin. After transition to flowering, no AtCRF expression could be detected in aerial meristems. However, genes from each clade examined were found in either floral or fruit tissues, often in vasculature tissue of those organs.

Cytokinin-regulated CRFs are involved in determining leaf vascular patterns

We have shown that CRFs from clades I and III are necessary for normal higher order vein patterning in leaves. The fact that genes from these clades are cytokinin inducible implies a link between cytokinin-regulated CRFs and leaf vein patterning (Table 1). Several CRFs could play a role in vascular development in leaves, similar to the cytokinin response regulator ARR1 shown to function as a mediator of hormone cross-talk between cytokinin and auxin in roots (Dello ioio et al. 2008). Interestingly CRFs have been shown to interact with ARR1 at the protein level (Cutcliffe et al. 2011). Previous study of the Arabidopsis histidine kinase family of cytokinin receptors has also linked them to vascular development as seen in triple receptor mutants (ahk2,3,4) with altered shoot and root vasculature (Nishimura et al. 2004, Hejatko et al. 2009). Although leaf vascular patterning of the ahk2,3,4 mutant was not quantified, its defect appears qualitatively similar to those found in crf2: reduced numbers of larger areoles. Several mutants related to auxin signaling and transport have irregular vascular patterning phenotypes often resulting in increased numbers of free-ending veins; frequently the result of vein discontinuity (Mattson et al. 2003, Dettmer et al. 2009). A significant increase of free-ending veins was observed in crf5 leaves; however, these do not appear to result from vein discontinuities since that would predict crf5 to have a reduced number of areoles that are larger in size, whereas the opposite was found. This suggests that the increase of free-ending veins is instead a result of an overall increase in vascular density, which was observed. Because AtCRF5 and AtCRF6 transcripts are abundant in the phloem, we have hypothesized that an increased density of veins found in clade III mutants may be an attempt to compensate for a diminished ability either to load or to unload the phloem.

Interestingly the observed areole phenotypes of crf2 leaves were the opposite of crf5: reduced numbers of larger areoles in crf2 vs. increased numbers of smaller areoles in crf5. While this shows that both clade I and III transcripts are necessary for normal areole patterning, they appear to regulate the process differentially. In fact, a similar examination of crf2,5 double mutants revealed no significant differences from the wild-type (Supplementary Table S1), implying that the two genes function in different pathways antagonistically regulating formation of areoles. As this result appears to mirror clade I and III expression differences seen during leaf development, it suggests that these genes play more complex roles than simply the regulation of areole number and size in vein patterning, such that analysis of multiple interclade mutant combinations is likely to be complex and potentially confounding for functional interpretation of individual genes.

Conclusions

In summary, a comparative genomic/transcriptomic examination of CRF protein sequences across land plants revealed that diversification of CRFs resulted in five clearly defined lineages represented in nearly every angiosperm genome. Despite limited phylogenetic resolution of sequences in CRF and AP2 domains, strong support for the presence of distinct CRF clades was found in C-terminal amino acid alignments, revealing novel and highly conserved regions for each clade. Species representing early branching lineages of angiosperms possess sequences...
from each clade, implying maintenance of an ancient functional diversification within this gene family. In contrast, CRF promoters from all clades have a significantly enriched TC-rich cis-element similar to motifs linked with phloem-specific expression. Both transcript and reporter gene analyses revealed that CRFs from every clade are expressed in vascular tissues, often the phloem, in agreement with promoter motif analysis. Together this suggests that while each clade may have specific functions including some unique expression patterns, all CRFs appear to be involved in vascular-related processes. Finally, an examination of CRF mutants revealed anomalous patterns in leaf vascular networks. The varying nature of mutant abnormalities found between clades supports the notion of clade-specific functional involvement in vascular patterning probably linked to cytokinin. This study has further elucidated the evolutionary relationships among angiosperm CRFs as well as general patterns of CRF expression. We believe this generates a better picture of the role these proteins play in plant development and hopefully provides the baseline for future work targeted at understanding specific roles of CRFs.

**Materials and Methods**

**Informatics**

Novel CRF gene sequences were identified in sequence databases (NCBI NR/NT, EST, GSS, HTGS and WGS), trace archives and several plant genome sequencing project databases with BLAST searches using blastx, tblastn and tblastx. Iterated profile searches were conducted using PSI-BLAST (Position-Specific Iterated BLAST) to target successive rounds of searching. All sequence databases were also queried using HMMER3 software employing hidden Markov probabilistic models capable of detecting remote homologs (Eddy 1998). CRF and AP2 domains from diverse CRF genes (as they were discovered) and variously broad consensus sequences of each domain were used as query sequences in multiple searches. Default settings of the BLAST programs were used except for the low complexity filter.

**Sequence analyses**

Promoter motif detection was performed with MEME (MEME version 4.1.0, http://meme.sdsc.edu/meme/) using standard settings and 1,000 bp of 20–50 different CRF sequences upstream of their ATG start sites. The most highly significant motifs identified from analysis of specific clades and all CRFs are presented as described in Fig. 4. All 403 CRF domain-containing proteins were aligned across their CRF and AP2 domains where possible. Amino acid alignments were performed initially with MAFFT (Katoh, 2008) and adjusted manually where necessary. Phylogenetic analyses were performed in PAUP* v4.0b11 using the Neighbor–Joining method, and in RAxML v7.3.0 using maximum likelihood as an optimality criterion. Only CRF and AP2 domain sequences were included in phylogenetic analyses (Supplementary Figs. S1–S3). The best fitting amino acid substitution model (PROTCATWAGF) was chosen by the RAxML model selection Perl script, which tests the fit of amino acid substitution models by optimizing model parameters and branch lengths on an initial JTT-based phylogenetic hypothesis. Non-parametric bootstrap procedures were used to estimate clade support (Felsenstein 1985).

**Hierarchical clustering analysis**

Clustering analyses were performed for Arabidopsis and Populus data sets individually. The 11 clade-specific Populus CRF genes (ERF78–87) were used to perform an ExCluster-Heatmap from all 384 microarray experiments available at the Populus Genome Integrative Explorer: PopGenIE v2.0 and are displayed in a Heatmap with the default settings of a Euclidean distance function and Ward hierarchical function. The seven clade-specific Arabidopsis AtCRF genes found on ATH1 microarrays were used to perform hierarchical clustering analysis of all 1,529 microarray experiments or subsets (anatomical/development and perturbations) available at Genevestigator using default settings and displayed as a Pearson correlation.

**Growth conditions**

*Arabidopsis thaliana* (Col-0). Sterilized seeds were germinated on Petri dishes containing 0.8% agar gel with MS salts (4.8 g l\(^{-1}\)) plus 1% sucrose medium of pH 5.7. Plants were grown under a 100 \(\mu\)mol m\(^{-2}\) s\(^{-1}\) light–dark cycle: 16 h, 22\(^\circ\)C/8 h, 18\(^\circ\)C in controlled-environment chambers. For extended growth, seedlings were transferred to soil (sunshine mix #8) and grown under the same conditions as above, but under \(~150\) \(\mu\)mol m\(^{-2}\) s\(^{-1}\) light.

*Solanum lycopersicum*. Micro-Tom cultivar plants were grown in Sunshine Mix #8 soil under a 150 \(\mu\)mol m\(^{-2}\) s\(^{-1}\) light–dark cycle: 16 h, 26\(^\circ\)C/8 h, 22\(^\circ\)C in controlled-environment chambers.

**Generation of transgenic plants**

The promoter (~2 kb upstream of ATG) sequences of AtCRF1–AtCRF6 and SICRF1; and promoter plus coding sequence of AtCRF1, AtCRF2, AtCRF4 and AtCRF5 were amplified using sequence-specific primers with att-B sites. PCR products were cloned into the pDONR221 entry vector (Invitrogen) and plasmids were generated using the Invitrogen GATEWAY™ system according to the manufacturer’s instructions. Destination vectors pBGWFS7 or pKGWFS7 (Karimi et al. 2002) were used to create transcriptional and translational fusion expression vectors. For Arabidopsis, plasmids were transformed into *Agrobacterium tumefaciens* C58 into electroporation, and plants were transformed by agrobacteria using the floral dip method (Clough and Bent 1998), with marker-selected homozygous lines used for analyses. For tomato, vector pKGWFS7 containing the SICRF1 promoter was sent to the Plant Transformation Research Center (PTRC) at University of California Riverside and transformed into Micro-Tom plants as a service. Homozygous plants were used for all analyses. Between six
and 15 independent transgenic lines of all plants were examined, from which a single representative line is shown.

Histochemical analysis

Analysis of GUS activity from various aged tissues noted in the text were vacuum infiltrated for 20 min with X-gluc buffer (Weigel and Glazebrook 2002) then further incubated at 37°C for 2–6 h for Arabidopsis and overnight for tomato. Tissue was then cleared in 70% ethanol at room temperature or 50°C, and examined. Mounted sections were fixed, embedded in paraffin, cut into 10 μm sections and viewed using a Nikon Eclipse 80i microscope. Photos were taken with a Qimaging Fast 1394 digital camera and are presented as composite images using Adobe Photoshop CS3 without altering the original integrity.

Extraction of vascular-enriched tissue

Fine-tip forceps slightly bent inward were used to gently grasp and remove the mid-vein from fully expanded Arabidopsis leaves (numbers 5–8 at bolting) and tomato leaflets. Mid-veins and the remaining blades were immediately frozen in liquid nitrogen, after which RNA was extracted, reverse transcribed and used for qPCR analysis. Each sample was composed of pooled RNA from multiple leaves.

RNA isolation, cDNA synthesis and quantitative PCR

RNA was isolated with a Qiagen RNeasy Kit according to the manufacturer’s instructions. RNA concentrations were equalized and pooled; reverse transcription was carried out using Qiagen qScript cDNA supermix. The resulting cDNA was diluted prior to qPCR performed with the SYBR-Green chemistry in an Eppendorf Mastercycler ep realplex. Arabidopsis reactions contained 9 μl of SYBR-Green supermix, 10.76 μl of cDNA template and 0.12 μl of 100 μM gene-specific primers. The qPCR program consisted of one cycle at 95°C, followed by 40 cycles of 15 s at 95°C, 45 s at 57°C and 30 s at 68°C. Tomato reactions contained 9 μl of SYBR-Green supermix, 2 μl of cDNA template, 3 μl of 4 μM gene-specific primers and 3 μl of sterile water. The qPCR program consisted of one cycle at 95°C, followed by 40 cycles of 15 s at 95°C, 30 s at 56°C and 35 s at 68°C. Relative expression data are the mean of two biological replicates each consisting of a pool from at least six individual plants from which at least four technical replicates were performed. Expression levels of reference genes for normalization were confirmed with reverse transcription–PCR as in Rashotte et al. (2006) and Shi et al. (2012).

Measurement of leaf vasculature

The homozygous T-DNA insertion mutant alleles analyzed were obtained from the ABRC stock center or as previously described: crf1-1, crf2-1, crf3-3 (salk_138253), crf5-6 (salk_024228), crf6-2 and crf9-1 (sail_770_B09) and compared with the wild type (Col) (Alonso et al. 2003, Rashotte et al. 2006). All lines were confirmed as homozygous T-DNA insertions using PCR with gene-specific and T-DNA-specific left border primers. For analysis, leaves 3–4 were excised at bolting, cleared in 95% ethanol overnight, then rehydrated in a graded ethanol series and stained overnight with 0.25% basic fuchsin. Stain was replaced with slightly alkaline (one drop of 1 N NaOH in 100 ml) water then rinsed until clear of excess stain. Leaves were further cleared of background stain with a 3:1 ethanol:HCl solution and photographed as described above. Images were processed and analyzed using LEAF GUI software (Price et al. 2011). For each leaf, the first area bounded by the mid-vein and two first-order lateral veins fully within the right blade was analyzed. Sample sets contained 9–30 individual leaves from ≥2 independently grown repetitions; mean, standard error and P-value based on one-tailed Student’s t-test were computed in Microsoft Excel 2010.

Supplementary data

Supplementary data are available at PCP online.

Funding

This work was funded by the US Department of Agriculture-National Research Initiative [grant 2008-35304-04457]; Alabama Agricultural Experiment Station-HATCH [grants 370220-310007-2055 and 103600-310010-2055].

Acknowledgments

We gratefully acknowledge members of the Rashotte lab for critical reading of the manuscript.

References


