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. Meyerowitz³

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 \sim 70 organs (Bowman et al. 1989). The AG gene has been

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

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

different from those of *ag* mutants: The petals are replaced by sepals, and the stamens are replaced by carpellike 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 λ gt10-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 EcoRI sites (the cDNA and genomic sequences agree with each other; for clone numbers, see Materials and methods). All clones have EcoRI sites (all are not shown) at both ends; only EcoRI, HindIII, and other sites used for demarcation purposes are shown. The symbols (\blacksquare) and (\bigcirc) 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 HindIII site (AAGCTT) found at this position in the other cDNAs. Enzyme keys: (C) ScaI; (H) HindIII; (N) NdeI; (P) HpaI; (R) EcoRI; (S) SspI; (X) XhoI. (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.

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 carboxyterminal 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-

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.

												.,	,		0
			(5)(SAA1	TC7	ATCI	TTC	CAI	CC	FCAC	TTC	CTCI	TTC	CTTTC	35
(!	5) TC	SATC	ATA	ATT	CAA'	rct'i	rgci		CC/	AGCI	AGC	GCI	"I'A'	AGAA	
A	rggi	AGGA	LAGO	TGO	GAC	STAC	STC.	ACGA	ICG(CAGA	GAC	TAG	SCA	GAAA	122
	-	B	, I	ര്	· _	ิธ์	I B	` ם '		-		<u> </u>		- 3	122
	R	a	G		8	N	R	v	•	8	8	8	ĸ	ĸ	13
C	FAGO	GAG	AGG	GAA	AAT	raga	GAT		GA	GAT	AGA	GAA	CAC		96
A					G									T G	167
L	a	R	G	ĸ	I	B	I	K	R	I	B	N.	. 1		30
I															
A	ATCO	STCA	AGI	TAC	CTTT	FCT C	GCAA	ACG	ACC	GCAA	TGO	STCI	TC	rcaag	141
	_			с_	_	-	_	-	_		~	Т	Α.	-	212
N	<u>R</u>	0	¥			C	K	ĸ	R	N	6	ы 	يا مص	N.	• •
A	AAGO	CTTA	TGA	ACI	CTC	CTGT	"CT"	GTO	TG	ATGC	CGA T	AGI	"IGC	CCTC	186
ĸ	λ	¥	B	Ъ.	g	v	L	с	D	_م ۲	B	v	λ	Ŀ	60
G	ICA.	CTT	CTC		TCO	TGG		TCT	CT	TGA	GTA	CGC	CAP	CAAC	231
						A				С					302
v	I	P	8	T	R	G	R	L	Y	B	Y	A	N_	<u> </u>	75
	#								~						276
AC	STGI	GAG	iGGG A	TAC a	:'AA']	TGA A		GTA	CAA	AGAA	AGC	TTG	rrc C	CGAT	2/6
s	v	R	ົດ	_	I	Ē	R	Y	ĸ	ĸ	λ	с	្ន	ס	90
_													*		
GC	CCGI	CAA	CCC	TCC	TTC	CGT	CAC	CGA	AGO	TAA	TAC	TCA	GTA	CTAT	321
		т			GA	A									392
X	v	N	P	P	-		T	B	λ	N	T	Q	Y	¥	105
<u> </u>	CCA	202	ACC	'CTC	ד ממדי	L CCT	יידירכ	CAC		CAT	TCC	200	тат	TCAC	366
Cr	GCA	MGM	G	G	.174	A A	C	GAG	A D	GAI	rce	G	C	ICAG	437
Q	Q	B	Ň	័ន	ĸ	Ľ	R	R	 Q	I	R	ֿם	ī	Q	120
				\$											
AA	ATTC		TAG	GCA	TAT	TGT	'TGG	GGA	ATC	CACT m	TGG	TTC	CTI	GAAC	411
	ទ់		<u> </u>	×	<u>ر</u>	v	-	·	-	· -	•	-			402
N	т.	N	R	н	I	τ.	G	R	8	Ŀ	G	8	L	N	132
тл		GGA	АСТ	CAA	AAA	- CCT	'AGA	AGG	ACG	TCT	TGA	ААА	AGG	AATC	456
	т				G		Т	А	ТΑ	G		G			527
F	т ж	B	L	ĸ	G N	L	T E	A G	TA R	G L	E	G K	G	I	527 150
P	T K	B	L	ĸ	G N	L	Т Е #	А С 9	TA R	G L	B	G K	G	I	527 150
F AC	T K SCCG	e Tgt	L CCG	K CTC	G N CAA	L AAA	T E .GAA	A G S TGA	TA R GCI	G L GTT	e Agt	G K GGGC	G Aga	I .GATA	527 150 401
F Ac	T K SCCG T	e Tgt	r CCG	K CTC A	G N CAA	L LAAA G	T E .GAA C	A G S TGA C	TA R GCI A	G L IGTT	e Agt	G K GGGC T	g Aga	I .GATA T	527 150 401 572
F AC	T K SCCG T R	e tgt v	L CCG	K CTC A	G N CAA K	L IAAA G K	T E .GAA C N	A G S TGA C B	TA R GCT A L	G L CGTT L	B AGT V	G K TGGC T A	G Aga B	I GATA T I	527 150 401 572 165
F Ac S	T K SCCG T R	e Tgt V	L CCG	K CTC A <u> </u>	G N CAA K	L IAAA G K	T B GAA C N H	A G S TGA C E	TA R GCI A L M	G L CGTT. L	E AGT V	G K GGCC T A	G 'Aga B	I GATA T I	527 150 401 572 165
F AC S GP	T K SCCG T R AGTA	B TGT V	L CCG R R GCA	K CTC A GAA	G N CAA K K	L AAAA G K # GGGA	T E GAA C N H AAT	A G J TGA C B GGA	TA R GCT A L M GTT	G L CGTT L CGCA	B AGT V ACA	G K GGCC T A .CAA	G Aga B Taa	I GATA T I CATG	527 150 401 572 165 546
F Ac S GP	T K GCCG T R A GTA A	B TGT V TAT C	L CCG R GCA	K CTC A <u>-</u> S A A	G N CAA X GAG A	L G K GGGA	T E GAA C N H AAT	A G TGA C B C GGA C	TA R GCI A L GTI C	G L CGTT L CGCA	B AGT V ACA A	G K GGC T A .CAA G	G Aga B Taa	I GATA T I SCATG	527 150 401 572 165 546 617
F AC S GA B	T K GCCG T R A GTA A Y	B TGT V TAT C	L CCG R GCA	K CTC A GAA A K	G N CAA K GAG A R	L G K # GGGA B	Т В С С М Н ААТ М	A G S TGA C E C GGA C E	TA R GCI L M GTI C L	G L CGTT L CGCA	E AGT V ACA H	G K GGC T A CAA S N	G Aga B Taa N	I GATA T I CATG	527 150 401 572 165 546 617 180
F AC S GA B	T K GCCG T R A GTA A Y	B TGT V TAT C M	L CCG R GCA	K CTC A GAA GAA K	G N CAA K GAG A R	L AAAA G K # GGGA B	T B GAA C N H AAT M I	A G TGA C E C GGA C E	TA R GCI A L GTI C L	G L GTT L GCA Q	E AGI V ACA H N	G K GGGC T A CAA G N D	G AGA B TAA N	I GATA T I CATG	527 150 401 572 165 546 617 180
F AC S GP E TP	T R GCCG T R A GTA A T		L CCG R GCA Q AGC	K CTC A GAA GAA K K	G N CAA X GAG A R GAI	L G K # GGGA E TAGC	T E GAA C N H AAT M I CGA	A G TGA C E C GGA C E AGG	TA R GCT A L GTT C L CGC	G L CGTT L CGCA Q	E AGT V ACA A H N ATT	G K GGGC T A CAA G N D GAA	G AGA B TAA N TCC	I GATA T I CATG M CGGAC	527 150 401 572 165 546 617 180 591 656
F AC g G F T F	T R GCCG T R A GTA A T X CCT		L CCCG R GCA Q AGCC CT A	K CTC A GAA GAA K C	G N CAA K GAG A R # JGAI	L AAAA G K # GGGA B TAGC TA A	T B GAA C N H AAT L L CCGA T	A G TGA C B C GGA C E AGG G	TA R GCT A L GTT C C C C C C C C C C C C C C	G L CGTT. L CGCA Q CCAG	E AGT V ACA A H N ATT AC L	G K T A CAA G N D GAA CAAG N	G AGA B TAA N TCC TCC	I GATA T I CATG M CGGAC CGGAC D	527 150 401 572 165 546 617 180 591 656 195
F AC g G F E T F Y	T K GCCCG T R A GTA A CCT T L	E TGT TAT C M GCG C R	L CCCG GCA GCA Q AGCC CT A S	K CTC A GAA A K C AAAA C K	G N CCAA K GAG R I GAT	L AAAA G K H GGGA B CAGC TA A T	T B GAA C N H AAT M I CCGA T B	A G TGA C B C GGA C E AGG G	TA R GCT L GTT C L CGC	G L CGTT L CGCA Q CCAG	E AGT V ACA A H N ATT AC L	G K GGCC T A CAAA G N D GAAA CAGG N G	G AGA E TAAA N TCC TCC P I	I GATA T I CATG M GGGAC CAC G D O	527 150 401 572 165 546 617 180 591 656 195 193
P AC S GP B TP TP	T K GCCCG T AGTA A Y ACCT T L	B TGT TAT C M GCG C R	L CCCG GCA GCA Q AGCC CT A S ATC	K CTCC A GAAA K CAAAA C K	G N CCAA K GAG A R SGAT I I TGT	L AAAA G K B GGGA E CAGC TA A T CGAT	T B GGAA C N H AAAT M I CCGA T B CCGA	A G TGA C C B C GGA C C B AGG AGG	TA R GCT A L M GTT C L CGC A 	G L CGTT L CGCA Q CCAG	B AGT V ACA H N ATI AC L T AGT	G K GGGC T A CAAA G N D GAAA CAGG N G TTTA	G AGA B TAA N TCC T P L CGA	I GATA T I CATG M CGGAC G AC G D Q ATCC	527 150 401 572 165 546 617 180 591 656 195 193 636
F AC B GP E TP TP CP	T K SCCCG T A GTA A T L AGCA A	B TGT TAT C TAT C M GGCG C R GGGA	L CCCG R GCA Q AGCA CT A S ATC	K GAA GAAA K C K GAG	G N CCAA K GAG A R # GAT I I TTGT	L AAAA G K GGGA B CAGCTA A T CGAT	T B GAA C N H AAAT I I CCGA T B ZACA	A G TGA C GGGA C B AGGG C C LAGG TCA	TA R GCT A L GTT C L CGC A 	G L DGTT L DGCA Q CCAG R R CGAC	B AGT V ACA A H N ATT ACT L T AGT	G K GGGC T A CAAA G N D GGAAA SAGG N G TTTA	G AGA B TAA N TCC T P L CGA	I GATA T I CATG M CGGAC CATG Q ATCC G G G	527 150 401 572 165 546 617 180 591 656 195 193 636 701
F AC S G A G A B T A C A C A C A C A C A C A C A C S S A C S S A C S S A C S S A C S S S A C S S S S	T K SCCCG T R A GTA A T L A GCA A O	B TGT TAT C M CGCG C R SGGA A B	L CCCG GCA GCA Q AGCC CT A S ATC	K CTC A GAAA K CAAAA C K GGAG	G N CCAA K GGAG A R SGAT I TGT	L AAAA G K GGGA B CAGC TA A T CGAT	T B GAA C N H AAAT K I CCGA T B CCGA C A C A C A C A C A C A C N A C N H A AAT C N H A AAT C N H A A A C N H A A A C N A A A C N A A A A C N N A A A A	A G TGA C E GGA C E AGG TCA G	TA R GCT A L GTT C L CGC A GAC T	G L DGTT L DGCA Q CCAG R CGAC	B AGT V ACA A H N ATT AC L T AGT	G K GGCC T A CAAA G N D CAAA CAGG N G TTTA	G AGA B TAA N TCC T P L CGA	I GATA T I CATG M GGGAC D C G G G G G G G G G G G G G G	527 150 401 572 165 546 617 180 591 656 193 636 701 210
F AC G G F T F C F C F Q	T R GCCG T A GTA A Y A GCA L A GCA A Q	B TGT TAT C M GCG C R GGGA A B	L CCCG GCA GCA Q AGCC CT A S ATC S	K CTC A GAAA K CAAA C K GAGG S	G N CCAA K GGAG A R SGAT I TGT V	L AAAA G K B GGGA E CAGC TA A T CGAT I	T B GAAA C N H AAAT M I CCGA T B YACA Q H	A G TGA C C GGA C C E AGG TCA G Q	TA R GCI A L M GTI C L CGC A GAC AGC T G	G L NGTT L NGCA Q CCAG R CGAC	B AGT V ACA A H N ACT L C L T AGT	G K GGGC T A CAAA G N D GGAAA S AGG N G TTTA Y	G AGA B TAAA N TCC T P L CGA B	I GATA T I CATG M GGGAC CATG D Q ATCC G G 8	527 150 401 572 165 546 617 180 591 656 195 193 636 701 210 208
F AC B GP E TP TP CP Q GC	T R GCCG T R AGTA A Y AGCA A CCT T L AGCA A Q GTGT	B TGT TAT C M GCG C R GCG R GGA A B TATC	L CCCG GCA GCA Q CCC A S ACC S TTC	K CTC A GAAA K C K CAAA C K C GAG S TCA	G N CCAA K GGAG A R # GGAT I TGT V TGA	L AAAA G K F GGGA E CAGC TA A T CGAT I I	T B GAAA C N H AAAT I CCGA T B CCGA C C ACA Q H GCTC	A G S TGA C B C G G A A G G C A A G G C A C G C A C C B C C C B C C C B C C C B C C C C	TA R GCI A L M GTI C C C C C C C C C C C C C C C C C C C	G L CGTT L CGCA Q CCAG R CGAC T TTTA	B AGT V ACA A H N ATT AC L T AGT V TAA	G K GGGC T A CAAA G N D GGAA A G G TTTA Y TCG	G AGA B TAA N TCC I P L CGA B GAA	I GATA T I CATG CATG CATG CATG G G G G G G G G G G G G G G G G G G	527 150 401 572 165 546 617 180 591 656 195 193 636 701 208 681
F AC B GP B TP TP CP Q GC	T K GCCG T A GTA A A CCT T L A GCA A Q GTGI	B TGT TAT C TAT C C C R GGGA A B CATC TA	L CCCG GCA GCA Q AGCC CT A S ATC 8 TTC	K CTC A GAA K C K CAA C K C GAG S S TCA	G N CCAA K GAG A R * GAG I TGI V TGA C	L AAAA G K GGGA E CAGCC TA A T CGAT I	T B GAAA C N H AAAT I CCGA T B VACA Q H GTC	A G TGA C GGGA C E AGGG TCA G C C C C C C C C C C C C C C C C C C	TA R GCT A L GTT C L CGC 	G L DGTT L DGCA Q CCAG R CCAG R CGAC T TTTA	B AGT V ACA A H N ATT AC L T AGT V TAA	G K GGCC T A CAAA G M D GGAA S AGG N G TTTA T TCG C	G AGA B TAA N TCCC T P L CGA B GAA	I GATA T I CATG CATG M CGGAC CAC G D Q LATCC G G B LATCC T	527 150 401 572 165 546 617 180 591 656 193 636 701 210 208 681 746
F AC B GP B TP TP CP CP CP C C C G C G C G C	T K GCCG T A GTA A A CCT T L A GCCA T L A GCA Q GTGI V	B TGT TAT C M GCG C R A GGGA A B SATC TA B	L CCCG GCA GCA Q AGCC CT A S ATC S TTC S	K CTC A GAA K C K C GAG S C K C GAG S TCA H	G N CCAA K GGAG A R * GAT I TGT V TGA C D	L AAAA G K GGGA B CAGCC TA A T CGAT I L CCCA	T B GAA C N H AAAT M I CCGA T B CACA Q H GTCC S	A G TGA C B GGGA C B AGG C C B AGG TCA G C C C C C C C C C C C C C C C C C C	TA R GCI A L GTI C L CGC C GCC T G GCC H	G L NGTT L NGCA Q CCAG CCAG R CCAG T T T T T T	B AGI V ACA A H N ATI ACI L T AGI V TAA	G K GGCC T A CAAA G N GAAA G G TTTA T C C R	G AGA B TAA N TCC T P L CGA B GAA	I GATA T I CATG M GGGAC CAC G D Q ATCC G G S S CTAT T Y	527 150 401 572 165 546 617 180 591 656 195 193 636 701 210 8681 746 225
F AC B GP B T7 T7 C F C C F Q GC G C G	T K GCCG T R AGTA A Y ACCT T L AGCA A Q GTGI V	B TGT TAT C M GCGG C R GGGA A B SATC TA S T	L CCCG GCA GCA Q AGCC CT A S ATC S TTC S	K CTC A GAA A K C K GAG S C K GAG S C TCA H S	G N CCAA K GGAG A R # GAT I TGT I TGT V TGA C D H	L AAAA G K GGGA E CAGCC TA A CGAT I ACCCA	T B GAA C N H AAAT M I CCGA T B CCGA T B CCGA T B CACA Q H GTCC S	A G TGA C B GGA C E AGG TCA G G C C C C C C C C C C C C C C C C C	TA R GCI A L M GTI C L CGC 	G L DGTT L DGCA Q CCAG CCAG CCAG T R CGAC T T T T	B AGT V ACA A H N ATT AC L T AGT V TAA	G K GGCC T A CAAA G M G G G TTA G TTCG C R	G AGA B TAA N TCC P L CGA B GAA	I GATA T I CATG M CGGAC CAC G D Q ATCC G G S S CTAT T Y	527 150 401 572 165 546 617 180 591 636 701 210 681 746 225 223
F AC B GP E T7 T C Z GC G G C AD	T K GCCG T K GTGTA A A CCTT L A GCA A Q GTGT V TTCC	B TGT TAT C M GCG C R A GGA A B SATC TA B T CGGT	L CCCG GCA GCA Q AGCC CT A S ATC S TTC S	K CTCA A GAAA K C K GAGA S C TCA H S CCT	G N CCAA GAG A R GAT I TGI V TGA C D H TTCT	L AAAA G X GGAA T CGAT I ACCA Q	T B GAA C N H AATT T CCGA T CCGA T CCGA T C CCGA T C CCGA T C CCGA T C CCGA T C CCGA T C CCGA T C C N N S C N N S C N N S C N S A A S S A S S S S S S S S S S S S S	A G TGA C B GGA C B C C B C C C C C C C C C C C C C C	TA R GCI L GTI C L CGC 	G L NGTT L NGCA Q CCAG CCAG R CCAG T R G T T T T A GCA	B AGT V ACA A H N ATT AGT T AGT V TAA N ATT	G K GGGC T A CAAA G N D GAAA AGG N G TTTA G C T T C C R	G AGA E TAAA N TCCC P L CGA E GAA N	I GATA T I CATG S GGAC G G G G C CATCC G G G S CTAT T Y CCAAA	527 150 401 572 165 546 617 180 591 636 193 636 701 210 2081 746 223 726
F AC G G F T T C Z G C C Z G C G C G C G C G C G C G C	T K GCCG T A GTAA Y A CCTT L A GCAA Q GTGI V TTCCC G	B GTGT TATTC C C C C C C C C C C C C C C C C	L CCCG CCA Q CCA CCA S ACC CCA S CCA S CCCA CCA S CCCCA CCA	K CTCA A GAAA K CAAAA C K GAGG GAG S TCA TCA TCA TCA S CCT	G N CCAA GAG A R GAT I TGI V TGA V TGA V TGA	L AAAA G K GGAA E CAGCA T T CGAT I ACCA Q	T B GAA C N H AAT T B CCGA T B CCGA T B CCGA T B CCGA C C ACC C S ACCC	A G GGAA C E C G G AGGG C C E AGGG C C C C C C C C C C C C C C C C C	TA R GCT L GTT C C GCC GCC GCC GCC T GCC H Q T C T C	G L CGTT L CGCA Q CCAG CCAG C CAG C CAG C T T T T A GCA	B AGT V ACA A H N ATT L T T AGT V TAA N ATT T_C	G K GGGC T A CAAA G N D GAAA A G C TTAA TCG C R CCTCC	G AGA B TAA N TCC T P L CGA B GAA N SCGG	I GATA T I CATG CATG CATG C CATCC G G G C A T T Y SCCAA	527 150 401 572 165 546 617 180 591 656 195 193 6301 210 208 681 746 225 223 726 791
F AC G G F T F C F C F Q G C G C G C G C G C G C G C G C I I	T K GCCG T A GTA A Y CCCT T L A GCA A Q GTGT V TTCCC P	B GTGT TATTC C C C C C C C C C C C C C C C C	L CCCG CCA CCA CCA CCA S ACCCA S CCA S CCA S CCCA CCA	K CTCA A GAAA K CAAAA C K GAGG GAG B TCA TCA TCA TCA L L	G N CCAA GAG A R GAT I TGT I TGT V TGA C D H TCT L	L AAAA G K GGAA E CAGCA T T A CCA I I A CCA G A T T GAT I S CAGA T I S CAGA T I S GAT I S GAT I S GAA A A A G K B C S G A A A A A G K B C S C K B C C S C K B C C C C K B C C C C K B C C C C C	T B GAA C N H AAT T B CCGA T B CCGA T B CCGA T B CCGA T B CCGA T B CCGA T B CCGA T C C C N N A C N N H AAT T C N N A A C N N H AAAT N C N A A A C N N H AAAT N C N A A C N N H AAAT N C N A A A C N N H AAAT N C N A A A C N N H AAAT N C N A A A C N N H AAAT N C N A A A C N N H AAAT N C N A A A C N N H AAAT N C N A A A C N N H AAT N C N A A A C N N H A A C C N N H A A T C N N H A A T C N A A T C N A A T C N N H A A T C N A A T C S A A A T C A A A T C S A A A T C S A A A T C S A A A T C S A A A T C S A A A T C S A A T C S A A T C S A A T C S A A T C S A A T C S A A T C S A A T C S A C A A C A C A C A C A C A C A C A	A G S TGGA C E C G G G G G G G G G G G G G G G A G G G A S G A S G A S G A S G A S G G A S S S G G A S C C S S S S S S S S S S S S S S S S	TA R GCTI C CGCC L CGCC L CGCC T C GACG T G C C C C C C C C C C C C C C C C C	G L CGTT L CGCA Q CCAG G CCAG G C CAG C T T T T T A G A Q Q	B AGT V ACA A H N ATT L T T AGT V TAA N ATT T C F_	G K GGGC T A CCAA G N D GGAA A G G TTTA T TCG C R TCG C R C TCCTC	G AGA B TAA N TCC T P L CGA B G GAA N CGGG AA G	I GATA T I CATG M GGGAC AC G AC G AC G C AC G C AC G C C T T Y SCCAA	527 150 401 572 165 546 617 180 591 656 193 636 193 636 193 6301 208 681 746 223 726 791 240
F AC G G F T F C F Q G C G C G C G C G C G C G C G C G C G	T K GCCCG T A A A A A C C C C C C C C C C C C C	B TGT V TAT C M GGCG C R GGCG R GGCGA B SATC TA B T C GGT V	L CCCG GCA Q AGCC CT A S ATC S TTC S CAA T T	K GAA GAAA C K GAAA C K GAG GAG B TCA TCA B CCT L	G N CCAA K GAG A R GAG T GTGI V TGA C D H TTCI L	L AAAA G K GGGA B CAGC TA A CTA A CCA I CCA Q CTGA B	T B GAA C N H AAT B CCGA T B CCGA T B CCGA C C ACA Q H GTC S ACCO P	A G TGA C B GGGA C B GGGA G G G G G G G G G G G G G G G G	TA R GCTI C CGCC L CGCC L CGCC T C GACC T GACC T C A C C C C C C C C C C C C C C C	G L DGTT L DGCA Q CCAG C CCAG R CCAG C R CCAG C T T T A G T A G T A Q N N	B AGT V ACA A H N ATT AGT T T AGT T T C F 8	G K GGGC T A CCAA G N D GGAA G G TTTA T TCG C R TCG C R C TCG C S	G PAGA B TAA N TCC I P L CGA B CGA N CGG AA G N	I GATA T I CATG M CGGAC AC G G G G G G G G G G G G G	527 150 401 572 165 546 617 180 591 656 193 636 193 636 193 636 208 681 746 223 726 223 726 238
F AC B G F B T F C F C G C G C G C G C G C G C G C G C	T K GCCCG T A A A A A C C C C C C C C C C C C C	B TGT TAT C TAT C C M GGG C R GGGA B SATC TA B T C GGT V AACC	L CCCG GCA Q AGCC CT A S ATC S TTCC B TCCC	K GAA GAAA C K GAGAG S C TCA TCA L TCT	G N CCAA K GAG A R GAT I TGI V TGI V TGA	L AAAA G K GGA E CAGC TA T T CACT C CACT C C C C C C C C C C	T B GAAA C N H AAAT B CCGAA T B CCGAA T B CCGAA C ACCA B CACCA P TGI	A G G TGA C B G G G G G G G G G G G G G G G G G G	TA R GCT L GTT C CGC C GC A GC T G A GC T C A GC T C A GC T C A GC T A C G C A GC T A C G C T A C C C C C C C C C C C C C C C C C	G L GGTT L GGCA Q CCAG R CCAG R CCAG T T T T T A GCA C T T T A GCA T T T T T T S S C T	B AGT ACA A H N ATT AGT T AGT T T AGT T T C F 	G K GGGC T A CAAA G N D GGAA N G C T T C C R C C C C C C C C C C C C C	G PAGA B TAA N TCC T P L CGA B CGA N CGG AA G N	I GATA T I CATG M CGGAC AC G C AC G C AC G S S CTAT T Y CCAA	527 150 401 572 165 546 617 180 591 656 193 636 193 636 208 681 746 225 223 726 238
F AC B T T T C F C G C G C G C G C G C G C G C G C G	T K GCCG T A GTA A A CCT T L A A CCT T L A GCA A A CCT T L A CCT T L A CCT T L A CCC A A A A A A A A A A A A A	E TGT TAT C TAT C R GGGG R GGGG R S GGGA S C GGGT V V AACC	L CCCG Q CCT A CCT A CCT A CCT A CCT A CCT A CCT A CCCG CCCG A CCCCG A CCCG A CCCG A CCCG A CCCG A CCCCG A CCCCG A CCCCG	K CTCCA B GGAA A K C A C C C C C C C C C C C C C C C	G N CCAA K GGAG R CGAT I GAT TGA C C D H TCA G	L AAAA G K GGGA B CAGCA T T CGAT I ACCA Q TTGA B ACT	T B B GAAA C N H AAAT M I CCGA T B CCGA T B CACA Q H GTCC B C ACCC P	A G TGAA C B GGGAA C E AGGG C C C C C C C C C C C C C C C C C	TA R GCTI C GTI C CGCC L CGCC L CGCC T C A GCA T C A GCA T C A C C C C C C C C C C C C C C C C	G L CGTT L CGCA CCAG CCAG CCAG T T CCAG C T T CGAC T T S CGAC T T S CGAC T T S T S S T S S T S S T S T S T S T	B AGI V ACA H N ATI C T T AGI V TAA N ATI C F	G K GGGC T A CAAA G N D GGAA N G C T T C R C C R C C C C C C C C C C C C C	G AGA B TAA N TCC I P L CGA B GAA S GAA N CGGG A A G N	I GATA T I CATG M CGGAC G G G G G G G G G G G G G G G G G	527 150 401 572 165 546 617 180 591 656 193 636 208 681 746 225 223 726 238
P AC B GP B T7 T C F Q GC G C A T I GF D	T K GCCG T A A A A CCT T L A A CCT T L A CCT T L A CCT T L A CCT T L A CCT T L A CCT T L A CCT A A A A A A A A A A A A A	E TGT TAT C TAT C C R GGGG R GGGG R GGGA E TA C GGGT V V AACC	L CCCG Q CCT A ACCCT A ATCC B CTTCC B CGAA T N TCCC A P	K GAA A K GAAA C K GAG G GAG B TTCA H S CCT L L	G N CCAA X GAG A R GAT I TGT I TGT V TGA TGT L C TCA G Q	L AAAA G K GGGA B CAGCA TA A T TA A CGAT I ACCA B AACT L	T B B GAAA C N H AAAT M I CCGA T B CCGA T C CGA C C C C C C C C C C C C C C C C C	A G TGAA C B GGGAA G G GGAA S GGAA T G GGTA T G G G G G G G G G G G G	TA R GCI L GTI C C C C C C C C C C C C C C C C C C C	G L GGTT L GGCA Q CCAG G CCAG G T T T AGCA A Q T T T S 8188 24 8	B AGI ACA A H N ATI AGI V TAA N ATI C F	G K GGGC T A CAAA G M D CGAAA N CGAAA N G G C C R C C C R C C C C R	G AGA B TAAA N TCC P L CGA CGA B GAA S GAA S GAA N S CGGG AA G N	I GATA T I CATG M CGGAC G G G G G G G G G G G G G G G G G	527 150 401 572 165 546 617 180 591 656 193 636 746 225 223 726 238 238
P AC B GP B T7 T C F Q GC G G A T I G F D	T K GCCG T A GTA A A CCCT L A GCCA Q T T CCC G P A A CCCA	B TGT V TAT C TAT C C R GCG C R GCG R S GCG T T S GCG T T S GCG T T S GCG T T S GCG C R R S S C C C R S S C C C C R S S S C C C C	L GCCG R GCA Q AGC CT A S ATC S TTCC S TTCC A P	K GAA A K GAAA C K GAAG S GAG S C C T C A T C A T C H S C C T C L	G N CCAA X GGAG A R GGAT I TGT I TGT V TGT D H TCT C C Q Q	L AAAA G K GGGA E AGCAT I ACCA E ACCA E AACT L	T B GAAA C N H AAAT M I CCGAA T B CACA Q H GTCC B ACCC P TGI	A G TGAA C E C C E AGGG C C E AGGG C C C C C C C C C C C C C C C C C	TA R GCI A L M GTI C C C C C C C C C C C C C C C C C C C	G L GGTT L GGCA Q CCAG G CCAG G T R GGAC T T A GGAC T T A GGAC T T A GCAG C CAG G T C CAG C C C C C C C C C C C C C C C C C	B AGT V ACA A H N ATT AGT T TAA N ATT T F	G K GGGC T A CAAA G N D CGAAA N CGAAA N G G A A C C C C R C C C C C C C C C C C C C	G AGA B TAA N TCCC P L CCGA B GAA N CCGG A A G N	I GATA T I CATG M CGGAC G G G G G G G G G G G G G G G G G	527 150 401 572 165 546 617 180 591 656 193 636 701 210 208 681 746 225 223 726 726 238 238



(:	2) (CCJ	CAC	CAC	\TT I	CTJ	TATC	CTT	TTGC	CTCI	[CA]	ATA	GATI	CCI	ATTO	SAT.	TCA			ATI	TTC		FAA	AT?			ACCT	CCA		86
() ()	*) - 2) -	C1	TCC	2					AA			ATT.		AG	AGGZ		AAG	AATC	AAA			ידי. עדאי				TC	AGAC		CAGA	154
(4) (CCI	TC	- AAA/	AGA	GAA	AAG	GAT	CTAP	AGA	AT	AA	CAA	AG	CCC	'AA'	TAT	CAAA	TCA	CA	CCA	AA	AAA	CCF	AAC	SAA/	AG-C	TAA	TTAA	127
()	2) /	AGTT	TCO	CTC		TCI	TCC		AAG	TAC	TAC	CCT	TTTC	STT(A-T	TAGO	GTI			TTT	FTTI	CCP	AA	TACO	CATC	AAT	TCT	238
1	4) <i>F</i> 2) i			-тс. Сът		AGCT A A C	בירא: מיסידי		AAGA	TC1	CT.	ГТ-Т РС-Т	UTTC TTC	ንጉጉር ንጥሮ	TIC	ארא , אא א	AAC:	L'AGG LCAT			2117- 1.07071	י א ה י	2 7 77	 ኮምጥ	0. 70	 ጥር1ጥ		 That	1710	321
()	4)					<u>C</u>	ACC	AA	AAGA	TAZ	GA'	FCT	TTC	200	AGA	AAA	AGC	AAT	ACC	CA	GTC	AT	GTT	TCT	GTG	TGT	CTG	TAT	ATAG	253
(2	2) z	TAA	A-(CAT		TAC	<u>:CC</u> 4	ATA	TTTG	TGT	TAT	GA	CATA	AA	AAGI	GGI	AAA	TTAA	GGT	'AAC	AAA	AAC	GAA-							386
(4	4) <u>Z</u>	TAA	AAC	CAT		TAC	: <u>CC</u> 1	[AA]	TAAG	GTI	TAC	ACA	AATA	GC.	FATA	AA	AGA	GGGA	AAA	TAF	GAI	AGC	GAT	נידידי	TTC	GGG	TGA	GGA	AAG	338
A	rgge	GAAG	AGO	SAAC	GAGI	AGA	GCI	rga.	AGAG	GAI	'AGI	GA	ACAA	AA?	FCAP	CAC	GAC	AGI	AAC	GTI	TGC		AGCO	TAC	GAA	CGG	TTT	GTT	GAAG	476
M	G	R	G	R	v	B	L	ĸ	R	I	B	N	K	I	N	R	0	v	T	F	A	ĸ	R	R	A N	G	L	L	K A	428
A	AAGC	TTA	TGP		GTC	TGI	TCT	CT(GTGA	TGC	TG	AAG	TTGC	TCI	rcai		rct1	ICTC	CAA	.000	TGG	AAA	AGCI	CTA	TGA	GTI	TTG	CAG	CTCC	566
				GC	Т		_		С				CT			G						С			С		С		A s	518
ĸ	A	Y	B	L	g	v	L	с	D	λ	B	v	8	L	I	v	P	9	N	R	G	K	L	Y	B	F	С	8	T	60
TT (ה גאר	#		א גי <i>ר</i> יר	~~~~	יא כית	אישייש	ma	າ <i>ດ</i> ຫ າ	001	~ ~ ~ ~	<u>م</u> س		0007		אתר	- - - - - -	א ר/חר	۸.Cm	~ ~ ~	~ ~ ~	~ ~ ~ ~		maa		202	a cm	m.a. a.	*	656
10	C		GCI		GAC	AC I	G	AA	JGIA	T	GAP	G	T T		4166	C	LCA1	. I GA	AGI	CAA			MCC	190	T.CAA	AGA	G	IGA	GAAC	608
3	N	M	L	K	т	L	D	R	Y	Q	K	с	8	Y	G	s	I	B	v	N	N	ĸ	P	λ	K	R	L	R	N	90
A	GCTA	CAG	AGA	ATA	TCT	GAA	в .GCT	TAT	AGGG	TAG	АТА	TG	AGAA	CCI	TCA	ACG	TCA	ACA	# GAG	ΑΑΑ	тст	тст	TGG	GGA	GGA	דידיד	'AGG	ACC	TTTG	746
				G	СТ	~ .		G	A				A	Т	G			G						A		C	Т		C	698
S	Y	R	B	Y	L	K	L	K	G	R	Y	B	N	L	Q	R	Q	Q	R	N	L	L	G #	R	D	L	G	P	L	120
AA	ATTC	AAA	GGA	GTI	AGA	.GCA	GCT	TG	AGCG	TCA	ACT	GGI	ACGG	CTC	TCT	CAA	GCA	AGT	TCG	GTC	CAT	CAA	GAC	ACA	GTA	CAT	'GCT'	TGA	CCAG	836
N	a	Ŧ		C T	P	0	τ.	P	ъ	0	т.	A	0	đ	т.	G T	•	v	ъ	CG S	т		m	~	v	T		P	•	788
A	5	n	Б	4	5	¥	5	Б	r	*	"	5	G	3		n	¥	•	л	с	*	n	1	¥	1	-	Ц	b	w	190
СЛ	CTC	GGA	TCT	TCA		TAA	AGA G	.GCI	AAT TAA	GTT	GCI	TG	AAC	CAA	TAG	AGC	TTT	'GGC.	AAT	GAA	GCT	GGA	TGA	TAT	GAT	TGG C	TGT	GAG	AAGT	926
г	g	Ď	L	0	N	ĸ	R	0	Т ж	L	L	R	T	N	R	x	L	Å	M	ĸ	L	D	מ	м	т	a	v	R	8 8	180
-		-	-	-	G	••	-	-	I	-	-	D	λ		N			<u>8</u>	-		-	R	-	-	-	•	•		H	
CA	ACTCA	ТАТ	G	-GG	AGG	ATG	GGA	AGC	GCGG	TGA	A	-CA	GAA	TGT	TAC	СТА	CGC	GCA	TCA'	TCA	AGC	TCA	GTC	TCA	GGG	ACT	ATA	CCA	GCCT	1010
н	с н	M	AGG -	a	a	w	P	a	1 0	B	- -	А	N	V.	T	v	т G А	A U	н	^	G N	Q	a	~	•	т	v	~ [′]	P	208
		I	G	Ū	5	"		G	0	D	Q	×	<u>n</u>	ī	λ		G	п	P	¥	^	н	5	¥	G		•	¥	8	210
CT	מטידינ	ልጥር	~~~	TCC	AAC	TCT	GCA	העע	l CCCC) ርጥል	ጥርእ	ጥ አ አ	mee	лат	מישמי	CTTC	TC N	CCN	እ እጥ(~~~	TCC	~ ~ ~	N N C	<u>.</u>	<u> </u>	ጥሮ እ	ccci	2010	2000	1100
C 1	IGA	A10	TG	100	C	T	UCA		T .	A	AG	CC	iicc	noi	G		A	.ucn	nn 1)	GG	T	GAC	GGT	G	G	ICA	T (A AA	1058
L	B	с	N_	_ P	T	L	Q	M	G	Y	D	N	P	v	с	8	B	Q	I	т	X	T	T	Q	λ	Q	X	Q	P	238
00		000	D	съп	maa	100	1 m/2	I	naom	ന്നര	9	H		m n c	mam			0.00		A	V		V		G		9		Q	240
GG	AAA	CGG	TTA C	CAT	C C	AGG. T	ATG C	GAI	GCT	G G	AGA GC	GAT	ATG ACT	TAC TCT	TCC	GAT CCC	GAA AAT	GCT AAA		CCA CTT	CAA	AAG CAA	ACC GTA	TTA CTG	TAT GTG	ATA GGG	TAT/ TCT	<u>baa</u> rcg'	FGGT	1190
G	N	G	Y	I	₽	G	W	м	L	Bn	a	2 4 8																		
											2	250	I																	
(2	AG	ATA	CAA	GAC	TTG	GAT	TTG	TAC	GACA		GTG	GCI	TAAT	ATA	ATG	GTC	CTG	AGG.	ATC	TTC	TAG	ACA	TTT	GTA	TCT	TTT	GGGI	AAT(CCTT	1277
(4) GT	GAT	CTT	'AGA	.GAA TCT	тат	GCA	TAT		294 TAA	таа	TGT	ጥልጥ	TGC	ACA	AGA	CTT	יחידים	ייוייזיי	ртс	TAG	۵۵۵	CDD	GTG	ദറന	ልጥል	CCTO	נמידיב	מידאמ	1235
()	CC	TTC	AAC	ATC	TCT	CTT	CTG	TTT	CAG	GAT	TTG	TTT	GTG	CCT	ATT	GTA	ATT	GCT	TAT	ATA	TGT	ATG	GTT	TGT	ATA	ATG	TGT	GAA	ATGT	1322
	עדי א	۸ <i>С</i> ۸	TCC	<u>م</u>	አጥሮ	നരന	ር እ ጥ		\$ 	۸															\$					
	τH	ncA	100	nu	A10	101	\$	C10	\$	Π																				
	_															_		•.												. .

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

											ATT	TAT	CG	IGTAC	13
GA'	TAC	TTT	ATT	CCT	TTT	ATC	ГАТ	TCT	TGA	AAA	AAA	GTT	AC	CAATT	58
CT'	TGA	GAA	GAA	GAA	GAA	ATC	AGA.	ATC.	AAG	AGA	AGG	AGA	GAC	GAAAG	103
AT	GGGi	AAG	AGG	GAG	AGT	GGA	GAT	GAA	GAG	GAT	AGA	GAA	CA/	AGATT	148
М	G	R	G	R	v	B	M	K	R	I	E	N	K	I	15
AA'	TAG	ACA	AGT	GAC	CTT	CTC	AAA	AAG.	AAG	ААА	CGG	TTT	GC.	IGAA G	193
N	<u>R</u>	0	Y.	T	F	8	K	R	R	N	G	L	L	ĸ	30
AA	AGC	TTA'	TGA	GCT	TTC	TGT	гст	TTG	CGA	TGC	CGA	AGI	TG	TCTC	238
ĸ	λ	Y	B	L	8	v	L	С	D	A	B	v	λ	L	45
AT	CAT	CTT	CTC.	AAG	CCG	TGG	CAA	GCT	ста	CGA	GTT	TGG	TAC	TGTT	283
I	I	F	8	s	R	G	ĸ	L	¥	B	F	G	S	v	60
GG	AAT	TGA	AAG	CAC	AAT	CGA	ACG	GTA	ТАА	TCG	TTG	TTA	CAA	ACTGC	328
G	I	E	8	T	I	E	R	Y	N	R	с	Y	N		75
TC'	гсти	AAG	CAA	таа'	TAA	GCC'	rga.	AGA	GAC	TAC	ACA	GAG	TTC	GTGT	373
8	L	8	N	N	K	P	B	B	T	T	Q	8	W	с	90
CA	GAG	GGTO	SAC	A A A	GCT	гаа	ATC	CAA	ата	CGA	АТС	TCT	TG	TCGT	418
0	B	v	T	K	L	ĸ	g	K	Y	B	8	L	v	R	105
AC	TAA	CAG	GAA	ኮጥጥ	GCT	rggi	AGA.	AGA	тст	TGG	AGA	AAT	GGC	TGTG	463
T	N	R	N	L	L	G	E	D	Ľ	G	E	Ж	G	v	120
۵۵۵	GA		SCA	AGCO	SCTO	GA	AG	GCA	GCT	CGA	AGC	CGC	יידרי	TACC	508
x	B	L	0	λ.	L	E	R	0	L	E	λ	A	L	T	135
••	-	-	_		_	_		_							
GCC	SAC	rcgi	ACA	SCG	~~~~	SAC	ACA	AGT	דאד	GAT	GGA	AGA	AAT	GGAA	553
GC(GAC'	rcg/ R	ACA	GCG(R	CAAC	GACI		AGT'	TAT	GAT	GGA E	AGA B	AAT M	IGGAA E	553 150
	GAC' T		ACA Q	GCG(R			ACAL Q		TAT M MGG	GAT M AGA	GGA E	AGA B		IGGAA E	553 150 598
GCC A GAC D	GAC' T CCT' L	rcg) R Fag(R	ACA Q GAAJ	GCG(R AAA(K	CAAC K GGAC E	GACI T GAGO R	ACA Q GCA	AGT V ACT	TAT MAGG	GAT MAGA	GGA E CAT I	AGA B AAA N	.AA1 M .CA7 K	IGGAA E AACAA Q	553 150 598 165
GCC A GAC D	GAC T CCT L	FCG R FAG(R R	ACA Q GAAJ K	GCG(R AAA(K	CAAC K GGAC E	GACI T GAGO R	ACA Q GCA Q	AGT V ACTI L	TAT MAGG G	GAT MAGA D TGC	GGA E CAT I TTT	AGA B AAA N N	АА] М САР К ААС	IGGAA E AACAA Q	553 150 598 165 643
GCC A GAC D CTC L	GAC T CCT L CAA	rcgi R Fag(R Gat: I	ACA Q GAAJ K FAA K	GCG(R AAA(K GTT: F	CAAC K GGAC E IGAA B	GACI T GAGO R AACO T	ACAJ Q GCAJ Q GGAJ E	AGT V ACT L AGG	TAT MAGG G CCA H	GAT MAGA D TGC	GGA E CAT I TTT F	AGA B AAA N CAA K	AAT M CAF K AAC T	TGGAA E AACAA Q CCTTT F	553 150 598 165 643 180
GCC A GAC D CTC L	GAC	rcg/ R FAG(R GAT: I	ACAO Q GAAJ K TAAO K	GCG(R AAA(K GTT: F GC()	CAAC K GGAC E IGAA B	GACI T GAGO R AACO T	ACAI Q GCAI Q GGAI B GCC	AGT V ACTI L AGGO G	TAT MAGG G CCA H	GAT MAGA D TGC A GGT	GGA E CAT I TTT F	AGA B AAA N CAA K	AAI M CAF K AAC T	TGGAA E AACAA Q CCTTT F	553 150 598 165 643 180 688
GCC A GAC D CTC L CAA	GAC' T CCT' L CAAC K AGAC	TCG/ R FAG(R GAT: I CTT/ L	ACAG Q GAAJ K TAAG K ATGG	GCG(R AAAA(K GTT: F GGC/ A	CAAC K GGAC E IGAA B AAAC N	GACI T GAGO R AACO T CTCO S	ACAJ Q GCAJ Q GGAJ E GGCQ A	AGT V ACT L AGG G G G G G G G C Z	TAT MAGG CCA H ATC	GAT AGA D TGC A GGT V	GGA E CAT I TTT F GGC	AGA B AAAA N CAA K CGG G	AAT M .CAF K AAC T GGA D	TGGAA E AACAA Q CCTTT F ATCCA P	553 150 598 165 643 180 688 195
GCC A GAC D CTC L CA2 Q AAC	GAC T CCT L CAAC K AGAC D	TCG/ R FAG(R GAT: I CTT/ L	ACAG Q GAAJ K TAAG K ATGG W	GCG(R AAA(K GTT: F GGC/ A	CAAC K GGAC E IGAA B AAAC N	GACI T GAGO R AACO T CTCO S GT/	ACAJ Q GCAJ GGAJ E GGCC A AGAC	AGT V ACTI L AGGO G GGCI A GCC	TAT MAGG CCA H ATC S	GAT MAGA D TGC A GGT V TCA	GGA E CAT I TTT F GGC A	AGA B AAA N CAA K CGG G G	AAT M CAF K AAC T GGA D	TGGAA E AACAA Q CCTTT F ATCCA P VATTG	553 150 598 165 643 180 688 195 733
GCC A GAC D CTC L CAA Q AAC N	GACT T CCTT L CAAC K AGAC D CAAT	rcg/ R FAG(R GAT: I CTT/ L FTC: 9	ACAG Q GAAJ K TAAG K ATGG W FGAJ B	GCGC R AAAC K GTT F GGC2 A TT F	CAAC K GGAC E IGAA B AAAAC N ICCC	GACI T GAGO R AACO T CTCO S GGTI V	ACAJ Q GCAJ Q GGAJ E GGCC A AGAC	AGT V ACT AGG G G G G G G C Z A G C Z P	TAT M G CCA H ATC S TTC	GAT MAGA D TGC A GGT V TCA H	GGA E CAT I TTT F GGC A TCC P	AGA B AAA N CAA CAA CGG G TAA N	AAT M CAF K AAC T GGA D TGT	TGGAA E AACAA Q CCTTT F ATCCA P CATTG L	553 150 598 165 643 180 688 195 733 210
GCC A GAC D CTC L CAZ Q AAC N	BAC T CCT L CAAC K AGAC D CAAT	rcg/ R FAGC R GAT: I CTT/ L FTC7 <u>S</u>	ACAG Q BAAJ K ITAAG K ATGG W ITGAJ B	GCGC R AAAC K GTT: F GGCI A TT: F	CAAC K GGAC B IGAZ B AAAC N ICCC P	GACI T GAGO R AACO T CTCO S GGTA V	ACAJ Q GCAJ Q GGAJ E GGC(A AGA(E	AGT V ACT L AGG G G G G G C Z P	TAT M G CCA H ATC S FTC S	GAT AGA D TGC A GGT V TCA H	GGA E CAT I TTT F GGC A TCC P	AGA B N CAA K CGG G TAA N	AAT K AAC T GGA D TGT V	TGGAA E AACAA Q CCTTT F ATCCA P CATTG L	553 150 598 165 643 180 688 195 733 210 778
GCC A GAC D CTC L CAZ Q AAC N GAT	BACT ECT L CAAC K AGAC D CAAT N C	ICGA R FAGC BAT I TTTA L ITCI <u>9</u> CAAC	ACAG Q GAAJ K ITAAG K ATGG W ITGAJ B CACC	GCGC R AAAC K GTT: F GGCJ A TT: F CGAJ R	CAAC K GGAC E IGAJ B AAAC N ICCC P	GACI T GAGO R AACO T CTCO S GGTA V CTTT	ACAI Q GCAI B GGCO A GGCO A CTT/ L	AGT V ACT L AGG G G G G G G G C Z P A CA Z	TAT MGG CCA H ATC S FTC S AAT.	GAT AGA D TGC A GGT V TCA H AGG	GGA E CAT I TTT F GGC A TCC P GTT F	AGA B AAA N CAA K CGG G TAA N TCA	AAT M CAF K AAC T GGA D TGT V ACA	ACAA Q CCTTT F NTCCA P NTTCA L ACAT H	553 150 598 165 643 180 688 195 733 210 778 225
GCC A GAC D CTC L CAZ Q AAC N GAT D	BACT T CCTT L CAAC K AGAC D CAAT N C C TTAC	rcg/ R FAG(R GAT' I CTT/ L TTCT <u>9</u> CAA(N	ACAG Q GAAJ K TAAG K ATGG W IGAJ B CACG T	GCG(R AAA(K GTT) F GGC/ A TT) F CGA/ B	CAAC K GGAC B GAAAC N ICCC P ACCC P	GACI T GAGO R AACO T CTCO S GGTI V F	ACAI Q GCAI Q GGAI B GGCC A AGAC E TTT/ L	AGT V ACT AGG G G G G G G G C A C A C A C A C A C	TAT MAGG CCA H ATC S TTC S AAT. I	GAT AGA D TGC A GGT TCA H AGG Q	GGA E CAT I TTT F GGC A TCC P GTT F AAA	AGA B AAAA N CAA K CGG G TAA N TCA Q CAG	AAT K CAF K AAC T GGA D TGT V ACA Q	RGGAA E AACAA Q CCTTT F ATCCA P VATTG L L ACAT H UCGTG	553 150 598 165 643 180 688 195 733 210 778 225 823
GCC A GAC D CTC L CAA Q AAC N GAT D TAC	BACT T CCTT L CAAC R AGAC D CAAT N TTGC C CTAC	rcg/ R FAGC R GAT: I CTT/ L TTC] _ S CAAC N CAAC	ACAG Q BAAJ K TAAG K ATGG W IGAJ B CACG T GCAJ	GCG(R AAA(K GTT' F GGC/ A TT' F CGA/ B AGGI G	CAAC K GGAC B AAAC N ICCC P ACCC P IGAA K	GACI T GAGO R AACO T CTCO S GTZ S GTZ F AGGO	ACAI Q GCAJ Q GGAJ B GGC(A AGA(B TTT/ L GTC) S	AGT V ACT AGG G G G G G G C C A C A C A C A C A C	TAT M G CCA H ATC S FTC S AAT. I GGT. V	GAT MAGA D TGC A GGT V TCA H AGG G ATC S	GGA E CAT TTT F GGC TCC P GTT F AAA	AGA B AAA N CAA K CGG G G TAA N TCA Q GAG S	AAT M CAF K AAC T GGA D TGT V ACA Q TAA	RGGAA E AACAA Q CCTTT F ATCCA P VATTG L L ACAT H CCGTG V	553 150 598 165 643 180 688 195 733 210 778 225 823 240
GCC A GAC D CTC L CAA Q AAC N GAT D TAC Y	GACT T L CAAC K AGAC D CAAT N C C TTGC C TAC Y	ICG/ R FAGO R GAT' I CTT/ L TTC/ S CAAO N CGTO V	ACA(Q BAAJ K FAA(K ATG(W IGAJ B CAC(T GCAJ Q	GCG(R AAAA(K F GGC/ A ATT' F CGA/ B AGG' G G RAAS	CAAC K GGAC B IGAZ B AAAAC N ICCCC P ICCC P ICCAZ C C C C C C C C C C C C C C C C C C	GACI T GAGO R AAACO T CTTCO S GGT/ V CTTTT F AGGO G	ACAI Q GCAI Q GGAI B GGCC A AGAC E TTT/ L S TCT S	AGT' V ACTI L AGGC G GGC/ A GCC' P A CAL A C A C	TAT MAGG GCCA HATC' S ITTC' S AAT. I GGT. V	GAT MAGA D TGC A GGT V TCA H AGG Q ATC S GGT	GGA E CAT I TTT F GGC A TCC F GTT F AAAA K	AGA B AAA N CAA K CGG G G TAA N TCA Q GAG S TTC	AAT M CAF K AAC T GGA D TGT V ACA Q TAA N N	ACAA Q CCTTT F NTCCA P VATTG L ACAT H SCGTG V 962	553 150 598 165 643 180 688 195 733 210 778 225 823 240
GCC A GAC D CTC L CAA Q AAC N GAT D TAC Y GCA	GAC T CCT L CAAC K AGAC D CAAC C C C T T G C T AGG T G G G	ICGJ R FAG(R GATT I ITT) <u>9</u> CAA(N CGT(V V GA(R	ACA(Q BAAJ TAA(K ATG(W IGA) B CACC T GCA) Q B ACC T	GCGC R AAAA K GTTT F GGCJ A TTT F CGAJ B CGAJ G G G G CAAT	CAAC K GGAC B IGAA B AAAAC P ICCCC P IGAA B ITTCC F	GACI T GAGC R AACC T CTCC 8 GGT/ F AGGC G CGTC V	ACAJ Q GCAJ B GGCA A GGCA TTT I S TCT S CCAJ Q	AGT' V ACTJ L AGG(G G GGC/ P ACAJ Q ITCC S AGGJ G G	TAT M AGG CCA H ATC S TTC S AAT. I GGT. V TTG	GAT MAGA D TGC A GGT V TCA H AGG G GT V	GGA E CAT I TTT F GGC A TCC F AAAA K TCT L	AGA B AAA N CAA K CGG G G TAA N TCA Q GAG S TTG S TTG	AAT M CAF K AAC T GGA D TGT V ACA N A A A A	rGGAA B AACAA Q CCTTT F ATCCA P CATTG L ACAT H SCGTG V 862 252	553 150 598 165 643 180 688 195 733 210 778 225 823 240
GCC A GAC D CTC L CAA Q AAC N GAT GAT GCZ A	GAC T CCT L CAAC K AGAC D CAAC C C C C C C C C C C C C C C C	ICGJ R FAG(R GAT' I CTTJ L CTTJ L CTTJ S CAA(N CGT(V V CGT(V V CGA(E E CTTJ)	ACAC Q GAAA K FAAC K K ATGC W IGAA B CACC T GCAA Q GACC T	GCGCI R AAAA K GTTT F GGCI A TTT F CGAI B G G G I AAGGI G I AAG S I I A	CAAC K GGAC E IGAJ B AAAAC N ICCCC P ACCCC P IGAJ E ITTCC F	GACI T GAGC R AAACC T CTTCC S GGT/ V CTTT F AGGC G G GTCC V	ACAJ Q GCAJ Q GGAJ E GGC(A AGA(E STC) S S CCAJ Q	AGT' V ACTJ L AGG(G G G G G C C S G C C S G G G G G G C C C C	TAT M AGG CCA H ATC S FTC S S TTC V FTG W	GAT AGA D TGC A GGT X TCA H AGG G GTC V	GGA E CAT I TTT F GGC P TCC P GTT F AAAA K TCT L	AGA B AAA N CAA K CGG G G TAA N TCA Q GAG S TTG B D TTG	AAT K CAF K AAC T GGA D TGT V ACA Q TAA N A A	ACGAA B AACAA Q CCTTT F ATCCA P AATTG L LACAT H CGTG V 862 252	553 150 598 165 643 180 688 195 733 210 778 225 823 240
GCC A GAC D CTC L CAA Q AAC N D TAC Y GCA A CTC	GACT T CCTT L CAAC K AGAC D CAAT C C C C C C C C C C C C C C C C C	ICGJ R FAGG R GAT I CTTJ L CTTJ L CTTJ SAAG CAG C CAG C C CAG C C C C C C C C C	ACAC Q GAAA K FAAC K ATGC K GCAA Q GCAA T FGAA T T GGAA	GCGCI R AAAAC K GTTT F GGCI A A TTTT F CGAI B A GGT G CGAI N TTAC	CAAC K GGAC E IGAA B AAAAC N ICCCC P ACCCC P IGAA E ITTCC F GCCCC	GACI T GAGC R AAACC T CTCC 8 GGT/ V CTTT F AGGC G CGTC V CACC TTTT	ACAJ Q GCAJ Q GGAJ E GGGCC A AGAC E TTTJ E GTCT S CCAJ Q GATC	AGT' V ACTI L AGGC G GGC A G G G G G G G G G G G C C A A T C C A C A A T C C A C A C A C	TAT M AGG CCA H ATC S TTC S S TTC S GT. I TTG V V ACG C FAG	GAT AGA D TGC GGT V TCA H AGG GGT V GGTC ATT	GGA B CAT I TTT F GGC P GTT F AAAA K TCT L AGG	AGA B AAA N CAA K CGG G TAA N TCA GAG S TTG E D CCA GGT	AAT M CAF K AAC T G G A T G T T T T T A A A A T A A A T T T	TGGAA B AACAA Q CCTTT F ATCCA P ATTCA L ACAT H CCGTG V 862 252 TCAGC	553 150 598 165 643 180 688 195 733 210 778 225 823 240
GCC A GAC D CTC L CAA Q AAC M TAC GAT D TAC Y CTC TCT	GAC T CCT L CAAC K AGAC D CAAC C CAAC C C C C C C C C C C C C	ICGA R GAT I CTTA L CTTA L CAAC N CGTC V CGAC E CGTC V CGAC C B CGTC ACAC CAAC	ACAC Q GAAJ K IAAC K V IGAJ B CACC T GAC T T GAC T GAT GAT	GGCGC R AAAC K GTTT F GGCJ A TTT GGCJ A CGAJ G G CGAJ G G TTAC STTC AAAC	CAAC K GGAC E IGAA B AAAAC P ICCCC P ICCCC P ICCCC F GCCCC CTTT	GAC/ T GAGC R AACC T CCTCC S GGT/ CCTTT F GGTC V CACC TTTT TTGC	ACAJ Q GCAJ Q GGAJ B GGCO A AGAG C TTTJ L C CAJ Q GATO C CAJ C CAJ	AGT' V ACTI L AGGC G GGC A ACAI Q C C S G G G G C C A ATT TAGI	TAT M AGG CCA H ATC' S ITC' S AATC' S GT. I TGG M ACGG TAG.	GAT AGA D TGC A GGT TCA H AGG GTC SGT STC ATT TTT	GGA E CAT I TTT F GGC TCC F AAA K TCT L AGGGTC	AGA B AAA N CAA CGG G TAA TCA GAG S TTG GAG S CCA GGT	AAT K CAF K AAC TGI V ACA Q TAA N A A A A TAA A TTI TTI	rGGAA B AACAA Q CCTTT F ATCCA P VATTG L ACAT H CCGTG V 862 252 TCCAGC TTTTTT	553 150 598 165 643 180 688 195 733 210 778 225 823 240 908 954 1000
GCC A GAC D CTC L CAA Q AAC N TAC Y GCA A CTC TCC AGC	GACT T CAAC K AGAC D CAAC D CAAC C C C C C C C C C C C C	rcg/ R FAGC R GAT I CTT// L CTT// L CAAC N CGAC N CGAC R CGAC R CAAC C R CAAC C CAAC STGT	ACAC Q Q BAAJ K FAAC K M IGAJ B CACC T GCAJ GCAJ GCAJ GAAJ IGAJ	GCGC R AAAC K GTT? F GGC/ A ATT? F CGA/ B GGA/ G GTAC STTC AAAC ATC?	CAAC K GGAC E IGAJ B AAAAC P IGAJ E ICCCC P IGAJ E ITTTC F GCCCC CTTTT TATJ	GACI T GAGG T CTTC G GGT/ CTTC G G G CGT/ V CACC TTC CACC TTC CACC	ACAJ Q GCAJ Q GGAJ E GGCC A AGAG C CAJ Q GCAJ Q GCAJ Q CAC C CAJ Q CAC C CAJ C CAJ C CAJ C CAJ C CAJ C CAJ C C CAJ C C C C	AGT V L AGGC G GGC2 A GGC2 P A A C A G G G C C A A T T C A A T T C C A A T T C C A A T C C A C C C A C T A C T A C T A C T A C T A C T A C C A C C A C C A C C A C G C C C A G G C C A G G C C A G G C C A C C C C	TAT M AGG CCA H ATCC S S TTC' S ATC' S GT. I TTG' V V GGT. ATG' I ATG' I ATG'	GAT AGA D TGC A GGT TCA H AGG GGT V GTC ATTT TTT GTG	GGA E CAT I TTT F GGC A TCC F AAA K TCT L AGGGTT GTCC CTT	AGA B AAA N CAA K CGG G G TAA N TCA Q GAG S TTG GAG S TTG GAT TGA	AAT KAAC TGI VACA A A A A A A TAA A A TTT TAA AGA	RGGAA B AACAA Q CCTTT F ATCCA P VATTG L ACAT H CCGTG V 862 252 TCCAGC TTTTTT TTCCCA	553 150 598 165 643 180 688 195 733 210 778 225 823 240 908 954 1000 1046
GCC A GAC D CTC L CAA Q AAC A GAT D TAC Y GCA A CTC TCC AGC	BAC' T DECT' L DEAAC K AGAC D DEAAT N CEAAT C TTGC C C TTAC C TTCTA D DECTA	rcg) R FAGG R GATT I TTT/ P CAAC N CAAC V CAAC C GTT ACAC GTT ACAC	ACAC Q GAAJ K FAAC K W IGAJ B GAC T GAC GAC GAC GAAJ GAAJ	GGCG R AAAA K F F GGCJ A ATTT F CGAJ B G G G TAATT N TTAC GTTC AAAC	CAAC K GGAC E IGAJ B AAAAC P IGAJ E ICCC P IGAJ E ITTC F GCCC CTTT ITTT	BAC/ T BAGC R R AACCC T CTCC B GGT/ F CTTCC G GGT/ V CTTCC G G CGTCC V CACCC TTTT TTGC ATTCC	ACAJ Q GCAJ GGAJ E GGCO A GGA SGCO SGCAJ Q GCAJ SACC SACC SCAJ SACC SCAJ SACC SCAJ SACC SCAJ SACC SCAJ SACCAJ SACCAJ SACCAJ SACCAJ SACCAJ SACCAJ SACCAJ SACCAJ SACCAJ SACCAJ SACCAJ SCAJ S	AGT' V L AGGC G GGC/ A GCC/ P A CA C S GCC/ S G G G C C A ATG'	TAT M AGG CCA H ATC S TTC S S TTC S S TTC V TTG V TTG V TTG C ATG TATG	GAT MAGA D TGC A GGT V TCA H AAGG G GTC ATT TTT GTG	GGA E CAT I TTT F GGC A TCC P GTT F AAAA K TCT L AGGG TCT GTCC CTT	AGA B AAAA N CAA K CGG G TAA N TCA Q GAG S TTG GAG S TTG CCA GGT TGA	AAT K AAC T GGA D TGT V ACA Q A A A A A TAA A A A TTT TTAA AGA	AGGAA B AACAA Q CCTTT F VATTG L AACAT H S62 252 TCCAGC TTTTTT ATTCC AATTC	553 150 598 165 643 180 688 195 733 210 778 225 823 240 908 954 1000 1046

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.

Ś

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 AGLs 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 crosshybridization, 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_{i} which was estimated to have an average abundance of 1 in 10⁴ 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)



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 dashedline 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 (\bigcirc) 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 ³⁵S-labeled antisense RNA probes from AGL1 and AGL2 cDNAs. The 3' portions of the cDNAs lacking the putative DNAbinding 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_{i} 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_{i} 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, \sim 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-

		- ·														
	A	G	AGL1		AGL5		AGL2		AGL4		AC	GL3	AGL6		DEF A	
Genes	I	C	M	K	M	K	M	К	M	K	Μ	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	С	I	С	I	С	I	С	I	С	I	С				

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

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 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.

type-specific gene expression (Herskowitz 1990). The

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 ³²P-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.



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 μ m.

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 calmodulindependent 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



Figure 8. A proposed relationship between AG, AGL genes, and DEFA. 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 EcoRI fragment (pCIT565) as a probe. The cDNA library was also screened at high stringency (65°C hybridization and a final wash with $0.2 \times$ SSPE) with several probes (Fig. 1B): probe 1, a 1.1-kb DraI AGL1 genomic fragment (from pCIT1202); probe 2, a 0.36-kb BgIII AGL2 genomic fragment (from pCIT1243). During the construction of the cDNA library, cDNAs with internal EcoRI site(s) were cleaved and then ligated into separate vector molecules. The AGL1, AGL2, AGL3, AGL4, and AGL5 cDNAs all contain at least one EcoRI 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

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 EcoRI-BgIII and 0.79-kb BgIII fragments (from pCIT1202); probe 4 (AGL2), a 2.0-kb HindIII fragment (from pCIT1243); and probe 5 (AGL4), a 1.0-kb EcoRI-BglII 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 EcoRI 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 Scal site (Fig. 1A); pCIT4210 (lacking the MADS box) contains a portion of an AGL6 cDNA 3' of the XhoI 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 BglII fragment from cosmid clone pCIT1210 was found to reveal a Bg/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 EcoRI fragment from AGL2 cosmid pCIT1243 revealed an EcoRI 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 (XbaI polymorphism) and a cross between Landsberg and Niederzenz (BglII polymorphism; Chang et al. 1988) were performed using an AGL2 BglII-EcoRI fragment (subclone pCIT1273 from cosmid pCIT1243) as a probe. For AGL3, a subclone (pCIT1291) with a ~7-kb Bg/II fragment from cosmid pCIT1216 uncovered an XbaI polymorphism between Columbia and Niederzenz ecotypes and a BglII polymorphism between Landsberg and Niederzenz ecotypes and was used to probe appropriate filters. The data from DNA blot hybridization experiments were analyzed using the MAP-MAKER 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 μ g 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 ³²P 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

A family of Arabidopsis transcription factors

probe synthesis: pCIT565 (AG, with putative DNA binding domain), pCIT4219 (AGL1), pCIT4221 (AGL2, 3' NdeI-EcoRI fragment), pCIT4233 (AGL4, SspI-HpaI fragment), pCIT2299 (AGL5), pCIT4210 (AGL6, 3' XhoI-EcoRI 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 ³⁵S-labeled RNA probes complementary to AGL mRNAs. The plasmids pCIT4219 (SP6, XhoI) and pCIT4221 (T7, NdeI) (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

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 CCAATbound 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 cfos 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.

- 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.



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.



Streamline your research with Horizon Discovery's ASO tool