Downloaded from genesdev.cshlp.org on August 9, 2022 - Published by Cold Spring Harbor Laboratory Press

AGL1–AGL6, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes

Hong Ma,[1] Martin F. Yanofsky,[2] and Elliot M. Meyerowitz[3]

Division of Biology, California Institute of Technology, Pasadena, California 91125 USA

The predicted products of floral homeotic genes, *AGAMOUS* (*AG*) from *Arabidopsis thaliana* and *DEFICIENS A* (*DEF A*) from *Antirrhinum majus*, have been shown previously to share strong sequence similarity with transcription factors from humans (SRF) and yeast (MCM1). The conserved sequence between these proteins is localized within a domain known to be necessary for the DNA binding and for the dimerization of SRF. We have isolated six new genes from *A. thaliana*, *AGL1–AGL6*, which also have this conserved sequence motif. On the basis of the sequence comparison between the *AG* and *AGL* genes, they can be assigned to two subfamilies of a large gene family. RNA dot blot analysis indicates that five of these genes (*AGL1*, *AGL2*, *AGL4*, *AGL5*, and *AGL6*) are preferentially expressed in flowers. In addition, in situ RNA hybridization experiments with *AGL1* and *AGL2* show that their mRNAs are detected in some floral organs but not in others. Our results suggest that these genes may act to control many steps of *Arabidopsis* floral morphogenesis. In contrast, the *AGL3* gene is expressed in vegetative tissues as well as in flowers, suggesting that it functions in a broader range of tissues. We discuss possible roles of this gene family during the evolution of flowers.

[*Key Words*: Floral-specific genes; flower development; gene family; MADS box; in situ hybridization]

Received October 30, 1990; revised version accepted December 28, 1990.

Although flower development has been described in some detail, very little is known about the molecular machinery that controls cellular differentiation in developing flowers. In recent years, the small mustard *Arabidopsis thaliana* has been used increasingly for plant molecular and genetic studies (Meyerowitz 1987, 1989), and a number of *Arabidopsis* floral homeotic mutants have been characterized (Koornneef 1987; Pruitt et al. 1987; Bowman et al. 1988, 1989; Haughn and Somerville 1988; Komaki et al. 1988; Kunst et al. 1989; Meyerowitz et al. 1989). Phenotypes of several homeotic mutants indicate that they alter floral organ identities (Komaki et al. 1988; Bowman et al. 1989; Kunst et al. 1989). One of these homeotic genes is *AGAMOUS* (*AG*). Homozygous *ag* mutant plants produce double flowers (Bowman et al. 1989; Meyerowitz et al. 1989). In the *ag* mutant flower, while four sepals and four normal petals develop in the outer two whorls, as in the wild type, six additional petals occupy the wild-type positions of stamens. In addition, a new flower appears in the position occupied in wild type by the ovary. The pattern of 4 sepals surrounding 10 (4 + 6) petals repeats until the whole flower has ~70 organs (Bowman et al. 1989). The *AG* gene has been cloned recently (Yanofsky et al. 1990), and DNA sequence analysis indicates that it encodes a protein that shares striking similarity in its amino-terminal portion with the DNA-binding domains of transcription factors from humans (SRF; Norman et al. 1988) and yeast (MCM1; Passmore et al. 1988), suggesting that the AG protein is a transcription factor. Another yeast regulatory gene, *ARG80* (Dubois et al. 1987), also has the same type of sequence motif.

**Present addresses: [1]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724 USA; [2]Department of Biology C-016, University of California at San Diego, La Jolla, California 92093 USA.
[3]Corresponding author.**

Approximately a dozen genes have been defined genetically to be required for normal floral morphogenesis in *Arabidopsis* (Koornneef 1987). The complex process of flower development is likely to require many more regulatory proteins to coordinate the formation of floral organs at the proper time and location. In *Drosophila*, many of the regulatory proteins that control early developmental fate share a conserved domain for similar functions, e.g., DNA binding (Ingham 1988). By analogy, it is possible that the conserved putative DNA-binding domain of AG is shared by other regulators of flower development. In fact, a recently cloned flower homeotic gene from *Antirrhinum majus* (snapdragon), *DEF A*, also encodes a protein with the same type of DNA-binding domain (Sommer et al. 1990). This conserved motif has since been called the MADS box (for MCM1, AG and ARG80, DEF A, and SRF). Mutations in the *DEF A* gene cause phenotypes in snapdragon flowers that are very

Downloaded from genesdev.cshlp.org on August 9, 2022 - Published by Cold Spring Harbor Laboratory Press

different from those of *ag* mutants: The petals are replaced by sepals, and the stamens are replaced by carpel-like tissues, while the outer sepals and inner carpels are normal (Sommer et al. 1990).

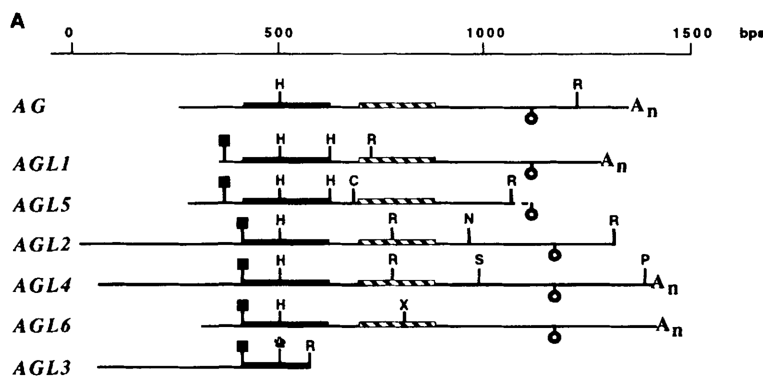
In an effort to gain further understanding of *Arabidopsis* flower development, we set out to isolate genes that share sequence similarity with *AG*. Here we report the isolation and characterization of six genes that share substantial sequence similarity with *AG* and *DEF A*. They are designated *AGL1–AGL6* for AG-like. We present the sequences of the *AGL* genes and their expression patterns. The possible functions of this large family of regulatory genes in flower development, and its possible role in the evolution of the flower, are discussed.

Results

Isolation of AGL *genomic and cDNA clones*

The region of amino acid sequence similarity between the AG protein (Yanofsky et al. 1990), the known transcription factors SRF (Norman et al. 1988) and MCM1 (Passmore et al. 1988), and the yeast regulatory protein ARG80 (Dubois et al. 1987) is localized within a 56-residue domain (the MADS box) in the amino-terminal region of these proteins. A highly conserved octapeptide, KKAYELSV, is found within the MADS box. A set of degenerate oligonucleotides was generated based on this octapeptide (see Materials and methods). Low-stringency hybridization of an *Arabidopsis* genomic DNA blot with this set of oligonucleotides as probes revealed ~20 bands (data not shown). These oligonucleotides were then used to probe a cosmid library (Yanofsky et al. 1990) made from *Arabidopsis* nuclear DNA, and 46 clones were isolated. Southern blot analysis showed that 12 of the clones hybridized to an *AG* cDNA clone under moderate stringency (data not shown). On the basis of patterns of restriction fragments and hybridization with the *AG* cDNA, we concluded that these 12 clones most likely represent four genes, named *AGL1–AGL4*. This was confirmed later by DNA sequencing (see below). Representative cosmids were chosen for further analysis.

AGL1 and *AGL2* genomic fragments were used to probe a λgt10-based cDNA library constructed from *Arabidopsis* floral poly(A)$^+$ RNA (Yanofsky et al. 1990). DNA sequence analysis (see below) revealed that among the cDNA clones isolated with an *AGL1* probe (probe 1, Fig. 1B) there were not only clones for *AGL1* but also for *AGL2* and for one additional gene, designated *AGL5* (Fig. 1A). Similarly, *AGL2* and *AGL4* cDNA clones (Fig. 1A) were isolated with an *AGL2* probe (probe 2, Fig. 1B). Because these clones hybridize to the *AG* gene at moderate stringencies, we probed the cDNA library with an *AG* cDNA at a moderate stringency (see Materials and methods). Only moderate to weak positives were analyzed; a total of 27 *AGL* clones, including *AGL3* and another



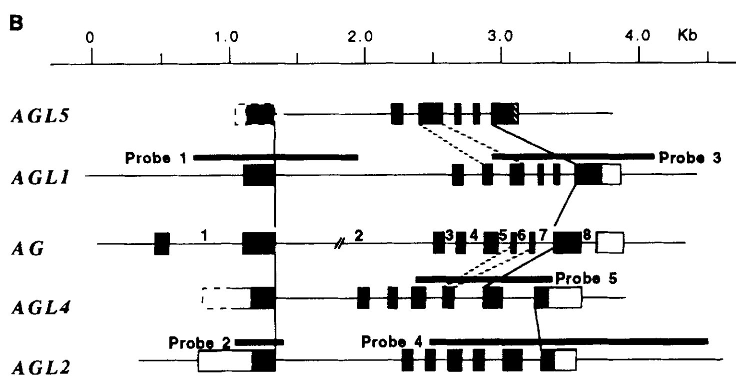


Figure 1. (*A*) Maps of *AGL* cDNAs. (*AG* and *AGL1*, *AGL2*, *AGL4*, and *AGL6*) Composite maps from two cloned fragments, the ends of which are marked by *Eco*RI sites (the cDNA and genomic sequences agree with each other; for clone numbers, see Materials and methods). All clones have *Eco*RI sites (all are not shown) at both ends; only *Eco*RI, *Hin*dIII, and other sites used for demarcation purposes are shown. The symbols (■) and (○) indicate the positions of translational initiation and termination codons, respectively. The A_n signs represent poly(A) tails. The solid bars indicate the position of the MADS boxes, and the hatched bars indicate the position of the conserved K boxes. The region at the carboxyl terminus of *AGL5*, represented by the dashed line, is from genomic sequence. The asterisk (*) in *AGL3* indicates the position of sequence AGGCTT, one base different from the *Hin*dIII site (AAGCTT) found at this position in the other cDNAs. Enzyme keys: (C) *Sca*I; (H) *Hin*dIII; (N) *Nde*I; (P) *Hpa*I; (R) *Eco*RI; (S) *Ssp*I; (X) *Xho*I. (*B*) *AGL1*, *AGL2*, *AGL4*, *AGL5* gene structures (for clone numbers, see Materials and methods). The boxes indicate exons (open boxes represent untranslated regions), and the lines between them represent introns. All of the introns have the canonical donor and acceptor sites, GT and AG, respectively. The boxes with dashed lines at the 5'-most portion of *AGL4* and *AGL5* represent regions of uncertainty because of the lack of genomic information; the hatched box at the 3' region of *AGL5* lacks cDNA confirmation. The bars above *AGL1*, *AGL2*, and *AGL4* represent regions of the corresponding genes used as probes to isolate cDNA clones. The dashed lines indicate the introns that are lacking in some genes; the other introns between the solid lines connecting different genes have conserved positions.

Downloaded from genesdev.cshlp.org on August 9, 2022 - Published by Cold Spring Harbor Laboratory Press

gene, *AGL6* (Fig. 1A), were isolated. *AGL5* genomic cosmids (probed with an *AGL5* cDNA) and additional *AGL2* and *AGL4* cosmids (probed with an *AGL2* cDNA) were isolated from the cosmid library. The 3′ portions of *AGL1*, *AGL2*, and *AGL4* cDNAs were isolated subsequently using gene-specific genomic fragments as probes.

The AGL *gene structures and nucleotide sequences*

We have determined the sequences of *AGL* cDNAs (Figs. 2–4; *AGL3* cDNA sequence is incomplete and not shown), as well as the entire genomic regions for *AGL1* and *AGL2* and most of the *AGL4* and *AGL5* genes. On the basis of the comparison between cDNA and genomic sequences, we have deduced the complete exon–intron structures for *AGL1* and *AGL2* and nearly complete structures for *AGL4* and *AGL5* (Fig. 1B). The intron positions are largely, though not entirely, conserved in all of the genes where the intron positions are known (Fig. 1B).

The *AGL1*, *AGL4*, and *AGL6* cDNAs each contain a large open reading frame (ORF), as well as 5′- and 3′-nontranslated regions and a poly(A) tail (Figs. 1A, 2–4). Although the cDNA clones for *AGL2* and *AGL5* do not include poly(A) tails (see Materials and methods), the entire protein-coding regions for these two genes have been identified (Figs. 1A, 2, and 3). The *AGL5* cDNA clone does not contain the termination codon for the longest ORF, but comparison of the genomic sequence matching the end of the *AGL5* cDNA with the carboxy-terminal sequences of *AG* and *AGL1* suggests the probable carboxyl terminus of the AGL5 protein (Fig.2). The sequence of the amino-terminal portion (including most of the MADS box) of the AGL3 protein (Fig. 5B) has been deduced from the cDNA sequence (data not shown). Additional AGL3 protein sequence (Fig. 5B) has been deduced from genomic sequence (data not shown) using the canonical intron donor (GT) and acceptor (AG) sites. The proteins encoded by the *AGL* genes are small (28.2–28.8 kD calculated molecular mass) and slightly basic, simi-

```
                             (1)GGATCA        6
       (5)GAATTCATCTTCCCATCCTCACTTCTCTTTCTTTC  35
(5)TGATCATAATTAATCTTGCTAAGCCAGCTAGGGCTTATAGAA  77
ATGGAGGAAGGTGGGAGTAGTCACGACGCAGAGAGTAGCAAGAAA    51
       GT    C     A G A TA        C        G    122
      E     G     S  H  D
M  E     Q     S           A  E  S  S  K  K      15
      G     A     N  E  V
CTAGGGAGAGGGAAAATAGAGATAAAGAGGATAGAGAACACAACA    96
A             G                          T  G    167
L
   G  R  G  K  I  E  I  K  R  I  E  N  T  T      30
I
AATCGTCAAGTTACTTTCTGCAAACGACGCAATGGTCTTCTCAAG    141
           C                        T A          212
N  R  Q  V  T  F  C  K  R  R  N  G  L  L  K      45
AAAGCTTATGAACTCTCTGTCTTGTGTGATGCCGAAGTTGCCCTC    186
           G                 C  T  G     T  T    257
K  A  Y  E  L  S  V  L  C  D  A  E  V  A  L      60
GTCATCTTCTCCACTCGTGGCCGTCTCTATGAGTACGCCAACAAC    231
                 A           C                   302
V  I  F  S  T  R  G  R  L  Y  E  Y  A  N  N      75
 #
AGTGTGAGGGGTACAATTGAAAGGTACAAGAAAGCTTGTTCCGAT    276
        A  A     A                    C     C    347
S  V  R  G  T  I  E  R  Y  K  K  A  C  S  D      90
                                      #
GCCGTCAACCCTCCTTCCGTCACCGAAGCTAATACTCAGTACTAT    321
     T        GA  A                              392
               S  V
A  V  N  P  P        T  E  A  N  T  Q  Y  Y      105
               T  I
CAGCAAGAAGCCTCTAAGCTTCGGAGGCAGATTCGAGATATTCAG    366
        G  G     A  C     A        G  C          437
Q  Q  E  A  S  K  L  R  R  Q  I  R  D  I  Q      120
          $
AATTCAAATAGGCATATTGTTGGGGAATCACTTGGTTCCTTGAAC    411
    TG  C  A  C   C    T     T                   482
   S              V
N     N  R  H  I     G  E  S  L  G  S  L  N      135
   L              L
TTCAAGGAACTCAAAAACCTAGAAGGACGTCTTGAAAAAGGAATC    456
  T           G     T   A TA G     G             527
                        G
F  K  E  L  K  N  L  E     R  L  E  K  G  I      150
                        S
                    #
AGCCGTGTCCGCTCCAAAAAGAATGAGCTGTTAGTGGCAGAGATA    401
  T        A     G   C C   A       T        T    572
                     N     L
S  R  V  R  S  K  K     E     L  V  A  E  I      165
                     H     M
                 #
GAGTATATGCAGAAGAGGGAAATGGAGTTGCAACACAATAACATG    546
  A  C     A  A        C   C     A  G            617
                     M           H  N
E  Y  M  Q  K  R  E     E  L  Q        N  M      180
                     I           N  D
              #
TACCTGCGAGCAAAGATAGCCGAAGGCGCCAGATTGAATCCGGAC    591
  T  C  CT C     TA T  ------   ACAGG  TAC  G    656
         A        A     G  A     L  N  P  D      195
Y  L  R     K  I     E        R
         S        T     -  -     T  G  L  Q      193
CAGCAGGAATCGAGTGTGATACAAGGGACGACAGTTTACGAATCC    636
  A  A                 TCAAGG            G  G    701
                     Q  G  T                     210
Q  Q  E  S  S  V  I           T  V  Y  E  S
                     H  Q  G                     208
GGTGTATCTTCTCATGACCAGTCGCAGCATTATAATCGGAACTAT    681
     TA     TC C        GG   G     C     T       746
      S     H  D        Q  H                     225
G  V     S        Q  S        Y  N  R  N  Y
      T     S  H        G  Q                     223
ATTCCGGTGAACCTTCTTGAACCGAATCAGCAATTCTCCGGCCAA    726
   G    T                     A T C   AA         791
   P                          Q  F     G         240
I     V  N  L  L  E  P  N  Q        S     Q
   A                          N  S     N         238
GACCAACCTCCTCTTCAACTTGTGTAA  753
        A     G        TG    818
                             248
D  Q  P  P  L  Q  L  V  End
                             246
(1)CTCAAAACATGATAACTTGTTTCTTCCCCTCATAACGATTAAGA 797
GAGAGACGAGAGAGTTCATTTTATATTTATAACGCGACTGTGTATTC 844
ATAGTTTAGGTTCTAATAATGATAATAACAAAACTGTTGTTTCTTTGCTTCAn
(5)TTCAGTCTAACATAAGCTTCTTTCCTCAGCCTGAGATCGATCTA 862
TAGTGTCACCTAAATGCGGCCGCGTCCCTCAACATCTAGTCGCAAGC 909
TGAGGGGAACCACTAGTGTCATACGAACCTCCAAGAGACGGTTACACAAA
```

Figure 2. *AGL1* and *AGL5* cDNA and deduced protein sequences. The complete coding region of *AGL1* is shown, and only the nucleotides in *AGL5* that are different from *AGL1* are shown below the *AGL1* sequence. The amino acid sequences are shown in boldface. The dashes in both nucleotide and amino acid sequences indicate gaps introduced to allow the best alignment. Where AGL1 and AGL5 amino acid residues are identical, they are shown once; where they are different, the AGL1 residues are shown above the AGL5 residues. For *AGL5*, the DNA sequence starting at nucleotide 782 and the last 12 amino acids residues are from genomic sequences. The calculated molecular masses for AGL1 and AGL5 are 28,337 and 28,158 daltons, respectively. The potential phosphorylation sites [RXX(T/S)] and glycosylation sites [NX(T/S)] are underlined. The positions of introns shared by *AGL1* and *AGL5* are represented by the number ( ) sign above the *AGL1* nucleotide sequence; the one intron that is present only in *AGL1* is indicated by the dollar ($) sign. The untranslated regions for *AGL1* and *AGL5* are shown separately, as indicated.

Downloaded from genesdev.cshlp.org on August 9, 2022 - Published by Cold Spring Harbor Laboratory Press


```
(2) CCCTCACACATTTCTTATCTTTTGCTCTCAATAGATTCCATTGATTCAAAACAAAATTTTCATTAAGATTTCACAACCTCCACACA   86
(4) ----------------------------------------GATTCA-CAAAAACTTTTC-TTCAGATT-CACAATCTCATCACAA   42
(2) --CTTCC----------------AAACACAATTAAAGAGAGGAAAAAGAATCAATAACCCTATAAATAAAAAATCAGACAAACAGA  154
(4) CCCTTCAAAAAGAGAAAAGATCTAAAGAATAAACAAGAGCCCTAATATCAAATCACAACCAAAAAAACCAAAGAAAG-CTAATTAA  127
(2) AGTTTCCTCTTCTTCTTCCTTAAGCTAGTACCTTTTGTTCTTGAAA-TTAGGGTTAATTTCTTTTTTCCAAATACCATCAATTCT  238
(4) AGTTTTCTCTCTAGCTATTCCTCTT---CTTTTCTTGTTCTTGAAAACTAGGGTTTACTT------------------------  184
(2) CCAGACCATAAAAACTCAAAAAGATCAGATCTTTCCTCTGAAAAAGAGATACCCAACTTATGTTTTTGTGTGTCTGTATATAG  321
(4) --------------CACCAAAAGATAAGATCTTTCCCCAGAAAAAGCAATACCCAAGTCATGTTTCTGTGTGTCTGTATATAG  253
(2) ATAAA-CATTACATACCCATATTTGTGTATAGACATAAAAAGTGGAAATTAAGGTAACAAAAAGAA------------------  386
(4) ATAAAACATTACATACCCTAATAAGGTTACACAAATAGCTATAAAAGAGGGAAAATAAGATAGGGATTTTTTGGGGTGAGGAAAG  338

ATGGGAAGAGGAAGAGTAGAGCTGAAGAGGATAGAGAACAAAATCAACAGACAAGTAACGTTTGCAAAGCGTAGGAACGGTTTGTTGAAG  476
                    C                             G        T  A     A  T       C    A  428
M  G  R  G  R  V  E  L  K  R  I  E  N  K  I  N  R  Q  V  T  F  A  K  R  R  N  G  L  L  K   30

AAAGCTTATGAATTGTCTGTTCTCTGTGATGCTGAAGTTGCTCTCATCATCTTCTCCAACCGTGGAAAGCTCTATGAGTTTTGCAGCTCC  566
          GC T              C          CT         G           C        C     C      A   518
K  A  Y  E  L  S  V  L  C  D  A  E  V  A  L  I  I  F  S  N  R  G  K  L  Y  E  F  C  S  S   60
                                       S     V                                           T
   #                                                                                     #
TCAAACATGCTCAAGACACTTGATCGGTACCAGAAATGCAGCTATGGATCCATTGAAGTCAACAACAAACCTGCCAAAGAACTTGAGAAC  656
 C               G  AA    T    G  T      C                              T    G            608
S  N  M  L  K  T  L  D  R  Y  Q  K  C  S  Y  G  S  I  E  V  N  N  K  P  A  K  E  L  E  N   90
                     E                                  #
AGCTACAGAGAATATCTGAAGCTTAAGGGTAGATATGAGAACCTTCAACGTCAACAGAGAAATCTTCTTGGGGAGGATTTAGGACCTTTG  746
        G  CT      G  A        A  T  G       G              A     C T        C           698
S  Y  R  E  Y  L  K  L  K  G  R  Y  E  N  L  Q  R  Q  Q  R  N  L  L  G  E  D  L  G  P  L  120
                                                                       #
AATTCAAAGGAGTTAGAGCAGCTTGAGCGTCAACTGGACGGCTCTCTCAAGCAAGTTCGGTCCATCAAGACACAGTACATGCTTGACCAG  836
           C                        A         G          C G              T                788
N  S  K  E  L  E  Q  L  E  R  Q  L  D  G  S  L  K  Q  V  R  S  I  K  T  Q  Y  M  L  D  Q  150
                                                            C  #
CTCTCGGATCTTCAAAATAAAGAGCAAATGTTGCTTGAAACCAATAGAGCTTTGGCAATGAAGCTGGATGATATGATTGGTGTGAGAAGT  926
   T         GG   G     T  C        TG     C          T            A       C  C       CA   878
L  S  D  L  Q  N  K  E  Q  M  L  L  E  T  N  R  A  L  A  M  K  L  D  D  M  I  G  V  R  S  180
               G           I        D  A              S           E                    H
CATCATATG---GGAGGATGGGAAGGCGGTGAA---CAGAATGTTACCTACGCGCATCATCAAGCTCAGTCTCAGGGACTATACCAGCCT 1010
 C    AGGA                T     TCAA     A  G    T GA    C   G     T                AT    968
H  H  M  -  G  G  W  E  G  G  E  -  Q  N  V  T  Y  A  H  H  Q  A  Q  S  Q  G  L  Y  Q  P  208
      I  G                    D  Q     I  A     G     P        H                       S  210
                             #
CTTGAATGCAATCCAACTCTGCAAATGGGGTATGATAATCCAGTATGCTCTGAGCAAATCACTGCGACAACACAAGCTCAGGCGCAGCCG 1100
      TG    C   T        T  A   AGCC      G     A        GG   T   GGTG   G     T C  A AA  1058
L  E  C  N  P  T  L  Q  M  G  Y  D  N  P  V  C  S  E  Q  I  T  A  T  T  Q  A  Q  A  Q  P  238
         D              I        S  H                    M  A  V     V     G     S     Q  240
GGAAACGGTTACATTCCAGGATGGATGCTCTGAGAATCATGTACTGTGATGAAGCTCACCCACAAAAGACCTTATATATATATAAAGTAT 1190
      C     C  T  C        G    GCGATACTTCTTCCCCCAATAAAGATCTTAAGCAAGTACTGGTGGGGTCTTCGTGGT 1148
G  N  G  Y  I  P  G  W  M  L  End 248
                                  250

(2)AGATACAAGACTTGGATTTGTAGACATAAGTGGCTAATATAATGGTCCTGAGGATCTTCTAGACATTTGTATCTTTTGGGAATCCTT 1277
   GCTTATATTAAGAATTC          1294

(4)GTGATCTTAGATCTTATGCATATGAATAATAATGTTATTGCACAAGACTTTTGCTTTTGTAGACACAAGTGGCTATAGCTGTAATAG 1235
   CCTTCAACATCTCTCTTCTGTTTCAGGATTTGTTTGTGCCTATTGTAATTGCTTATATATGTATGGTTTGTATAATGTGTGAAATGT 1322
                         $                                                   $
   TAACATCGACCATGTCTCATCTGGTGAn
                    $        $
```

Figure 3. *AGL2* and *AGL4* cDNA and deduced protein sequences. The coding regions are shown in the same way as in Fig. 2, with *AGL2* sequences above *AGL4* sequences. The dashes in both nucleotide and amino acid sequences indicate gaps. The calculated molecular masses for AGL2 and AGL4 are 28,456 and 28,579 daltons, respectively. The 5′-untranslated regions are shown with *AGL2* above *AGL4* sequences, as indicated with (2) and (4). The two small ORFs are highlighted in boldface, and the flanking conserved sequences are underlined. The potential phosphorylation sites and glycosylation sites of the predicted proteins are underlined. Intron positions for both *AGL2* and *AGL4* are denoted with number (#) signs. The 3′-untranslated region for the two cDNAs are shown separately. In the 3′-untranslated region of *AGL4*, the dollar ($) signs indicate the positions of four observed polyadenylation sites, and potential polyadenylation signals are underlined for both *AGL2* and *AGL4*.

lar to the AG protein (Yanofsky et al. 1990). Table 1 shows the percentages of identical residues between AG and the AGL proteins in different regions. Like AG, the AGL proteins all have the MADS box, as discussed below.

The regions in the *AGL2* and *AGL4* cDNAs 5′ of the long ORFs each have a short ORF beginning with an ATG codon (Fig. 3). These short ORFs differ by only one nucleotide and potentially encode the identical heptapeptide MFLCVCI. The sequences flanking these short

Downloaded from genesdev.cshlp.org on August 9, 2022 - Published by Cold Spring Harbor Laboratory Press


```
                                          ATTTATCGTGTAC    13
GATACTTTATTCCTTTTATCTATTCTTGAAAAAAAGTTACCAATT    58
CTTGAGAAGAAGAAGAAATCAGAATCAAGAGAAGGAGAGAGAAAG   103
ATGGGAAGAGGGAGAGTGGAGATGAAGAGGATAGAGAACAAGATT   148
M  G  R  G  R  V  E  M  K  R  I  E  N  K  I      15
AATAGACAAGTGACCTTCTCAAAAAGAAGAAACGGTTTGCTGAAG   193
N  R  Q  V  T  F  S  K  R  R  N  G  L  L  K      30
AAAGCTTATGAGCTTTCTGTTCTTTGCGATGCCGAAGTTGCTCTC   238
K  A  Y  E  L  S  V  L  C  D  A  E  V  A  L      45
ATCATCTTCTCAAGCCGTGGCAAGCTCTACGAGTTTGGTAGTGTT   283
I  I  F  S  S  R  G  K  L  Y  E  F  G  S  V      60
GGAATTGAAAGCACAATCGAACGGTATAATCGTTGTTACAACTGC   328
G  I  E  S  T  I  E  R  Y  N  R  C  Y  N  C      75
TCTCTAAGCAATAATAAGCCTGAAGAGACTACACAGAGTTGGTGT   373
S  L  S  N  N  K  P  E  E  T  T  Q  S  W  C      90
CAGGAGGTGACAAAGCTTAAATCCAAATACGAATCTCTTGTTCGT   418
Q  E  V  T  K  L  K  S  K  Y  E  S  L  V  R     105
ACTAACAGGAATTTGCTTGGAGAAGATCTTGGAGAAATGGGTGTG   463
T  N  R  N  L  L  G  E  D  L  G  E  M  G  V     120
AAGGAACTGCAAGCGCTCGAGAGGCAGCTCGAAGCCGCTCTTACC   508
K  E  L  Q  A  L  E  R  Q  L  E  A  A  L  T     135
GCGACTCGACAGCGCAAGACACAAGTTATGATGGAAGAAATGGAA   553
A  T  R  Q  R  K  T  Q  V  M  M  E  E  M  E     150
GACCTTAGGAAAAAGGAGAGGCAACTAGGAGACATAAACAAACAA   598
D  L  R  K  K  E  R  Q  L  G  D  I  N  K  Q     165
CTCAAGATTAAGTTTGAAACGGAAGGCCATGCTTTCAAAACCTTT   643
L  K  I  K  F  E  T  E  G  H  A  F  K  T  F     180
CAAGACTTATGGGCAAACTCGGCGGCATCGGTGGCCGGGGATCCA   688
Q  D  L  W  A  N  S  A  A  S  V  A  G  D  P     195
AACAATTCTGAATTTCCGGTAGAGCCTTCTCATCCTAATGTATTG   733
N  N  S  E  F  P  V  E  P  S  H  P  N  V  L     210
GATTGCAACACCGAACCCTTTTTACAAATAGGGTTTCAACAACAT   778
D  C  N  T  E  P  F  L  Q  I  G  F  Q  Q  H     225
TACTACGTGCAAGGTGAAGGGTCTTCGGTATCAAAGAGTAACGTG   823
Y  Y  V  Q  G  E  G  S  S  V  S  K  S  N  V     240
GCAGGTGAGACTAATTTCGTCCAAGGTTGGGTTCTTTGA   862
A  G  E  T  N  F  V  Q  G  W  V  L  End  252
CTCTCTGTTGATTAGCCCACGATGCCACGGTCAGGCCAATTTCAGC   908
TCTCTACAGTTGTTCTTTTTTCAAATTAGATTTCTGGTTTTTTTTT   954
TCCTATAAGAAAAACTTTTGCACTAGATGTTTGTCATTTAATTTCC  1000
AGCTCGTGTGAATCTATATTCGCATGTATGTGCTTTGAAGAATTTC  1046
                    $
TCCTCTTACTCCTACTTGATCTAAAACATTATTTTTGTTTTGGGTTTAn
                                                $
```

Figure 4. *AGL6* cDNA and deduced protein sequence. The calculated molecular mass for AGL6 is 28,744 daltons. The potential phosphorylation sites and glycosylation sites are underlined. The positions of two observed polyadenylation sites are denoted with a dollar ($) sign below the nucleotide.

ORFs are also highly conserved between *AGL2* and *AGL4* (36/45 and 17/18 identity for 5′ and 3′, respectively). However, the nucleotides immediately adjacent to the ATGs of these small ORFs do not match the plant initiation consensus sequence: (A/T)(C/A)AAC-AATGGC (Lütcke et al. 1987). The presence of short ORFs upstream of the protein-coding region has been observed previously for other genes in yeast (Hinnebusch 1984; Werner et al. 1985; Forsburg and Guarente 1989), in animals (Kozak 1987), and in plants (Ma et al. 1990; Schmidt et al. 1990). In yeast, it is known that the short ORFs in the *GCN4* and *CPA1* (Hinnebusch 1988) mRNA are required for proper translational regulation.

Map positions of AGL1, AGL2, *and* AGL3

As a step toward determining whether the *AGL*s correspond to any genes identified previously by genetic analysis, we localized *AGL1*, *AGL2*, and *AGL3* relative to other molecular markers using restriction fragment length polymorphisms (RFLPs; Chang et al. 1988). *AGL1* maps on chromosome 3 near the lower end, ~0.6 cM from the marker 460 on the RFLP map constructed by Chang et al. (1988); *AGL2* maps on chromosome 5 about 0.6 cM centromere-proximal from the chalcone synthase gene; and *AGL3* maps on chromosome 2 near the upper end, ~2.6 cM from the marker 246. From these mapping results, *AGL1–3* do not appear to coincide with any gene identified previously by mutations. The *AGL4*, *AGL5*, and *AGL6* genes did not reveal any RFLPs between the ecotypes used in our crosses, and have not been mapped.

AGL1, AGL2, AGL4, AGL5, *and* AGL6 *are expressed preferentially in flowers*

Because cDNA clones for all six *AGL* genes have been isolated from a cDNA library constructed from floral poly(A)$^+$ RNA, it follows that these genes are expressed in flowers. Their expression patterns were further characterized using RNA dot blot hybridizations. RNAs from immature seed pods, flowers, stems, and leaves were spotted onto nylon filters, and identical filters were probed with each of the labeled 3′ portions of *AGL1*, *AGL2*, *AGL4*, *AGL5*, and *AGL6* cDNAs (lacking the sequences encoding the MADS box to minimize cross-hybridization) and the only available *AGL3* cDNA (including the sequences encoding the MADS box). As a control, radiolabeled *AG* cDNA was also used to probe one of the RNA filters. The results (Fig. 6) agree with the previous finding (Yanofsky et al. 1990) that *AG* is expressed in flowers but not in leaves or stems. Five of the *AGL* genes (except *AGL3*) are expressed preferentially in flowers, and the expression continues, albeit diminished, in immature seed pods (Fig. 6). Faint signals were also detected with *AGL1* and *AGL6* in stems, and with *AGL2* in leaves. *AGL3* is expressed in stems and leaves, as well as in flowers and seed pods (Fig. 6). As controls for cross-hybridization, in vitro transcripts of the sense orientation from *AGL1*, *AGL2*, *AGL4*, *AGL5*, and *AGL6* cDNA were synthesized and spotted on the same filter strips. No cross-hybridization between any of the *AG* and the *AGL* probes and in vitro transcripts was observed under the conditions used (Fig. 6). The approximate levels of *AGL* expression are slightly lower than that of *AG*, which was estimated to have an average abundance of 1 in 10^4 poly(A)$^+$ RNA molecules in floral tissues (Yanofsky et al. 1990). This result agrees with the observed frequency of *AGL* cDNA clones in the cDNA library.

In situ RNA hybridizations with AGL1 *and* AGL2

To determine whether these genes are expressed in an organ-specific manner, the expression patterns of *AGL1* and *AGL2* were characterized in more detail by in situ hybridization. Wild-type *A. thaliana* (Landsberg *erecta*)

Downloaded from genesdev.cshlp.org on August 9, 2022 - Published by Cold Spring Harbor Laboratory Press

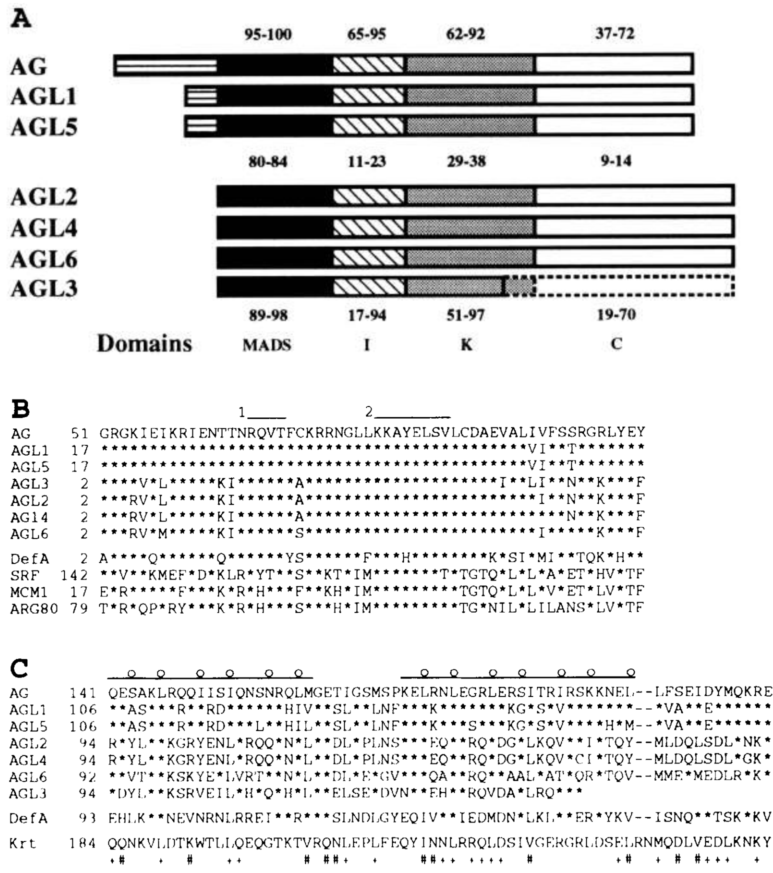


Figure 5. The comparison of deduced AG and AGL structures and alignment of conserved domains. The asterisks (*) represent identity to the first sequence of the same group; dashes indicate gaps introduced for alignment purposes. The alignment was done using the FASTP program (Lipman and Pearson 1985). (*A*) The comparison of different regions of AG and the AGL protein. The four regions (MADS, I, K, and C; see Table 1 for more detailed data on percent identities) are represented by differently shaded boxes. The numbers at the top are percent identities between AG, AGL1 and AGL5 (subfamily I); the numbers at the bottom are those between AGL2, AGL4, AGL6, and AGL3 (subfamily II); and the numbers in the middle are those between any member of subfamily I and any of subfamily II. The AGL3 information is partly derived from genomic sequence based on similarity to other AGLs and canonical intron donor and acceptor sites; the dashed-line boxes represent presumed unknown regions. The sequences in the MADS and K boxes are compared in *B* and *C*, respectively. (*B*) The alignment of the AG (Yanofsky et al. 1990) and AGL MADS boxes with those of DEF A (Sommer et al. 1990), SRF (Norman et al. 1988), MCM1 (Passmore et al. 1988) and ARG80 (Dubois et al. 1987) proteins. The conserved phosphorylation site (1) and the peptide (2) used to derive degenerate oligonucleotide sequences are indicate with lines above the AG sequence. (*C*) The alignment of the plant protein K boxes and region of the human type II keratin (Krt; see Tyner et al. 1985). Two possible helices are indicated by lines above the regions, and the circles (○) above the AG sequence indicate residues at positions a and d in the coiled coil heptapeptide repeat structure (Steinert and Roop 1988). The plus (+) signs below the keratin sequence represent an identity of the keratin residue to the corresponding residue in at least three of the AG, AGL, and DEF A sequences; the number (#) signs represent similar residues between keratin and at least four of the AG and AGL sequences.

inflorescence sections were hybridized with ^{35}S-labeled antisense RNA probes from *AGL1* and *AGL2* cDNAs. The 3′ portions of the cDNAs lacking the putative DNA-binding domain were used to avoid cross-hybridization. As shown in Figure 7, *AGL1* is expressed in carpels, particularly in ovules but not in stamens, petals, or sepals. *AGL2* is expressed mainly in carpels; in addition, the *AGL2* probe also detects a weak signal in stamens. Within the stamens, the *AGL2* signal is restricted to the anthers and is not observed in the filaments. Similar to *AGL1*, the *AGL2* signal in carpels is concentrated in ovules. The expression of *AGL1* and *AGL2*, as detected by in situ hybridization, begins in stage 10 flowers (Smyth et al. 1990) after all of the floral organs are recognizable and the ovules are visible. This onset of *AGL1* and *AGL2* expression is much later than that of *AG*, which is seen before the separation of petal and stamen primordia from the central floral primordium (G. Drews, J. Bowman, and E.M. Meyerowitz, unpubl.), in stage 3 flowers (Smyth et al. 1990). The expression of *AG*, *AGL1*, and *AGL2* all extend into later stages of flower development, including immature seed pods (Figs. 6 and 7). For *AGL1* and *AGL2*, the in situ signals are stronger in older organs (Fig. 7).

Discussion

AG *and the* AGLs *constitute a gene family*

We have identified and characterized six genes from *A. thaliana*, designated *AGL1–AGL6*. The deduced AGL proteins all share striking sequence similarity (Fig. 5) with each other and with the products of the floral homeotic genes *AG* and *DEF A*. Sequence analysis indicates that *AG* and *AGLs* are members of a diverse gene family. Table 1 shows the percentage of amino acid sequence identity between these deduced proteins in four regions. The most conserved region, the MADS box (M), is located either at or very near the amino terminus in the AGLs. The second conserved domain (the K box; see below), not found in SRF and MCM1, is near the center of the proteins, ~35 residues from the MADS box. On the basis of sequence comparison, *AG*, *AGL1*, and *AGL5* can be assigned to one subfamily, and *AGL2*, *AGL4*, and *AGL6* can be assigned to another subfamily. It is worth noting that the sequence similarity shared between members of the same subfamily is not restricted to the two conserved regions but extends throughout the entire length of the proteins (Fig. 5; Table 1). The subfamily assignment is also supported by the exon–intron struc-

Downloaded from genesdev.cshlp.org on August 9, 2022 - Published by Cold Spring Harbor Laboratory Press

Table 1. *Percentage of amino acid identity in different domains of AG, the AGLs, and DEF A proteins*

	AG		AGL1		AGL5		AGL2		AGL4		AGL3		AGL6		DEF A	
Genes	I	C	M	K	M	K	M	K	M	K	M	K	M	K	M	K
AG			**95**	68	**95**	62	**82**	32	**84**	30	**80**	37	**84**	38	**71**	24
AGL1	71	39			**100**	92	**82**	33	**80**	33	**82**	29	**82**	33	**68**	20
AGL5	65	37	95	72			**82**	36	**80**	35	**82**	33	**82**	33	**68**	23
AGL2	14	14	11	11	11	10			**98**	97	**95**	59	**95**	53	**62**	29
AGL4	17	14	14	10	14	9	94	70			**93**	57	**93**	55	**62**	29
AGL3	14	—	14	—	14	—	42	—	39	—			**89**	51	**61**	24
AGL6	23	10	23	12	23	11	34	19	29	19	17	—			**57**	23
	I	C	I	C	I	C	I	C	I	C	I	C				

The percentages for the MADS box (M, shown in boldface, 56 residues) and the second conserved domain (K, 66 residues except for AGL3, of which only 49 residues are known) are shown above the diagonal (blank space); the percentages for the sequences between the two conserved domains (I, 34–36 residues) and the carboxy-terminal regions (C, 78–98 residues) are shown below the diagonal. Because the AGL3 carboxy-terminal sequence is not known, the percentage of identity could not be calculated. See Fig. 5A for the domain organization of AG and the AGLs. The percentages of the two conserved domains (M and K) for *Antirrhinum* protein DEF A are also shown.

tures. An analysis of third-base silent changes in the MADS box indicates that the two most similar pairs (AGL1 and AGL5; AGL2 and AGL4) have much smaller percentages of difference (<30%) than other pairs, supporting the subfamily structure. Although the incomplete *AGL3* sequence does not allow definitive assignment of the *AGL3* gene, the partial sequence data indicate it is more similar to the *AGL2*, *AGL4*, and *AGL6* subfamily than the *AG*, *AGL1*, and *AGL5* subfamily. Figure 8 illustrates the relationship between AGLs and AG based on the sequence information. Sequence comparison of the *DEF A* gene to *AG* and the *AGL* genes suggests that *DEF A* does not belong to either of the two subfamilies (Table 1; Fig. 8). Low-stringency hybridization of *Arabidopsis* genomic DNA with degenerate oligonucleotides revealed ~20 bands, more than accounted for by the seven genes that have been isolated; therefore, it is likely that there are additional members of this gene family in *Arabidopsis*. We propose that members of this gene family are derived from a single ancestral gene and have arisen by gene duplication and subsequent sequence divergence and intron loss (Fig. 8).

AGLs *likely encode transcription factors*

The deduced AG and AGL amino acid sequences, as well as the product of the snapdragon floral homeotic gene *DEF A* (Sommer et al. 1990), contain a sequence motif of ~56 amino acids (the MADS box; see Fig. 5B) that is also found in the transcription factors SRF (Norman et al. 1988) and MCM1 (Passmore et al. 1988). A region of ~90 residues containing the MADS box is known to be sufficient for DNA binding and is involved in dimerization of SRF (Hayes et al. 1987; Norman et al. 1988). Recent evidence (Tan and Richmond 1990) suggests that the yeast MCM1 MADS box is also sufficient for specific DNA binding. The human SRF is involved in the regulation of the proto-oncogene c-*fos* (Treisman 1986, 1987) and a sarcomeric actin gene (Boxer et al. 1989). The yeast MCM1 gene product (GRM/PRTF) regulates mating type-specific gene expression (Herskowitz 1990). The phenotypes of *Arabidopsis ag* mutants (Bowman et al. 1989) and *Antirrhinum defA* mutants (Sommer et al. 1990) suggest that these genes play regulatory roles in specifying the identity of floral organs. The fact that the AGL proteins also contain the MADS box suggests that they are also transcription factors, possibly regulating floral morphogenesis. The AGL proteins may function in different floral cells, controlling branches of the floral morphogenesis regulatory hierarchy. Alternatively, they may function at different times, directing different stages of flower development. Although all of the AGL proteins are probably transcription factors, their expression patterns (see below) suggest they control different sets of genes.

In addition to the MADS box, the AG and AGL proteins share a second domain, which has a low but significant similarity to a portion of keratin sequences (the K

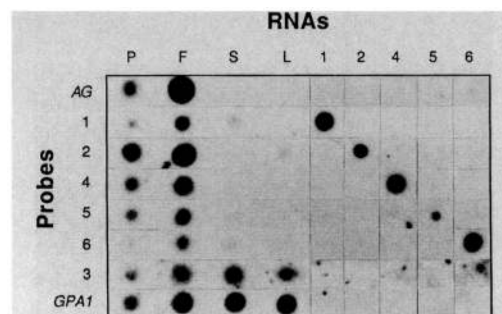


Figure 6. RNA dot blot autoradiogram. RNAs from immature seed pods (P), floral buds (F), stems (S), and leaves (L), as well as in vitro-synthesized RNAs from *AGL1* (*1*), *AGL2* (*2*), *AGL4* (*4*), *AGL5* (*5*), and *AGL6* (*6*), were spotted onto eight nylon filter strips. Each strip was hybridized with one ^{32}P-labeled cDNA: *AG*, *AGL1* (*1*), *AGL2* (*2*), *AGL4* (*4*), *AGL5* (*5*), *AGL6* (*6*), *AGL3* (*3*), and *GPA1*. *GPA1* is expressed in both floral and vegetative tissues (Ma et al. 1990). Similar amounts of radioactivity were used in all of the hybridizations. A single autoradiographic exposure was used.

Downloaded from genesdev.cshlp.org on August 9, 2022 - Published by Cold Spring Harbor Laboratory Press

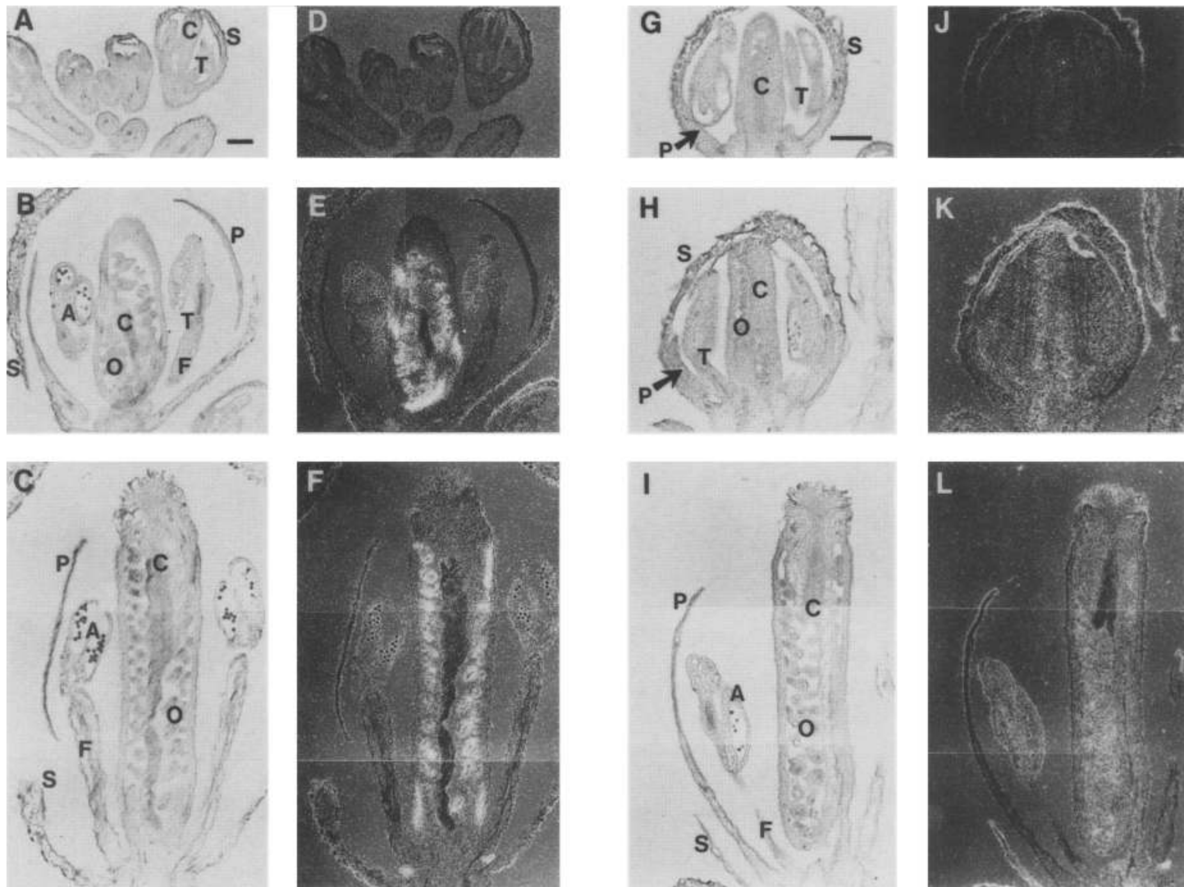


Figure 7. RNA in situ analysis with *AGL1* (*A–F*) and *AGL2* (*G–L*). *A–C* and *G–I* were photographed in bright field, and the others in dark field. For each gene, at least three developmental stages (for a description of stages 1–12, see Smyth et al. 1990) are shown: (*A* and *D*) stages 5–8; (*G* and *J*) stage 9; (*B* and *E*) stage 12; (*H* and *K*) stage 10; (*C, F, I,* and *L*) after flower opens but just before pollination; stage 13. Floral organ designations: (S) sepal; (P) petal; (T) stamen; (A) anther; (F) filament; (C) carpel; (O) ovule. The grains seen on the surface of some sepals and around the pollen grains are also seen when the sense-strand probes were used and, therefore, probably represent nonspecific sticking of the probes. Prints of one enlargement were used for *A–F, I,* and *L,* and a different enlargement for *G, H, J,* and *K.* (*A* and *G*) Bars, 100 μm.

box; Fig. 5C). Keratins are major components of intermediate filaments (Steinert and Roop 1988). It is known that the region of keratin with similarity to AG and AGL proteins is part of the coiled coil sequence that forms the central rod-shaped domain of keratin (Fig. 5C). The AGL sequences in this domain can potentially form two amphipathic helices.

Phosphorylation and glycosylation may modulate the activity of AGLs

SRF (Prywes et al. 1988; Ryan et al. 1989) is known to be phosphorylated. Furthermore, the phosphorylation of SRF has been shown to affect its activity (Prywes et al. 1988). Other transcription factors have been suggested to be regulated by phosphorylation, such as the yeast heat shock factor (Sorger and Pelham 1988) and GAL4 protein (Mylin et al. 1989). The difference between the apparent size (Tan and Richmond 1990) and sequence-derived size (Passmore et al. 1988) of the MCM1 protein suggests that it is also post-translationally modified. There is a conserved potential site (RQVT; Fig. 5C) for calmodulin-dependent phosphorylation [RXX(S/T); see Cohen 1988] in all of the AGLs, and most have additional sites as well (Figs. 2–4). Furthermore, it was reported that SRF is a glycoprotein (Schröter et al. 1990). Several other eukaryotic transcription factors are also glycosylated (Jackson and Tjian 1988; Lichtsteiner and Schibler 1989). The AGL proteins all have potential glycosylation sites (NXT or NXS; see Fishleigh et al. 1987; Figs. 2–4). The presence of these sites suggests that the activity of AGLs may be modulated by phosphorylation and/or glycosylation, perhaps in response to environmental and developmental signals.

Expression and functional implications of AGLs

Five of the *AGLs* are expressed preferentially in flowers and young seed pods but not (or at low levels) in leaves or stems. At this level, they are similar to known floral

Downloaded from genesdev.cshlp.org on August 9, 2022 - Published by Cold Spring Harbor Laboratory Press

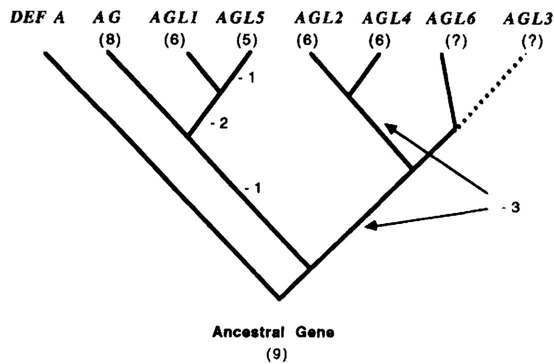


Figure 8. A proposed relationship between *AG*, *AGL* genes, and *DEF A*. The number of introns is indicated in parenthesis below each gene where it is known. The dashed line leading to *AGL3* represents the uncertainty of the position of *AGL3* due to the incompleteness of its sequence information. The time at which introns were lost during the evolution of *AGL2*–*AGL4*, and *AGL6* cannot be deduced because intron information is not available for *AGL3* and *AGL6*.

homeotic genes, *AG* from *A. thaliana* and *DEF A* from *A. majus*. The in situ hybridization results, on the other hand, indicate that the patterns of *AGL1* and *AGL2* expression within the flower are slightly different from that of *AG*. The *AG* expression begins in morphologically undifferentiated cells (G. Drews, J. Bowman, and E.M. Meyerowitz, unpubl.) in early developing flowers at stage 3 (Smyth et al. 1990). Later in development, *AG* is expressed in both carpels and stamens, including anthers and filaments (G. Drews, J. Bowman, and E.M. Meyerowitz, unpubl.). The onset of *AGL1* and *AGL2* expression is at stage 10, much later than that of *AG*. In addition, *AGL1* is expressed preferentially in carpels, not in stamens, petals, or sepals. The *AGL2* signal is found primarily in carpels and is lower in stamens. In carpels, the expression of both *AGL1* and *AGL2* is concentrated in ovules. Although *AGL2* and *AG* are both expressed in the stamens, *AGL2* mRNA is found only in the anthers, not in the filaments. These expression patterns suggest that *AGL1* and *AGL2* may regulate different genes from those that are controlled by *AG*. The fact that AGL1 and AGL2 share amino acid similarity with transcription factors, and that they are expressed in ovules suggest that they are both involved in regulating ovule development. The flower-specific expression of the *AG*, *AGL*, and *DEF A* genes suggests that the MADS box has been used repeatedly for flower development. On the other hand, because *AGL3* RNA is expressed in vegetative tissues as well as floral tissues, it is likely that this class of transcription factors is not exclusive to flower development.

Flowering plants appeared suddenly and very recently in evolution at ~140 million years ago (Lower Cretaceous). Some have suggested that floral organs are modified leaves, through a series of structural and functional changes during evolution (Stebbins 1976). In fact, the combination of mutations in three *Arabidopsis* homeotic genes, *AG*, *APETALA2*, and *APETALA3*, leads to "flowers" consisting of only leaf-like organs (J. Bowman, D.R. Smyth and E.M. Meyerowitz, in press). The presence of MADS-box containing genes in yeast, plants, and humans argues that this class of genes predates flowering plants. Therefore, it is likely that MADS-box regulatory gene or genes was present and functioning when flowers arose. Furthermore, during the evolution of flower structures, it is possible that these genes were duplicated, and some members diverged to take on new functions in floral morphogenesis, either interacting with different proteins, or binding to DNA with slightly different specificities. The family of genes described here includes at least one member (*AGL3*) that is expressed in both vegetative and floral tissues, presumably fulfilling a more widespread function. Additional members (*AG*, *AGL1*, *AGL2*, *AGL4*, *AGL5*, and *AGL6* and possibly others) of this gene family, presumably arisen by gene duplications, and have evolved to be preferentially expressed in flowers. At least two members of this gene family (*AG* from *Arabidopsis* and *DEF A* from *Antirrhinum*) are known to control flower development based on genetic analyses (Bowman et al. 1989; Sommer et al. 1990). Future analyses are required to test our hypotheses about the function of the other MADS box genes.

Materials and methods

Library screening, clones, and subclones

A genomic cosmid library (Yanofsky et al. 1990) made from nuclear DNA of *A. thaliana* (Landsberg *erecta*) was screened with radiolabeled degenerate oligonucleotides according to previously published procedures (Bürglin et al. 1989). The oligonucleotides 5'-ACNGANAGYTCUTANGCYTTYTT-3'(N = A, G, C or T; U = A or G; Y = C or T) are based on the conserved heptapeptide KKAYELSV in the MADS box (Yanofsky et al. 1990). More than 70 positives were detected among colonies of four genomes worth, and cosmid DNA of 46 of the positives were purified. Additional screening of the cosmid library with *AGL2* and *AGL5* cDNAs was done as described previously for hybridization of genomic DNA (Chang et al. 1988). Representative cosmids were characterized further: *AGL1*, pCIT1202 and pCIT1210; *AGL2*, pCIT1243; *AGL3*, pCIT1216; *AGL4*, pCIT1247 and pCIT4244; and *AGL5*, pCIT4243. *AGL6* cosmids have not yet been isolated. Portions of the cosmids pCIT1202, 1243, 4244, and 4243 are shown in Figure 1B.

About 1×10^6 plaques of a cDNA library constructed from floral bud poly(A)$^+$ RNA (Yanofsky et al. 1990) were screened at a moderate stringency (as described previously by Chang et al. 1988, except that the hybridization and washes were done with 5× SSPE at 52°C) with a 977-bp *AG* cDNA *Eco*RI fragment (pCIT565) as a probe. The cDNA library was also screened at high stringency (65°C hybridization and a final wash with 0.2× SSPE) with several probes (Fig. 1B): probe 1, a 1.1-kb *Dra*I *AGL1* genomic fragment (from pCIT1202); probe 2, a 0.36-kb *Bgl*II *AGL2* genomic fragment (from pCIT1243). During the construction of the cDNA library, cDNAs with internal *Eco*RI site(s) were cleaved and then ligated into separate vector molecules. The *AGL1*, *AGL2*, *AGL3*, *AGL4*, and *AGL5* cDNAs all contain at least one *Eco*RI site, therefore, each of the cDNA had to be isolated as two or more separate fragments. The 3' portions of *AGL3* and *AGL5* cDNAs have not yet been isolated. The portions encoding the amino terminus containing the conserved DNA-binding domain (Fig. 1A) were isolated first. The portions

Downloaded from genesdev.cshlp.org on August 9, 2022 - Published by Cold Spring Harbor Laboratory Press

of the cDNAs encoding the carboxy-terminal half for *AGL1*, *AGL2*, and *AGL4* (Fig. 1A) were isolated subsequently using the respective gene-specific genomic probes (probes 3–5, Fig. 1B): probe 3 (*AGL1*), a 0.41-kb *Eco*RI–*Bgl*II and 0.79-kb *Bgl*II fragments (from pCIT1202); probe 4 (*AGL2*), a 2.0-kb *Hind*III fragment (from pCIT1243); and probe 5 (*AGL4*), a 1.0-kb *Eco*RI–*Bgl*II fragment (from pCIT4244). The *AGL1*, *AGL2*, and *AGL4* genomic sequences were determined, and each agrees with the corresponding cDNA sequences on both sides of the *Eco*RI sites. Furthermore, additional *AGL6* cDNA clones were isolated using the first cDNA clone (pCIT3209) as a probe. The following cDNAs are shown in Figure 1A: *AGL1*, pCIT2241(5') and 4219 (3'); *AGL2*, pCIT3228 (5') and 4221 (3'); *AGL3*, pCIT2280 (5'); *AGL4*, pCIT3227 (5') and 4233 (3'); *AGL5*, pCIT2242 (5'); and *AGL6*, pCIT3209. In addition, pCIT3216 is identical to pCIT2242; pCIT2299 (lacking the MADS box) is a subclone of pCIT2242 containing a region 3' of the *Sca*I site (Fig. 1A); pCIT4210 (lacking the MADS box) contains a portion of an *AGL6* cDNA 3' of the *Xho*I site (Fig. 1A); and pCIT4214 is the same as pCIT4233 except at the 3' end where pCIT4214 lack 23 nucleotides and the poly(A) tail.

DNA sequencing and RFLP analysis

Genomic and cDNA fragments were subcloned into pGEM3Zf(+) and pGEM7Zf(+) vectors (Promega) for sequencing. Sequencing was done using the Sequenase Kit (U.S. Biochemical) according to the provided protocol. Both strands were sequenced, unless otherwise noted.

RFLP mapping was done as described by Chang et al. (1988). For *AGL1*, a 1.8-kb *Bgl*II fragment from cosmid clone pCIT1210 was found to reveal a *Bgl*II polymorphism between the Columbia and Niederzenz ecotypes of *Arabidopsis*, and it was used to probe filters carrying DNAs from one of the crosses used to generate a RFLP map (Chang et al. 1988). Similarly, a ~13-kb *Eco*RI fragment from *AGL2* cosmid pCIT1243 revealed an *Eco*RI polymorphism between Columbia and Niederzenz ecotypes and was used to probe filters from the same cross. Additional hybridization of filters with DNAs from a cross between Columbia and Niederzenz (*Xba*I polymorphism) and a cross between Landsberg and Niederzenz (*Bgl*II polymorphism; Chang et al. 1988) were performed using an *AGL2 Bgl*II–*Eco*RI fragment (subclone pCIT1273 from cosmid pCIT1243) as a probe. For *AGL3*, a subclone (pCIT1291)with a ~7-kb *Bgl*II fragment from cosmid pCIT1216 uncovered an *Xba*I polymorphism between Columbia and Niederzenz ecotypes and a *Bgl*II polymorphism between Landsberg and Niederzenz ecotypes and was used to probe appropriate filters. The data from DNA blot hybridization experiments were analyzed using the MAPMAKER computer program (Lander et al. 1987) to obtain linkage information with respect to existing markers on the RFLP map (Chang et al. 1988).

RNA analyses

Poly(A)$^+$ RNA was isolated from developing seed pods (3–5 days after pollination), floral buds (stages 1–12, Smyth et al. 1990), floral stems, and leaves, according to procedures described previously (Crawford et al. 1986). For RNA dot blot analysis, 15 μg of total RNA from each of the four tissues was spotted onto a nylon filter (Hybond N, Amersham), and hybridized with *AG* and *AGL* cDNAs labeled with ^{32}P using random priming methods. In addition, one filter was probed with labeled cDNA of the *GPA1* gene, which is expressed in stems and leaves, as well as in flowers (Ma et al. 1990). The following plasmids (see Fig. 1A; unless otherwise noted, the entire insert was used) were used for probe synthesis: pCIT565 (*AG*, with putative DNA binding domain), pCIT4219 (*AGL1*), pCIT4221 (*AGL2*, 3' *Nde*I–*Eco*RI fragment), pCIT4233 (*AGL4*, *Ssp*I–*Hpa*I fragment), pCIT2299 (*AGL5*), pCIT4210 (*AGL6*, 3' *Xho*I–*Eco*RI fragment), pCIT2280 (*AGL3*), and pCIT857 (*GPA1*; see Ma et al. 1990). To avoid cross-hybridization, the probes for *AGL1*, *AGL2*, *AGL4*–*AGL6* correspond to carboxy-terminal portion of the proteins (less conserved) and 3' nontranslated regions. Hybridizations were done as before (Yanofsky et al. 1990). 40 pg (about the amount of *AG* mRNA present in the total flower RNA) of in vitro synthesized RNA from *AGL* cDNAs (except *AGL3*) were also spotted on the same filters. The following cDNAs (see Fig. 1A) were used to synthesize RNA with the respective polymerase: pCIT4219 (*AGL1*, T7), pCIT4221 (*AGL2*, SP6), pCIT4214 (*AGL4*, T7), pCIT3216 (*AGL5*, SP6) and pCIT3209 (*AGL6*, T7). The plasmids were linearized so that only the inserts were used as templates for RNA synthesis.

In situ analysis was performed according to a previously described procedure (Barker et al. 1988; G. Drews, J. Bowman, and E.M. Meyerowitz, unpubl.). Inflorescences with young buds (stages 1–12; Bowman et al. 1989; Smyth et al. 1990) and flowers before pollination were fixed, embedded, and sectioned. The sections were hybridized with ^{35}S-labeled RNA probes complementary to *AGL* mRNAs. The plasmids pCIT4219 (SP6, *Xho*I) and pCIT4221 (T7, *Nde*I) (Fig. 1A) were used for *AGL1* and *AGL2* probes, respectively.

Acknowledgments

We thank S. Kempin for help with RFLP mapping, L. Medrano for excellent technical assistance, G. Drews and J. Bowman for advice on the in situ hybridization experiments, D. Spector and L. Fields for help with photography of in situ slides, A. Tagle for assistance in preparation of this manuscript, and J. Bowman, L. Brockman, C. Chang, G. Drews, B. Hamilton, T. Jack, L. Sieburth, V. Sundaresan, A. van der Bliek, and D.Weigel for valuable comments on this manuscript. This work was supported by grants from the National Science Foundation (DCB 8703439) and the Lucille P. Markey Charitable Trust. H.M. was supported by a postdoctoral fellowship from the Helen Hay Whitney Foundation, and M.F.Y. was supported by a postdoctoral fellowship in plant biology from the National Science Foundation (DCB 8608456).The sequence data described in this paper have been submitted to the GenBank database, and the accession numbers are *AGL1*, M55550; *AGL2*, M55551; *AGL4*, M55552; *AGL5*, M55553; and *AGL6*, M55554.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Barker, S.J., J.J. Harada, and R.B. Goldberg. 1988. Cellular localization of soybean storage protein mRNA in transformed tobacco seeds. *Proc. Natl. Acad. Sci.* **85:** 458–462.

Bowman, J.L., M.F. Yanofsky, and E.M. Meyerowitz. 1988. *Arabidopsis thaliana*, a review. *Oxf. Surv. Plant Mol. Cell Biol.* **5:** 57–87.

Bowman, J.L., D.R. Smyth, and E.M. Meyerowitz. 1989. Genes directing flower development in *Arabidopsis. Plant Cell* **1:** 37–52.

Boxer, L.M., R. Prywes, R.G. Roeder, and L. Kedes. 1989. The sarcomeric actin CArG-binding factor is indistinguishable

Downloaded from genesdev.cshlp.org on August 9, 2022 - Published by Cold Spring Harbor Laboratory Press

from the c-*fos* serum response factor. *Mol. Cell. Biol.* **9:** 515–522.

Bürglin, T.R., M. Finney, A. Coulson, and G. Ruvkun. 1989. *Caenorhabditis elegans* has scores of homeobox-containing genes. *Nature* **341:** 239–243.

Chang, C., J.L. Bowman, A.W. DeJohn, E.S. Lander, and E.M. Meyerowitz. 1988. Restriction fragment length polymorphism map for *Arabidopsis thaliana. Proc. Natl. Acad. Sci.* **85:** 6856–6860.

Cohen, P. 1988. Protein phosphorylation and hormone action. *Proc. R. Soc. Lond. Ser. B.* **234:** 115–144.

Crawford, N.M., W.H. Campbell, and R.W. Davis. 1986. Nitrate reductase from squash: cDNA cloning and nitrate regulation. *Proc. Natl. Acad. Sci.* **83:** 8073–8076.

Dubois, E., J. Bercy, and F. Messenguy. 1987. Characterization of two genes, *ARGRI* and *ARGRIII*, required for specific regulation of the arginine metabolism in yeast. *Mol. Gen. Genet.* **207:** 142–148.

Fishleigh, R.V., B. Robson, J. Garnier, and P.W. Finn. 1987. Studies on rationales for an expert system approach to the interpretation of protein sequence data: Preliminary analysis of the human epidermal growth factor receptor. *FEBS Lett.* **214:** 219–225.

Forsburg, S.L. and L. Guarente. 1989. Identification and characterization of HAP4: A third component of the CCAAT-bound HAP2/HAP3 heteromer. *Genes & Dev.* **3:** 1166–1178.

Hayes, T.E., A.M. Kitchen, and B.H. Cochran. 1987. Inducible binding of a factor to the *c-fos* regulatory region. *Proc. Natl. Acad. Sci.* **84:** 1272–1276.

Herskowitz, I. 1990. A regulatory hierarchy for cell specialization in yeast. *Nature* **342:** 749–757.

Hinnebusch, A.G. 1984. Evidence for translational regulation of the activator of general amino acid control in yeast. *Proc. Natl. Acad. Sci.* **81:** 6442–6446.

———. 1988. Novel mechanisms of translational control in *Saccharomyces cerevisiae. Trends Genet.* **4:** 169–174.

Ingham, P.W. 1988. The molecular genetics of embryonic pattern formation in *Drosophila. Nature* **335:** 25–34.

Jackson, S.P. and R. Tjian. 1988. O-Glycosylation of eukaryotic transcription factors: Implications for mechanisms of transcriptional regulation. *Cell* **55:** 125–133.

Komaki, M.K., K. Okada, E. Nishino, and Y. Shimura. 1988. Isolation and characterization of novel mutants of *Arabidopsis thaliana* defective in flower development. *Development* **104:** 195–203.

Koornneef, M. 1987. Linkage map of *Arabidopsis thaliana* (2n = 10). In *Genetic maps 1987: A compilation of linkage and restriction maps of genetically studied organisms* (ed. S.J. O'Brien), pp. 742–745. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15:** 8125–8148.

Kunst, L., J.E. Klenz, J. Martinez-Zapater, and G.W. Haughn. 1989. *AP2* gene determines the identity of perianth organs in flowers of *Arabidopsis thaliana. Plant Cell* **1:** 1131–1135.

Lander, E.R., P. Green, J. Abrahamson, A. Barlow, M. Daly, S.E. Lincoln, and L. Newburg. 1987. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1:** 174–181.

Lichtsteiner, S. and U. Schibler. 1989. A glycosylated liver-specific transcription factor stimulates transcription of albumin gene. *Cell* **57:** 1179–1181

Lipman, D.J. and W.R. Pearson 1985. Rapid and sensitive protein similarity searches. *Science* **227:** 1435–1441.

Lütcke, H.A., K.C. Chow, F.S. Mickel, K.A. Moss, H.F. Kern, and G.A. Scheele. 1987. Selection of AUG initiation codons differs in plants and animals. *EMBO J.* **6:** 43–48.

Ma, H., M.F. Yanofsky, and E.M. Meyerowitz. 1990. Molecular cloning of *GPA1*, a G protein α subunit gene from *Arabidopsis thaliana. Proc. Natl. Acad. Sci.* **87:** 3821–3825.

Meyerowitz, E.M. 1987. *Arabidopsis thaliana. Annu. Rev. Genet.* **21:** 93–111.

———. 1989. *Arabidopsis*, a useful weed. *Cell* **56:** 263–269.

Meyerowitz, E.M., D.R. Smyth, and J.L. Bowman. 1989. Abnormal flowers and pattern formation in floral development. *Development* **106:** 209–217.

Mylin, L.M., J.P. Bhat, and J.E. Hopper. 1989. Regulated phosphorylation and dephosphorylation of GAL4, a transcriptional activator. *Genes & Dev.* **3:** 1157–1165.

Norman, C., M. Runswick, R. Pollock, and R. Triesman. 1988. Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-*fos* serum response element. *Cell* **55:** 989–1003.

Passmore, S., G.T. Maine, R. Elble, C. Christ, and B.K. Tye. 1988. A *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of MATa cells. *J. Mol. Biol.* **204:** 593–606.

Pruitt, R.E., C. Chang, P.P. Pang, and E.M. Meyerowitz. 1987. Molecular genetics and development of *Arabidopsis*. In *45th Symposium of the Society for Developmental Biology* (ed. W. Loomis), pp. 327–338. A.R. Liss, New York.

Prywes, R., A. Dutta, J.A. Cromlish, and R.G. Roeder. 1988. *Proc. Natl. Acad. Sci.* **85:** 7206–7210.

Ryan, W.A., Jr., B.R. Franza, Jr., and M.Z. Gilman. 1989. Two distinct cellular phosphoproteins bind to the c-*fos* serum response element. *EMBO J.* **8:** 1785–1792.

Schmidt, R.J., F.A. Burr, M.J. Aukerman, and B. Burr. 1990. The maize regulatory gene *opaque-2* encodes a protein with a leucine zipper motif that binds to zein DNA. *Proc. Natl. Acad. Sci.* **87:** 46–50.

Schröter, H., C.G.F. Mueller, K. Meese, and A. Nordheim. 1990. Synergism in ternary complex formation between the dimeric glycoprotein $p67^{SRF}$, polypeptide $p62^{TCF}$ and the c-fos serum response element. *EMBO J.* **9:** 1123–1130.

Smyth, D.R., J.L. Bowman, and E.M. Meyerowitz. 1990. Early flower development in *Arabidopsis. Plant Cell* **2:** 755–767.

Sommer, H., J.-P. Beltran, P. Huijser, H. Pape, W.-E. Lonnig, H. Saedler, and Z. Schwarz-Sommer. 1990. *Deficiens*, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: The protein shows homology to transcription factors. *EMBO J.* **9:** 605–613.

Sorger, P.K. and H.R.B. Pelham. 1988. Yeast heat shock factor is an essential DNA-binding protein that exhibits temperature-dependent phosphorylation. *Cell* **54:** 855–864.

Stebbins, G.L. 1976. Seeds, seedlings, and the origin of Angiosperms. In *Origin and early evolution of angiosperms* (ed. C.B. Beck), pp. 300–311. Columbia University Press, New York.

Steinert, P.M. and D.R. Roop. 1988. Molecular and cellular biology of intermediate filaments. *Annu. Rev. Biochem.* **57:** 593–625.

Tan, S. and T.J. Richmond. 1990. DNA binding-induced conformational change of the yeast transcriptional activator PRTF. *Cell* **62:** 367–377.

Treisman, R.H. 1986. Identification of a protein binding site that mediates transcription response of the c-fos gene to serum factors. *Cell* **46:** 567–574.

———. 1987. Identification and purification of a polypeptide that binds the c-*fos* serum response element. *EMBO J.* **6:** 2711–2717.

Downloaded from genesdev.cshlp.org on August 9, 2022 - Published by Cold Spring Harbor Laboratory Press

Tyner, A.L, M.J. Eichman, and E. Fuchs. 1985. The sequence of a type II keratin gene expressed in human skin: Conservation of structure among all intermediate filament genes. *Proc. Natl. Acad. Sci.* **82:** 4683–4687.

Werner, M., A. Feller, and A. Pierard. 1985. Nucleotide sequence of yeast gene *CPA1* encoding the small subunit of arginine-pathway carbamoylphosphate synthetase: Homology of the deduced amino acid sequence to other glutamine aminotransferases. *Eur. J. Biochem.* **146:** 371–381.

Yanofsky, M.F., H. Ma, J.L. Bowman, G.N. Drews, K.A. Feldmann, and E.M. Meyerowitz. 1990. The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcriptional factors. *Nature* **346:** 35–39.

Downloaded from genesdev.cshlp.org on August 9, 2022 - Published by Cold Spring Harbor Laboratory Press

AGL1-AGL6, an Arabidopsis gene family with similarity to floral homeotic and transcription factor genes.

H Ma, M F Yanofsky and E M Meyerowitz

Genes Dev. 1991, **5:**
Access the most recent version at doi:10.1101/gad.5.3.484

References This article cites 45 articles, 18 of which can be accessed free at:
http://genesdev.cshlp.org/content/5/3/484.full.html#ref-list-1

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

Copyright © Cold Spring Harbor Laboratory Press