The floral homeotic protein APETALA2 recognizes and acts through an AT-rich sequence element

Thanh Theresa Dinh¹, Thomas Girke¹, Xigang Liu¹, Levi Yant²*, Markus Schmid² and Xuemei Chen¹,³‡

SUMMARY
Cell fate specification in development requires transcription factors for proper regulation of gene expression. In Arabidopsis, transcription factors encoded by four classes of homeotic genes, A, B, C and E, act in a combinatorial manner to control proper floral organ identity. The A-class gene APETALA2 (AP2) promotes sepal and petal identities in whorls 1 and 2 and restricts the expression of the C-class gene AGAMOUS (AG) from whorls 1 and 2. However, it is unknown how AP2 performs these functions. Unlike the other highly characterized floral homeotic proteins containing MADS domains, AP2 has two DNA-binding domains referred to as the AP2 domains and its DNA recognition sequence is still unknown. Here, we show that the second AP2 domain in AP2 binds a non-canonical AT-rich target sequence, and, using a GUS reporter system, we demonstrate that the presence of this sequence in vivo is important for the restriction of AG expression in Arabidopsis. Furthermore, we show that AP2 binds AG in vitro and directly regulates AG expression through this sequence element. Computational analysis reveals that the binding site is highly conserved in the second intron of AG orthologs throughout Brassicaceae. By uncovering a biologically relevant sequence element, this work shows that AP2 domains have wide-ranging target specificities and provides a missing link in the mechanisms that underlie flower development. It also sets the foundation for understanding the basis of the broad biological functions of AP2 in Arabidopsis, as well as the divergent biological functions of AP2 orthologs in dicotyledonous plants.

KEY WORDS: AGAMOUS, ANT, AP2, APETALA2, DNA binding, Flower development

INTRODUCTION
The flower is an evolutionary innovation that contributes to the success of angiosperms. Dicotyledonous flowers are composed of four major types of organs: sepal, petal, stamen and carpel. Four major classes of homeotic genes, A, B, C and E, specify the four floral organ types in a combinatorial manner (reviewed by Krizek and Fletcher, 2005). A-class genes, APETAL1 and APETAL2 (AP1 and AP2), together with the E-class genes, confer sepal identity in the first whorl. Petal identity is determined by the activities of A-, B- and E-class genes in the second whorl. The C-class gene, AGAMOUS (AG), together with B- and E-class genes, specifies stamen identity in the third whorl. Carpel identity is conferred by C- and E-class activities in the fourth whorl. With the exception of AP2, all floral homeotic genes encode MADS-domain-containing proteins for which DNA binding and dimerization/multimerization specificities have been extensively characterized (reviewed by Immink et al., 2010). By contrast, the activity of AP2 as a transcription factor in flower development is poorly understood. In addition to promoting sepal and petal identities, AP2 restricts AG expression to the inner two whorls (Bowman et al., 1991; Drews et al., 1991). Previous studies using a GUS reporter system have shown that the 3-kb AG second intron contains sequence elements required for its proper expression, including responsiveness to repression by AP2 (Bomblies et al., 1999; Deyholos and Sieburth, 2000).

AP2 is the founding member of a family of 144 genes that encode at least one AP2 DNA-binding domain in Arabidopsis; the biological functions of this family range from development to stress and defense responses (Jofuku et al., 1994; Weigel, 1995; Okamuro et al., 1997; Riehmann and Meyerowitz, 1998). This DNA-binding domain was thought to be plant specific but computational analysis has identified this DNA binding domain in species such as Tetrahymena (Wuitschick et al., 2004) and Plasmodium (Yuda et al., 2009). Plant AP2 domain-containing genes were categorized into five subfamilies (Sakuma et al., 2002). Members of the AP2-like subfamily generally contain two AP2 DNA-binding domains, AP2R1 and AP2R2 (Kim et al., 2006; Shigyo et al., 2006). In the second to fifth subfamilies, the ERF-like, DREB-like, RAV-like and others, members contain one AP2 DNA-binding domain and are involved in abiotic and biotic stress responses.

The DNA-binding properties of AP2 domain proteins from various subfamilies have been studied. Members of the ERF-like and DREB-like subfamily bind well-documented GC-rich motifs (Ohme-Takagi and Shinshi, 1995; Stockinger et al., 1997; Hao et al., 1998; Liu et al., 1998). The AP2 domain in a RAV-like family member binds a CAACA motif (Kagaya et al., 1999). ANT is the only protein in the AP2-like subfamily for which DNA-binding properties have been studied (Nole-Wilson and Krizek, 2000; Krizek, 2003). Despite the crucial role of AP2 in flower development and the diversity of its targets (Yant et al., 2010), the DNA binding specificity of AP2 has never been characterized.

To better understand the molecular functions of AP2, we sought to determine its binding consensus sequence in vitro and characterize the relevance of the sequence in vivo. Here, we report that AP2R2 specifically binds the TTTGTT or AACAAA motif in vitro. We show that these motifs within the 2nd intron of AG are important for restricting AG expression to the inner two whorls in vivo. In silico analysis of 2nd intron sequences from AG orthologs

¹Department of Botany and Plant Sciences, Center for Plant Cell Biology, Institute of Integrative Genome Biology, University of California, Riverside, CA 92521, USA.
²Max Planck Institute for Developmental Biology, Department of Molecular Biology, D-72076 Tübingen, Germany.
³Howard Hughes Medical Institute, University of California, Riverside, CA 92521, USA.
*Present address: Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA.
‡Author for correspondence (xuemei.chen@ucr.edu)

Accepted 19 March 2012
AP2 binds an AT-rich sequence

uncovers strong conservation of this element in the *Brassicaceae* family. Furthermore, we found that AP2 directly regulates *AG* in young flowers through these elements. Curiously, the AP2 full-length protein binds DNA with no apparent specificity in vitro, suggesting that other factors influence its DNA binding specificity in vivo. These findings establish a missing link in the mechanisms underlying flower development, shed light on the molecular function of AP2, and set the foundation for further appreciation of the molecular basis for the broad biological functions of AP2.

**MATERIALS AND METHODS**

**Plasmid construction**

To express the AP2R1, AP2R2 and AP2R1R2 domains of AP2 in *E. coli*, the corresponding coding regions from the AP2 cDNA were amplified by PCR (supplementary material Table S1) and cloned into the pET21-A vector using BamHI and EcoRI sites (Novagen). To express the full-length AP2 protein, the entire coding region of AP2 was cloned in-frame to an N-terminal MBP and His tag using BamHI and EcoRI sites in the pcDNA7 XF0510 MBP-LIC vector (a gift from Dr Xiaofeng Cao, Institute of Genetics and Developmental Biology, Beijing, China).

For in vivo analysis of the *AG* 2nd intron, the region of the *AG* 2nd intron in the KB31 construct (Bombiles et al., 1999) was amplified and cloned into PCR2.1 (Invitrogen). Site-directed mutagenesis was performed (supplementary material Table S1) to introduce mutations into each of the two AP2-binding sites. The wild-type and mutant KB31 fragments were then cloned into pD991 (Tilly et al., 1998) using BamHI and HindIII sites. The 35S::AP2m3-GR construct was generated as described previously (Yant et al., 2010).

**Protein expression and purification**

The pET21A-AP2R1, AP2R2 and AP2R1R2, and the MBP-AP2 full-length protein plasmids were transformed into *E. coli* BL21. Protein expression and purification were carried out as previously described (Smith et al., 2002; Husbands et al., 2007) and purified proteins were quantified against BSA.

**Selection affinity and amplification binding (SAAB) assay**

Either 200 ng or 500 ng of doubly affinity-purified (with Ni²⁺ beads and T7 antibody) and desalted AP2R1, AP2R2 or AP2R1R2 was subjected to a SAAB assay as previously described (Smith et al., 2002; Husbands et al., 2007). Briefly, the protein-bead mixture was divided into six tubes. In the first tube, a pool of random, double-stranded oligonucleotides (supplementary material Table S1) was added and incubated for 4 hours with the protein-bead mixture. The DNA bound by the protein-bead complex was eluted and PCR was performed to amplify the bound sequences. An aliquot of the PCR reaction was added to some reactions at 1-2× the amount of the AP2R2 protein to obtain super-shifts.

**GUS staining and microscopy**

Inflorescences were stained for GUS activity and processed for sectioning as previously described (Sieburth and Mayerowitz, 1997). Slides were viewed under a Leica DMR compound microscope and images were taken with a Spot digital camera (Diagnostic Instruments).

**Induction and expression analysis of 3SS::AP2m3-GR**

The inflorescences of 35S::AP2m3-GR ap2-2 plants were treated once with a solution of 10 μM dexamethasone (DEX)/0.015% Silwet with or without 10 μM cyclohexamide (CHX) (Fisher). Six hours later, the treated inflorescences were dissected to remove stage 8 and older flowers. Total RNA was isolated from the dissected inflorescences and subjected to DNaseI treatment and reverse transcription. RT-PCR was performed on the cDNAs using primers specific for *AG* and UBQ5 (supplementary material Table S1). Real-time RT-PCR was performed on the same cDNAs using a real-time PCR SYBR Green system (BioRad). Three technical replicates were performed for each real-time RT-PCR. Three biological replicates of DEX induction and real-time RT-PCR were performed. Error bars represent the standard deviation from three technical replicates.

**Chromatin immunoprecipitation**

Chromatin immunoprecipitation (ChIP) experiments were performed on two biological replicates following previously described protocols (Gomez-Mena et al., 2005; Mathieu et al., 2009; Yant et al., 2010). The input and ChIP samples were subjected to real-time PCR (BioRad). Three technical replicates were performed. The data were analyzed as previously described (Wierzbicki et al., 2008).

**Sequence analysis**

All 2nd intron sequences from *AG* orthologs were downloaded from GenBank (supplementary material Table S2). The start and end positions, provided by Hong et al. (Hong et al., 2003) and those specified in the GenBank annotations files, were used to parse the introns from their source sequences and to bring them into their proper sense orientation (supplementary material Table S2). Sequence manipulations and analyses were performed with custom scripts that are based on the Biostrings package of the statistical programming environment R (Morgan et al., 2009; R Development Core Team, 2010). Multiple sequence alignments (MSAs) were computed with the dialign2-2 software from Morgenstern (Morgenstern, 2004) using the default parameters in the DNA mode. A sliding window analysis was performed to visualize the degree of conservation in the final MSA. For this, the relative conservation of each base was calculated at each position where a value of 1.0 indicates perfect conservation of one base (disregarding gaps) and a value of 0 indicates equal representation of all four bases. For plotting purposes, these conservation values were smoothed by calculating their mean for a sliding window size of 20 nucleotides along all MSA positions. Pattern searches were performed with the matchPattern function of the Biostrings package (Morgan et al., 2009).

**RESULTS**

**AP2R2 binds a novel consensus sequence**

To begin uncovering the molecular mechanisms underlying the role of AP2 in development, we sought to identify its binding sequence by performing a SAAB assay with AP2R1, AP2R2 or AP2R1R2 doubly purified based on their N- and C-terminal tags. When the SAAB assay was performed for the purified AP2R2 (supplementary material Fig. S1A), amplified DNA from the bound fraction could be detected starting from cycle 4 (supplementary
Fig. 1. AP2R2 binds the TTTGTT and/or AACAAA motif in vitro. (A) The sequences of the DNA probes (only one strand is shown). The consensus sites are underlined. αβ, a probe containing two sites; Δα or Δβ, probes containing only one site. (B) AP2R2 binds the αβ probe containing both sites. The shifted band that represents binding (lanes 1 and 2) is indicated by an arrow. The binding was lost upon the addition of 20× cold competitor (lane 3). The + sign indicates that 200 ng of AP2R2 protein was included; for lane 1, 100 ng AP2R2 was used. In lanes 4 and 5, 200 ng and 400 ng, respectively, of T7 and His antibodies were added to the reactions. The bands indicated by the stars and double stars probably represent the super-shifts. As a control, we also added 400 ng His and T7 antibodies to the reaction in the absence of AP2R2 and we did not observe any of the same band shifts as seen in the presence of AP2R2 (lane 6). Lane 7 represents the free probe lane. (C) One site is sufficient for binding by AP2R2. A shift was observed, as indicated by the arrow, with probes containing one site (lanes 2 and 5). Binding was lost upon addition of the cold competitors (lanes 3 and 6). AP2R2 binding was abolished when both sites were mutated (lane 8). Lanes 1, 4 and 7 are the free probe lanes.

When the same procedure was applied to purified AP2R1 (supplementary material Fig. S2A), no bound DNA was detectable by PCR, indicating that AP2R1 did not bind DNA in vitro (supplementary material Fig. S2B). The lack of DNA recovered from the AP2R1 SAAB assay was not due to loss of the protein during the procedure because the AP2R1 protein was present on the beads throughout the experiment (supplementary material Fig. S2C). The SAAB assay was also performed for purified AP2R1R2 (supplementary material Fig. S3A). DNA bound to AP2R1R2 was detectable from cycles 3 to 6 (supplementary material Fig. S3B); however, sequencing of cloned DNA bound to AP2R1R2 from cycle 6 did not reveal any obvious consensus sequence (supplementary material Fig. S3C). Ten out of 25 unique sequences contained one of the sites bound by AP2R2 (supplementary material Fig. S3C; and data not shown), one had a site with one nucleotide change (clone 3), and some of the other sequences were GC rich.

Next, we tested whether both sites were necessary for AP2R2 binding. We mutated the AACAAA site to AGGTGA and the TTTGTT site to TCCACT (Fig. 1A). The EMSA showed that AP2R2 was still able to bind probes that had one intact site (Fig. 1B, lanes 1, 2 and 5). As the experiments were conducted with AP2R2 purified from E. coli, there was a possibility that the observed shift may be due to a contaminating protein instead of AP2R2 (although AP2R2 was the only protein detected by Coomassie staining in the protein fraction; supplementary material Fig. S1A). If the observed binding was specifically caused by AP2R2, the addition of His and T7 antibodies would generate a supershift as AP2R2 had both a T7 and a His tag. We observed that the inclusion of the His and T7 antibodies resulted in super-shifted bands when compared with AP2R2 alone (Fig. 1B, compare lanes 4 and 5 to lane 2, stars), confirming that it was indeed AP2R2 itself that bound the probe. As a control, we also added 400 ng His and T7 antibodies to the reaction in the absence of AP2R2 and we did not observe any of the same band shifts as seen in the presence of AP2R2 (Fig. 1B, lane 6).

Next, we tested whether both sites were necessary for AP2R2 binding. We mutated the AACAAA site to AGGTGA and the TTTGTT site to TCCACT (Fig. 1A). The EMSA showed that AP2R2 was still able to bind probes that had one intact site (Fig. 1C, lanes 2 and 5, arrow). The shift was lost upon addition of the corresponding cold competitor (Fig. 1C, lanes 3 and 6). Loss of both sites, however, completely abolished binding (Fig. 1C, lane 8). These results demonstrate that AP2R2 binds AACAAA and/or TTTGTT in vitro.

ANT, the only protein characterized in terms of DNA-binding properties in the AP2-like subfamily of AP2 domain-containing proteins, has been shown to bind a loose and long consensus sequence,
In silico analysis indicated that the KB31 region contained two AP2R2-binding sites at the 3' end, which we termed A and B (Fig. 3A). Thus, we proceeded to test whether AP2R2 was able to bind this region of the AG second intron in vitro. Using primers encompassing this region (supplementary material Table S1), we generated a 167 bp probe (Fig. 3A) to perform an EMSA. Indeed, a shift was found with AP2R2 and the binding was stronger as we added increasing amounts of MBP-AP2 full-length protein (labeled as MBP-AP2) on the 167 bp fragment. MBP-AP2 full-length protein bound both the wild-type (lanes 2-7; arrow) and mutated (lanes 12 and 13) probes, and the binding was lost upon addition of the cold competitor (lanes 8 and 9). Binding was minimal for the ΔΔ probe containing mutated versions of both sites (rectangles). This 167 bp region was amplified with primers encompassing this region (supplementary material Fig. S5). We observed binding of AP2R2 to the ΔΔ probe (supplementary material Fig. S5B, lane 2) but not to the canonical GCC-box or two other random sequences (supplementary material Fig. S5B, lanes 5, 7 and 9).

**AP2 full-length protein lacks obvious DNA-binding specificity in vitro**

Next, to determine whether the full-length AP2 protein could bind the consensus sequence TTTGTT, we cloned the AP2 full-length protein into a vector containing an N-terminal His and MBP tag (MBP-AP2), and purified and desalted it (supplementary material Fig. S6A). Owing to the large size of MBP-AP2, 400 ng was used to perform the EMSA. Using the same probes as in the AP2R2 EMSAs (Fig. 1A), we found that MBP-AP2 was able to bind probes containing one or both consensus sequences (supplementary material Fig. S7A, lanes 2 and 6, arrow). Interestingly, MBP-AP2 was also able to bind the probe with both sites mutated (supplementary material Fig. S7A, lane 9, arrow). To assess the binding specificity of the AP2 full-length protein further, we performed EMSAs with a probe containing the canonical GCC-box and two random probes (supplementary material Fig. S5A). Interestingly, MBP-AP2 could bind all of the probes in vitro (supplementary material Fig. S7B, lanes 2, 5 and 8, arrow). To address whether the MBP tag may be binding the probes, an equal amount of MBP was added to the reactions for the EMSA (supplementary material Fig. S6B). MBP alone did not bind any DNA sequences (supplementary material Fig. S7A, lanes 3, 7 and 10; S7B, lanes 3, 6 and 9).

**AP2R2 binds AG 2nd intron in vitro**

Considering that AP2R2, but not AP2 full-length protein, specifically bound the consensus sequence, we sought to test whether the AP2R2-binding site had biological relevance. Previous studies have shown that AP2 represses AG expression (Drews et al., 1991), but it is still unknown whether AG is a direct target of AP2, although AP2 has been shown to bind the AG 2nd intron in vivo in our previous ChIP-seq analysis (Yant et al., 2010). Characterization of the AG 2nd intron has identified a 750 bp region that, when fused to the GUS reporter in a construct termed KB31, confers AP2 responsiveness to GUS (Bomblies et al., 1999; Deyholos and Sieburth, 2000). In silico analysis indicated that the KB31 region contained two AP2R2-binding sites at the 3’ end, which we termed A and B (Fig. 3A). Thus, we proceeded to test whether AP2R2 was able to bind this region of the AG second intron in vitro. Using primers encompassing this region (supplementary material Table S1), we generated a 167 bp probe (Fig. 3A) to perform an EMSA. Indeed, a shift was found with AP2R2 and the binding was stronger as we added increasing amounts of MBP-AP2 full-length protein (labeled as MBP-AP2) on the 167 bp fragment. MBP-AP2 full-length protein bound both the wild-type (lanes 3-8) and mutated (lanes 12 and 13) probes, and the binding was lost upon addition of the cold competitor (lanes 9 and 14). Binding was observed starting at 243 ng of the protein (lane 6), and was most obvious with 355 ng (lane 7) and 429 ng (lane 8) of the protein. MBP (M) alone did not bind any probe (lanes 2 and 11). Free=0.1 pmol/reaction (lanes 1 and 10). Triangles depict increasing amounts of protein added. The arrow marks the shift.
amounts of the protein (Fig. 3B, lanes 2-7, arrow). Furthermore, the binding was lost upon the addition of 20× or 40× cold competitor (Fig. 3B, lanes 8 and 9, respectively) to the binding reaction. To test whether the observed binding required the two sites within the 167 bp sequence, we performed site-directed mutagenesis to mutate both sites (Fig. 3A). With the ΔAB probe, binding was diminished greatly (Fig. 3B), showing that AP2R2 binds the 167 bp region via the two elements in vitro. In addition, and consistent with prior results, gel shifts showed binding of the MBP-AP2 full-length protein to both the wild-type and the ΔAB probe (Fig. 3C, lanes 3-8, 12 and 13, arrow). This binding was abolished upon addition of 40× cold competitor (Fig. 3C, lanes 9 and 14) and MBP alone did not bind either probe (Fig. 3C, lanes 2 and 11).

**The AP2R2 binding sites are important for the restriction of AG expression in vivo**

To evaluate the importance of the AP2R2 binding sites in the AG 2nd intron in vivo, we used the KB31 GUS reporter, which had been shown to report faithfully the endogenous domains of AG expression and to respond to the regulation by AP2 (Bombilies et al., 1999; Deyholos and Sieburth, 2000). We cloned the 750 bp KB31 region from the AG 2nd intron containing either the wild-type or mutant (ΔAB) sites into a GUS expression vector with a minimal (~60) 35S promoter (Tilly et al., 1998). The constructs were introduced into rdr-6-11 to prevent post-transcriptional gene silencing of the transgenes (Dalmary et al., 2000; Mourrain et al., 2000). For the wild-type construct, GUS staining of inflorescences from 99 independent T1 transgenic plants showed that 74 recapitulated the proper AG expression patterns (Fig. 4A,C,E). Twenty-two inflorescences did not show any GUS staining and three did not recapitulate the proper AG expression patterns. For the ΔAB construct, however, 71% of the 27 independent transformants showed expansion of the GUS expression domain to the outer two whorls (Fig. 4B,D,F) in all stages of flower development, with the remainder showing the correct expression patterns. Therefore, the A and B sites are important for the restriction of AG expression to the inner two floral whorls.

**AP2 directly regulates AG in young flowers through the binding sites**

The fact that the AP2R2 binding sites in the 2nd intron of AG are important for the restriction of AG expression to the inner two whorls implies that AP2 is a direct regulator of AG. To address whether AP2 acts on AG directly, we used a rat glucocorticoid receptor (GR)-induction system that has been widely used in Arabidopsis as a method to establish direct relationships between a transcription factor and its targets (Sablokowsky and Meyerowitz, 1998; Wagner et al., 1999; Ito et al., 2004; William et al., 2004). Because AP2 is targeted by miR172 and transgenes containing miRNA target sites are readily silenced in vivo, we fused a miR172-resistant AP2 cDNA (AP2m3) (Chen, 2004) to GR. The 35S::AP2m3-GR construct was transformed into the progeny of ap2-2/+ plants. After obtaining single-locus insertion transformants of the ap2-2/+ genotype, single and continuous treatments of 10 μM DEX were performed to determine the functionality of the transgene. A single DEX treatment of inflorescences was not sufficient to induce the AP2m3 phenotype (data not shown) (Chen, 2004). Continuous treatments (once a day for 1 week), however, led to the induction of the AP2m3 phenotype, thus showing that the transgene was functional (Fig. 5A,B).

To determine whether AP2 directly represses AG expression, we subjected 35S::AP2m3-GR ap2-2 inflorescences to a single treatment of cyclohexamide (CHX) with or without DEX. After 6 hours, inflorescence tissue was micro-dissected to remove stage 8 and older flowers. RT-PCR was performed to measure AG mRNA levels. We found that upon induction of AP2m3-GR, AG mRNA levels decreased in young flowers (Fig. 5C). Real-time RT-PCR of three biological replicates revealed a 50% decrease in AG transcript levels upon AP2m3-GR induction (Fig. 5D). Therefore, AG is likely to be a direct target of AP2.

Next, we sought to determine whether AP2 regulates AG through the two AP2R2-binding sites. If AP2 acts through the two sites, we would expect KB31, but not KB31ΔAB, to be repressed by AP2. KB31 and KB31ΔAB transgenic lines harboring a single transgene locus were identified and crossed into the 35S::AP2m3-GR ap2-2 background. Homozygous KB31 35S::AP2m3-GR ap2-2 or KB31ΔAB 35S::AP2m3-GR ap2-2 inflorescences were treated with DMSO or DEX for 6 hours and GUS expression was determined by real-time RT-PCR. DEX induction caused a decrease in GUS mRNA levels in KB31 but not in KB31ΔAB (Fig. 5E,F). Therefore, AP2 represses AG through the two AP2R2-binding sites in vivo.

**AP2 binds AG 2nd intron in vivo**

To test whether AP2 is associated with the AG 2nd intron in vivo, we performed ChIP assays using anti-AP2 antibodies (Mlotshwa et al., 2006). The antibodies were directed against a C-terminal region of AP2 that is predicted to be absent in the ap2-2 mutant. From a ChIP-seq experiment conducted with these antibodies on whole inflorescences, genome-wide AP2-binding sites were uncovered (Yant et al., 2010). The ChIP-seq effort identified a region in the 5′ end of the KB31 fragment that was bound by AP2, which we named region II (Fig. 6A). This region did not overlap with the region containing our binding sites, AB (Fig. 6A). To specifically test
whether AP2 binds the AB region, especially in young flowers, we performed ChIP experiments with dissected inflorescences containing stages 7 and younger flowers and used region II as the positive control. We were able to find enrichment of AP2 within the AB region as well as region II in two biological replicates (Fig. 6; supplementary material Fig. S8A and Table S1), indicating that some part of that region may be bound by AP2. We divided the region into four parts (1-4; supplementary material Fig. S8A and Table S1), including a region directly upstream and downstream. We found that AP2R2 bound part 3 of region II (supplementary material Fig. S8B) and upon further inspection, the 5’ end of part 3 contained an AP2R2 motif. The apparent exception appeared to be part 2, which was embedded within a short region that was conserved throughout the introns (supplementary material Fig. S10). At the A-site, the TTTGTT (or AACAA) motif was present multiple times in each of the introns (supplementary material Figs S9, S10). The B-site was not found at invariant positions as revealed by the sliding window conservation analysis (supplementary material Fig. S11). However, it is possible that AP2-mediated restriction of AG expression is conserved in species closer to Arabidopsis. Thus, we sought to determine whether the AP2R2-binding consensus sequence was conserved in Brassicaceae. Using AG 2nd intron sequences from 29 Brassicaceae species (Hong et al., 2003), we performed multiple sequence alignment as well as a sliding window conservation analysis to identify regions in the 2nd intron that are conserved both in sequence and in position. Although the TTTGTT (or AACAA) motif was present multiple times in each of the introns (supplementary material Fig. S10), the A-site (Fig. 3A) was embedded within a short region that was conserved throughout Brassicaceae both in sequence and in position within the introns, as revealed by the sliding window analysis (supplementary material Figs S9, S10). The B-site was not found at invariant positions among the introns (supplementary material Fig. S10). At the A-site, 28 of the 29 Brassicaceae species showed a perfect match to the TTTGTT pattern. The apparent exception appeared to be Thysanocarpus (AY253255) with a single nucleotide change in the motif. When the analyses included the 2nd intron of AG homologs from Antirrhinum majus (AY935269), Lycopersicon esculentum (AY254705) or Cucumis sativus (AY254702 and AY254704) belonging to Veronicaceae, Solanaceae or Cucurbitaceae, respectively, the divergence in these sequences was too high to compute reliable multiple sequence alignments of the introns, thus precluding any conclusions on the conservation of this motif outside of Brassicaceae.

**DISCUSSION**

**DNA binding specificities of AP2 domain proteins**

Genes encoding one or more AP2 DNA-binding domains are categorized under five subfamilies: DREB-like, ERF-like, RAV-like, AP2-like and others (Sakuma et al., 2002). The AP2-like subfamily, which can be further divided into two lineages, ANT and euAP2, is the only subfamily that contains two AP2 domains.
Single AP2 domain containing proteins of the other subfamilies bind to highly specific, mostly GC-rich sequence motifs (Ohme-Takagi and Shinshi, 1995; Stockinger et al., 1997; Hao et al., 1998; Liu et al., 1998). Only the target sequence of a single member of the AP2-like subfamily, ANT, has been reported. ANT binds a long and loose consensus sequence that is also GC rich (Nole-Wilson and Krizek, 2000). In contrast to conventional GC-rich target sequences of these characterized AP2-domain proteins, AP2R2 is highly specific for the AT-rich consensus sequences TTTGTT or AACAAA. The AP2 domain in RAV1 also binds a non-GC rich sequence CAACA (Kagaya et al., 1999). Two AP2 domain proteins from Plasmodium were found to bind the consensus sequences TGCAATGCA and GTGCAC, which are different from the target sequences of all plant AP2 domain proteins characterized to date (De Silva et al., 2008). Collectively, these studies show that AP2 domains have wide ranging target specificities. Consistent with this, the AP2 domain of AtERF1 and the AP2 domains of ANT appear to use largely non-conserved amino acids for DNA binding (Allen et al., 1998; Krizek, 2003) (supplementary material Fig. S11). The fact that AP2R1 does not appreciably bind any DNA sequences in vitro raises the possibility that some AP2 domains function in processes other than DNA binding.

It is useful to compare and contrast the DNA-binding specificities of ANT and AP2 as representatives of the two lineages within the AP2-like subfamily. In vitro selection of DNA sequences bound by ANT-AP2R1R2 led to the identification of a long consensus sequence (Nole-Wilson and Krizek, 2000; Krizek, 2003). In our study, we found that AP2R1R2 bound DNA in vitro, but no consensus sequence could be identified. We note that both ANT-AP2R1R2 and AP2R1R2 have poor DNA-binding specificities (as exemplified by the loose ANT consensus sequence and the lack of obvious consensus motifs from AP2R1R2-bound sequences). We also note that ANT-AP2R1R2 and AP2R1R2 have differences in their binding preferences. Although ANT-AP2R1R2 binds GC-rich sequences, AP2R1R2 probably prefers the TTTGTT or AACAAA motif as this motif was in 10 out of 25 clones from the SAAB assay. Moreover, not all clones from the SAAB assay were GC rich. Consistently as well, this could be explained by the addition of 10 amino acids in ANTR1 versus AP2R1 and a single amino acid addition in ANTR2 versus AP2R2 (supplementary material Fig. S11).

**AP2 DNA-binding specificities in vivo**

AP2 full-length protein was able to bind all probes that were tested in vitro (Fig. 3; supplementary material Figs S7, S8). This led us to question the specificity of AP2 DNA binding in vivo, especially in relation to its biological functions. AP2 has diverse biological functions such as seed development (Jofuku et al., 2003; Ohto et al., 2009), shoot apical meristem maintenance (Würschum et al., 2006), control of floral timing (Yant et al., 2010), preventing replum overgrowth during fruit development (Ripoll et al., 2011), establishment of floral meristem identity (Schultz and Haughn, 1993; Shannon and Meeks-Wagner, 1993), floral organ specification (Bowman et al., 1989; Kunst et al., 1989) and the regulation of homeotic gene expression (Drews et al., 1991). Perhaps the lack of strong inherent DNA-binding specificities underlies the diverse biological roles of AP2. Whole-genome ChIP-seq experiments identified more than 2000 sites that are bound by AP2 in vivo (Yant et al., 2010), highlighting the potential of AP2 in influencing the expression of a large number of genes. However, computational analyses failed to uncover a consensus sequence that is enriched in regions bound by AP2 in vivo (data not shown). In addition, owing to its AT-rich nature, we were not able to state that the AP2R2-binding site was statistically significant among the targets. However, it is interesting to note that in the sum sequence space of the 2275 bound sites there are 445 instances of AACAAA and 473 of TTTGTT (Yant et al., 2010). The lack of ability to find an AP2 consensus sequence could be reflective of its diverse roles in development. In addition, the discrepancy in the two sites (II and AB) that we found in the two ChIP experiments (Yant et al., 2010) (this study) could be due to tissue differences.

Despite the large number of in vivo binding sites, AP2 is still selective in its DNA binding in vivo, in contrast to its largely non-specific DNA binding in vitro. One potential mechanism underlying the in vivo specificity is that it might be conferred by other DNA-binding proteins that interact with AP2. In this scenario, the largely non-specific DNA binding by AP2 enhances the binding of other transcription factors at specific sites. The promiscuous binding of MBP-AP2 to all DNA sequences in vitro lends itself to this hypothesis as it could be feasible that AP2 full-length itself, as a regulator of diverse functions, would have specific binding abilities, depending on its protein-binding partners, that may modulate its activity in vivo. Another potential mechanism is that other factors interact with AP2R1 to allow AP2R2 to specifically interact with DNA. We prefer this explanation as the AP2R2-binding sites in the AG 2nd intron are indeed important for the function of AP2 in vivo. Moreover, the AP2R2 consensus sequence was recovered in 10 out of 25 clones in the AP2-R1R2 SAAB assay, suggesting that there is some inherent affinity of the AP2 domains for the TTTGTT or AACAAA consensus sequence. Both scenarios may occur in vivo, in which case the AP2R2 consensus sequence would only present at some of the in vivo AP2-binding sites.

**AP2 directly regulates AG**

AP2 has long been known to be essential in establishing the inner two whorl-specific pattern of AG expression (Drews et al., 1991). In ap2 loss-of-function mutants, AG expression expands into the outer two whorls. Using the GUS reporter system, elements responsive to AP2 regulation have been mapped to at least two regions in the AG 2nd intron (Bomblies et al., 1999; Deyholos and Sieburth, 2000). However, it was not known whether AP2 regulates AG expression directly. We found that a 750 bp AP2-responsive region contains two AP2R2 consensus sequences. Site-directed mutagenesis experiments indicated that the two sites were important for the negative regulation of AG by AP2. In addition, we found that this negative regulation was direct through an inducible system (AP2m3-GR) as well as ChIP experiments. Therefore, AP2 is a direct, negative regulator of AG.

Our data also suggest that AP2 represses AG most effectively during early stages of flower development. Initially, when AP2m3-GR whole inflorescences (composed of both young and old flowers) were used in the induction experiments, no obvious changes in AG mRNA levels were seen. However, upon microdissection of the inflorescences after induction to retain only flowers of stages 7 and younger, we observed a 50% decrease in AG mRNA levels upon AP2 induction. It is feasible that AP2 only negatively regulates AG during early stages of flower development as it has been shown that a myriad of other genes, such as CURLY LEAF (CLF), LEUNIG, SEUSS and RABBIT EARS also negatively regulate AG (Goodrich et al., 1997; Liu et al., 1998; Franks et al., 2002; Krizek et al., 2006). It is possible that in the outer two whorls, AP2 establishes the initial repression of AG, and other
AP2 binds an AT-rich sequence

The ‘A’-site is highly conserved in Brassicaceae

The euAP2 lineage predates the divergence of gymnosperms and angiosperms, but the biological functions of AP2 and its orthologs differ amongst flowering plants (reviewed in Litt, 2007). In Arabidopsis, AP2 specifies petrihan identities and restricts C-function to the inner two whorls. However, characterized AP2 orthologs from Antirrhinum and petunia do not appear to share the role of AP2 in flower development (Maes et al., 2001; Keck et al., 2003). For example, LIP1 and LIP2, AP2 orthologs in Antirrhinum, promote sepal identities but do not control petal identity or restrict the expression of PLENA (C-class gene) (Keck et al., 2003). In fact, mutations with ectopic C function in the outer whorls in Antirrhinum and petunia map to a microRNA, miR169 (Cartolano et al., 2007). Interestingly, the petunia ortholog of LIP/AP2, PhAP2A, was able to rescue the ap2-1 mutant when expressed in Arabidopsis (Maes et al., 2001). The ability of the petunia AP2 protein to regulate AG in the Arabidopsis context suggests that the DNA-binding properties of the petunia AP2 are similar to those of Arabidopsis AP2 and implies that divergence in C-class regulatory sequences or in AP2-interacting proteins may be responsible for the divergence in the ability of AP2 to regulate C-class genes.

In this study, we show that AP2R2 recognizes an AT-rich motif in vitro and that two such motifs within the AG 2nd intron mediate the regulation of AG by AP2 in vivo. Given the AT richness of introns, this sequence motif is present multiple times in the introns of AG and AG orthologs from other species. The positions of the motifs relative to other transcription factor binding sites may influence the ability of AP2 to act upon them. We show that the A site recognized by AP2R2 in the AG 2nd intron is conserved both in sequence and in position in Brassicaceae. This implies that AP2-mediated regulation of C-class gene expression is conserved in this family. Although the motif is present in the 2nd introns of AG orthologs from non-Brassicaceae species, the overall large divergence in 2nd intron sequence precluded confident alignments to determine whether the positions of the motifs are conserved.

Acknowledgements

We are grateful to Drs Aman Husbands, Harley Smith and Gina Maduro for advice on SAAB assays and EMSAs. We thank Xiaofeng Cao from the Institute of Genetics and Developmental Biology, Chinese Academy of Sciences for the pcmCSG7 XF0510 M8P-LIC vector and Patricia Springer for sharing plasmids and equipment. We thank Lijuan Ji, Shengben Li, Rae Eden Yumul and Yuanjuan Zhao for comments on the manuscript.

Funding

The work was supported by grants from the National Institutes of Health (GM61146), the Howard Hughes Medical Institute and the Gordon and Betty Moore Foundation to X.C. T.D. was supported by a National Science Foundation Career grant (DGE0504249). Deposited in PMC for release after 6 months.

Competing interests statement

The authors declare no competing financial interests.

Supplementary material

Supplementary material available online at http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.077073/-/DC1

References


R Development Core Team (2010). R: A language and environment for statistical computing. Vienna, Austria.


