Maydica 47 (2002): 287-301

MAIZE MADS-BOX GENES GALORE[1]

T. Münster[2], W. Deleu[2], L.U. Wingen[2,7], M. Ouzunova[3], J. Cacharrón[2,4], W. Faigl[2], S. Werth[2], J.T.T. Kim[2,5], H. Saedler[2], G. Theißen[2,6,*]

[2] Max-Planck-Institut für Züchtungsforschung (MPIZ), Abteilung Molekulare Pflanzengenetik, Carl-von-Linné-Weg 10, D-50829 Köln, Germany
[3] KWS Saat AG, Postfach 1463, D-37555 Einbeck

Received May 30, 2002

ABSTRACT - MADS-box genes encode a family of transcription factors which control diverse developmental processes in flowering plants ranging from root to flower and fruit development. A large screening for MIKC-type MADS-box gene cDNAs in maize yielded sequences for 31 different genes, 29 of which are of the MIKC-type. 15 of these MIKC-type genes were novel. Together with three published genes for which a cDNA did not appear in our screen 32 different MIKC-type genes have thus been identified in maize now. All of these genes are members of subfamilies known from eudicots. However, it appears that in many subfamilies there are more gene members in maize than in eudicot model plants such as *Arabidopsis*. Phylogeny reconstructions involving all published MADS-box genes identified two major reasons for this. First, after the establishment of the defined gene subfamilies in a common ancestor of eudicots and monocots, a number of gene duplications occurred in the lineage that led to extant monocots after the eudicots had branched off, but before the separation of the lineages that led to extant maize and rice. Based on our gene collection we could estimate that there must have been at least 20 different MIKC-type genes in the most recent common ancestor of maize and rice about 50-70 million years ago. In contrast, the same data set supports only the presence of at least 11 different genes in the last common ancestor of monocots and eudicots about 200 million years ago. Second, phylogenetic trees in line with chromosomal mapping data revealed that the event that gave rise to the ancient segmental allotetraploidy of the maize genome established typically two young paralogs for many orthologous rice MADS-box genes. By chromosomal mapping also candidate genes for some interesting maize developmental gene loci could be identified. The genes reported here are a rich ressource for further studies on the evolutionary dynamics of a complex gene family, the developmental genetics of maize, and a rational crop design employing developmental control genes as tools.

KEY WORDS: Flower development; Gene family; Evo-devo; Crop design.

INTRODUCTION

MADS-box genes (Schwarz-Sommer *et al.*, 1990) encode a family of transcription factors which play crucial roles in developmental control and signal transduction in higher eukaryotes, including plants (for recent reviews, see Riechmann and Meyerowitz, 1997; Theißen *et al.*, 2000; Ng and Yanofsky, 2001; Theißen, 2001b). For example, loss-of-function of some flowering plant MADS-box genes such as *DEFICIENS* (*DEF*) from *Antirrhinum majus* (snapdragon) and *AGAMOUS* (*AG*) from *Arabidopsis thaliana* (thale cress) causes homeotic transformations of floral organs (Sommer *et al.*, 1990; Yanofsky *et al.*, 1990). *DEF* and *AG* thus work as organ identity (homeotic selector) genes during flower development. Floral organ identity genes can be subdivided into at least five different classes, termed A-, B-, C-, D- and E-function genes, which provide five different homeotic functions, with A specifying sepals, A+B+E petals, B+C+E stamens, C+E carpels, and D ovules (Weigel and Meyerowitz 1994; Angenent and Colombo, 1996; Theißen, 2001b; Theißen and Saedler, 2001). Except some A-function genes (such as *APETALA2* from

---

[4] Present address: Dr. Cacharron & Ortmann Wirtschaftsberatung, Othestr. 2, 51702 Bergneustadt.

[5] Present address: Universität zu Lübeck, Institut für Neuro- und Bioinformatik, Seelandstr. 1a, D- 23569 Lübeck, Germany.

[6] Present address: Friedrich-Schiller-Universität Jena, Lehrstuhl für Genetik, Philosophenweg 12, D-07743 Jena, Germany.

[7] Present address: Max-Planck-Institut für Züchtungsforschung, Zentrum zur Identifizierung von Genfunktionen durch Insertionsmutagenese (ZIGIA), Carl-von-Linné-Weg 10, D-50829 Köln, Germany.

[1] Data deposition: The nucleotide sequence data of the cDNAs reported in this paper have been deposited in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers given in Table 1.

[*] For correspondence (fax: +49-3641-949552; e-mail: guenter.theissen@uni-jena.de).

Arabidopsis), all floral organ identity genes that could be cloned so far belong to the family of MADS-box genes (for reviews see THEIßEN *et al.*, 1996, 2000; RIECHMANN and MEYEROWITZ 1997; NG and YANOFSKY 2001; THEIßEN, 2001b). The MADS-type floral homeotic genes of *Arabidopsis* are *APETALA1* (*AP1*; A-function), *APETALA3* and *PISTILLATA* (*AP3* and *PI*; B-function), *AG* (C-function), and *SEPALLATA1*, *SEPALLATA2* and *SEPALLATA3* (*SEP1-3*, formerly known as *AGL2*, *AGL4*, and *AGL9*; E-function) (PELAZ *et al.*, 2000; HONMA and GOTO, 2001). D-function genes, termed *FBP7* and *FBP11*, are known from *Petunia* (ANGENENT and COLOMBO, 1996), but an *Arabidopsis* D-function gene has not been described so far.

Besides providing floral homeotic functions, MADS-box genes have many other roles within the gene networks that "control" reproductive development in eudicotyledonous flowering plants such as *Arabidopsis* (for recent reviews see RIECHMANN and MEYEROWITZ, 1997; THEIßEN and SAEDLER, 1999; THEIßEN *et al.*, 2000; NG and YANOFSKY, 2001; THEIßEN, 2001b). The MADS-box genes *FLC* and *SVP*, for example, are 'flowering time genes' which repress the floral transition until internal factors reflecting plant age, or environmental factors such as cold (vernalization) act to overcome this repression (MICHAELS and AMASINO, 1999; HARTMANN *et al.*, 2000; SHELDON *et al.*, 2000). *SOC1* (also known as *AGL20*), another MADS-box gene, is an important integrator of vernalization and two other pathways controlling flowering time in *Arabidopsis* (LEE *et al.*, 2000).

Flowering time genes exert their function by influencing 'meristem identity genes', either directly or indirectly. Meristem identity genes such as the MADS-box genes *AP1*, *FRUITFULL* (*FUL*) and *CAULIFLOWER* (*CAL*) 'control' the transition from inflorescence to floral meristems. Within floral meristems, 'cadastral genes' are involved in establishing the boundaries of floral organ identity gene functions, thus defining the different floral whorls. Besides its role as a floral organ identity gene, *AG* has also a cadastral function, because it prevents the A-function from being expressed in the 3rd and 4th floral whorls. The floral organ identity genes specify the organ identity within each whorl of the flower by activating 'realizator genes'.

MADS-box genes are also involved in developmental processes that follow fertilization of the flower, i.e. seed and fruit development. For example, *SHATTERPROOF1* and *2* (*SHP1* and *2*; formerly known as *AGL1* and *AGL5*, respectively) encode functionally redundant proteins that are required for the proper development of the fruit dehiscence zone of *Arabidopsis* (LILJEGREN *et al.*, 2000), and *FUL* is required for the normal pattern of cell division, expansion and differentiation during morphogenesis of the silique (GU *et al.*, 1998).

Moreover, transcription of a number of MADS-box genes outside flowers and fruits suggests that members of this gene family play regulatory roles also during vegetative development, such as embryo, root, or leaf development (e.g. MA *et al.*, 1991; HUANG *et al.*, 1995; ROUNSLEY *et al.*, 1995; for a review, see THEIßEN *et al.*, 2000). Analysis of a mutant generated by transgenic technology indicated that one gene, *ANR1*, is a key component of the signal transduction chain by which nitrate stimulates lateral root proliferation (ZHANG and FORDE, 1998). The existence of MADS-box genes in gymnosperms, ferns and mosses, which do not form flowers or fruits, further demonstrates that the role of these genes in plants is not restricted to flower or fruit development (TANDRE *et al.*, 1995; MÜNSTER *et al.*, 1997; WINTER *et al.*, 1999; KROGAN and ASHTON, 2000; HENSCHEL *et al.*, 2002).

Changes in the number, expression and interaction of developmental control genes have very likely contributed to the evolution of plant form (THEIßEN *et al.*, 2000). Since MADS-box genes play important and diverse roles in the gene networks that control plant development, understanding the phylogeny of MADS-box genes may strongly improve our understanding of plant evolution (THEIßEN and SAEDLER, 1995; THEIßEN *et al.*, 1996, 2000; LAWTON-RAUH *et al.*, 2000; MA and DEPAMPHILIS, 2000; NIKLAS, 2000). It is of quite some interest, therefore, to find out as to when during evolution the diversity of MADS-box genes present e.g. in *Arabidopsis* has been generated. Did it appear during eudicot evolution, or is it considerably older? Do changes in the expression and function of these genes reflect morphological innovations during angiosperm evolution? To answer these questions, the phylogeny of the MADS-box genes has to be reconstructed, and superimposed on the phylogeny of land plant taxa. As a prerequisite, the MADS-box gene family has to be characterized in phylogenetically informative key taxa, including the major groups of flowering plants (angiosperms). Flowering plants comprise two large clades (subclasses), the eudicots and monocots, which are nested within the ancestral magnoliids (basal angiosperms) (CRANE *et al.*, 1995).

In *Arabidopsis thaliana*, we are already quite close to a complete knowledge of all MADS-box

genes: according to current estimates, there are more than 80 different MADS-box gene members in the *Arabidopsis* genome (RIECHMANN *et al.*, 2000). Annotation for all these genes is underway in several bioinformatics labs. For about half of these genes, cDNAs have been isolated. For 15 genes (all mentioned above) the function has already been defined by mutant phenotypes, and for many others respective work is in progress.

In the other major groups of flowering plants (monocotyledonous plants and basal angiosperms), the characterization of MADS-box genes is lagging behind, although publication of draft sequences of the rice genome suggest that the situation may change soon, at least for one monocot species (GOFF *et al.*, 2002; YU *et al.*, 2002). Studies on several monocot genes suggest that the floral homeotic B and C functions are strongly conserved between monocots and eudicots (MENA *et al.*, 1996; KANG *et al.*, 1998; AMBROSE *et al.*, 2000), but also novel gene functions, e.g. concerning inflorescence development, seem to have originated within monocots (CACHARRÓN *et al.*, 1999). However, a large scale comparison between the MADS-box gene families from monocots and eudicots has been lacking so far due to our very incomplete knowledge about MADS-box genes in monocots. Such a comparison, however, could not only improve our understanding of plant evolution, but may also be interesting from an agronomic point of view.

Most human food and many other useful products such as timber and lumber are derived from seed plants. Understanding the genetic basis of seed plant development may thus help us to transform these plants according to our desires. The time to flowering, and the number and structure of the inflorescences and flowers, for example, are critical parameters that strongly influence where a crop plant can be grown and how many fruits or grains it may produce. It might be possible to create crop plants that flower earlier or later than wild-type by changing the expression of floral meristem identity genes. In addition, plants that produce more fruits or grains might be established by changing the expression of floral organ identity genes or other developmental control genes that influence the architecture of flowers or inflorescences. With MADS-box genes at hand it might thus be possible to 'design' crop plants to have novel agronomic features. Since the most important crop plants are monocots (*Triticum aestivum*, wheat; *Oryza sativa*, rice; *Zea mays* ssp. *mays*, maize, or corn; *Hordeum vulgare*, barley; *Sorghum bicolor*, sorghum), a better knowledge about the MADS-box gene family in monocots may provide the tools for a future design of cereal crops by transgenic technology, or mutagenesis and marker assisted breeding (MEYEROWITZ, 1994; THEIßEN, 2000, 2001a, 2002).

Previous phylogeny reconstructions revealed that the MADS-box gene family is composed of several defined gene clades (DOYLE, 1994; PURUGGANAN *et al.*, 1995; THEIßEN *et al.* 1996, 2000). All the plant MADS-box genes that have been functionally characterized so far are members of a monophyletic superclade of genes with a conserved structural organization, the so called MIKC-type domain structure, including a MADS (M-), intervening (I-), keratin-like (K-) and C-terminal (C-) domain (MA *et al.*, 1991; THEIßEN *et al.*, 1996; HASEBE and BANKS, 1997; MÜNSTER *et al.*, 1997). The highly conserved MADS-domain is the major determinant of DNA-binding, but it also performs dimerization and accessory factor binding functions (SHORE and SHARROCKS, 1995). The relatively weakly conserved I-domain constitutes a molecular determinant for the selective formation of DNA-binding dimers (RIECHMANN and MEYEROWITZ, 1997). The K-domain is defined by a conserved regular spacing of hydrophobic residues, which is proposed to allow for the formation of an amphipathic helix involved in protein dimerization (MA *et al.*, 1991; SHORE and SHARROCKS, 1995). The most variable region, both in sequence and length, is the C-domain at the C-terminus of the MADS-domain proteins, which is involved in transcriptional activation, or the formation of multimeric transcription factor complexes (CHO *et al.*, 1999; EGEA-CORTINES *et al.*, 1999).

The MIKC-type gene superclade can be further subdivided into several well defined gene clades whose members share similar expression patterns and highly related functions. Most importantly, all A-, B-, C-, D- and E-function genes known so far fall into separate clades, namely *SQUAMOSA*- (A-function), *DEFICIENS*- or *GLOBOSA*- (B-function), *AGAMOUS*- (C- and D-function) and *AGL2*-like genes (E-function) (DOYLE, 1994; PURUGGANAN *et al.*, 1995; THEIßEN and SAEDLER, 1995; ANGENENT and COLOMBO, 1996; THEIßEN *et al.*, 1996, 2000; MÜNSTER *et al.*, 1997). Therefore, the establishment of the mentioned gene clades by gene duplication, diversification and fixation was probably an important step towards the establishment of the floral homeotic functions, and thus flowers (THEIßEN *et al.*, 1996, 2000).

Previous studies have led to the conclusion that the last common ancestor of ferns and seed plants (gymnosperms + angiosperms) about 400 million years ago (MYA) had at least 2 different MIKC-type genes, but no orthologs of any of the MADS-box genes present in angiosperms (MÜNSTER *et al.*, 1997; THEIßEN *et al.*, 2000). In the last common ancestor of angiosperms and gymnosperms about 300 MYA, however, there were very likely already at least 7 different MADS-box genes present, including orthologs of floral homeotic B- and C-function genes (WINTER *et al.*, 1999; BECKER *et al.*, 2000; THEIßEN *et al.*, 2000).

In eudicotyledonous angiosperms, 11 different ancient paralogous MADS-box gene subfamilies have been defined by phylogeny reconstructions so far (THEIßEN *et al.*, 2000). For some of these gene groups members from monocots have not been reported yet. In other cases, just one gene from monocot species is known, where several clade members are known from eudicots. Moreover, no MIKC-type MADS-box gene from a monocot has been reported yet which does not correspond to a well defined subfamily known from dicots. This all may either reflect a lower complexity and diversity of the MADS-box gene family in monocots, or it may simply be due to the less intensive characterization of the MADS-box gene family in this taxonomic group.

In order to better characterize the MADS-box gene family in a monocotyledonous model system, we did a large screen for MADS-box gene cDNAs in maize (*Zea mays* ssp. *mays*). Here we report the up to now most comprehensive cloning of MADS-box genes from a monocotyledonous plant. Our data demonstrate that at least the 'MIKC branch' of the MADS-box gene family of maize is of similar complexity as the one of eudicots. Phylogeny reconstructions were used to determine the minimal number and type of MADS-box genes that was already present in the last common ancestor of monocots and eudicots, and of maize and rice. The agronomic perspectives of our findings are discussed.

MATERIALS AND METHODS

Plant material

Maize inbred lines C and T232 (EDWARDS *et al.*, 1987), and recombinant inbred (RI) lines TxCM and COxTx (BURR and BURR, 1991) were used throughout this work. Plants were grown in climate chambers under moderate long-day (14 hours light at 25°C, 10 hours darkness at 18°C) or under standard greenhouse conditions.

Isolation of cDNAs by screening of a library

A maize ear cDNA library described previously (CACHARRÓN *et al.*, 1999) was screened several times by plaque hybridization with radioactive probes derived from 14 different cDNAs representing the maize genes *ZMM2, 7, 15, 16, 17, 18, 20, 21, 23, 24, 27, 31, ZmMADS2* and the lily gene *LRGLOA*. Hybridization was performed under conditions of moderate stringency (58°C, 5xSSC) following standard procedures (SAMBROOK *et al.*, 1989). Plasmid clones obtained after in vivo excision contained cDNAs of diverse genes. For each gene, representative cDNAs were completely sequenced on both strands by the MPIZ DNA core facility ('ADIS') on PE Biosystems Abi Prism 377 and 3700 sequencers using BigDye-terminator chemistry. Oligonucleotides used were purchased from MWG, Metabion or LifeTech.

Isolation of cDNAs by RACE

Partial cDNAs were isolated by 3'-RACE as generally described (FROHMAN *et al.*, 1988; THEIßEN *et al.*, 1995; MÜNSTER *et al.*, 1997). As template for the synthesis of cDNA pools poly A^+-RNA isolated from the following tissues or organs of *Zea mays* ssp. *mays* lines C and T232 was used: roots, juvenile and adult leaves, coleoptiles, male and female inflorescences, developing kernels, and pollen. In some cases, upstream sequences overlapping with the 3'-fragment were isolated by 5'-RACE, employing a commercially available kit (5'/3'-RACE-Kit, Boehringer Mannheim, Germany). Sequences of primers used during the RACE procedures are available upon request. For each gene, at least three different cDNA sequences were independently cloned, and both strands were sequenced on automatic sequencers by the Automatic DNA Isolation and Sequencing team (ADIS) of our institute.

Mapping of genomic loci

Chromosomal map positions of MADS-box genes were determined by applying RFLP technology to maize recombinant inbred lines TxCM and COxTx (BURR *et al.* 1988; BURR and BURR, 1991) as generally described previously (FISCHER *et al.*, 1995). Gene-specific hybridization probes were prepared employing the different cDNAs as templates. Comparative maps for parts of maize chromosomes were depicted by information available at the Recombinant Inbred Maize Genome Database (http://burr.bio.bnl.gov/acemaz.html).

Sequence Alignments and Construction of Phylogenetic Trees

Accession numbers of all sequences used are available at the MADS homepage (http://www.mpiz-koeln.mpg.de/mads). Multiple alignments of conceptual amino acid sequences were generated by using the PILEUP program of the GCG package (version 10.0) with a gap creation penalty of 8 and a gap extension penalty of 2 (default parameters). Based on alignments of the MADS-domain (60 amino acids) plus the 110 amino acids downstream of the MADS-domain (termed "MADS+110" or "170" domain sequence; refs. THEIßEN *et al.*, 1996; WINTER *et al.*, 1999), phylogenetic trees were constructed by the neighbor-joining (NJ) method (SAITOU and NEI, 1987), version 3.5, as implemented by the PHYLIP program package (FELSENSTEIN, 1993). NJ was chosen because it is known to be quite efficient in obtaining reliable trees from large sets of data (ZHANG and NEI, 1996). Distance matrices were generated using the protein distance algorithm, version 3.55c, which is based on the PAM model of amino acid transition (DAYHOFF, 1979). To assess support for the inferred relationships, 100 bootstrap samples were generated as described (MÜNSTER *et al.*, 1997).

TABLE 1 - *Cloned MIKC-type MADS-box genes from maize.*

Gene[a]	Subfamily[b]	Marker designation[c]	Map position[d]	Accession No.[e]	Reference[f]	Defining clone[g]
SILKY1	DEF	n.p.	6L020	AF181479	AMBROSE *et al.*, 2000	/
ZAG1	AG	ucsd78A	6L080	L18924	SCHMIDT *et al.*, 1993	pSW203
ZAG2	AG	ucsd81A	3L084	L18925	SCHMIDT *et al.*, 1993	pSW611
ZAG3	AGL6	ucsd78C	4L098	L46397	MENA *et al.*, 1995	/
ZAG5	AGL6	ucsd72E	5L151	L46398	MENA *et al.*, 1995	/
ZAP1	SQUA	ucsd106A	2L200	L46400	MENA *et al.*, 1995	pSW85
ZMM1	AG	mpik31	10L054	AJ430630	THEIßEN *et al.*, 1995	pWFE074
ZMM2	AG	mpik24	8L056	AJ430631	THEIßEN *et al.*, 1995	pSW123
ZMM3	AGL2	mpik25	9S034	Y09301	FISCHER *et al.*, 1995	pWF4012
ZMM4	SQUA	mpik22A	1L224	AJ430641	FISCHER *et al.*, 1995	pSW02
ZMM5	TM3	kws2	9L140	AJ430642	This work	pSW290
ZMM6	AGL2	mpik23A	1L131	AJ430692	FISCHER *et al.*, 1995	pSW06
ZMM7	AGL2	mpik27	7L095	Y09302	FISCHER *et al.*, 1995	pSW293
ZMM8	AGL2	mpik28	9L115	Y09303	CACHARRÓN *et al.*, 1999	pSW24
ZMM14	AGL2	mpik43	1S055	AJ005338	CACHARRÓN *et al.*, 1999	pWF1552
ZMM15	SQUA	n.d.	n.d.	AJ430632	This work	pSW31
ZMM16	GLO	mpik39	3L110	AJ292959	MÜNSTER *et al.*, 2001	pSW436
ZMM17	GGM13	mpik40	5S107	AJ271208	BECKER *et al.*, 2002	pSW450
ZMM18	GLO	mpik41	8L079	AJ292960	MÜNSTER *et al.*, 2001	pSW559
ZMM19	STMADS11	mpik42	4L145	AJ430633	This work	pWF4123
ZMM20	STMADS11	mpik44	1S053	AJ430634	This work	pWFE7a8
ZMM21	STMADS11	mpik45	9S023	AJ430635	This work	pWF4105
ZMM23	AG	kws6	3L095	AJ430637	This work	pSW128
ZMM24	AGL2	kws7	1L245	AJ430638	This work	pWF1557
ZMM25	AG	mpik47	3L109	AJ430639	This work	pSW71
ZMM26	STMADS11	mpik48	5L215	AJ430693	This work	p423sw
ZMM27	AGL2	n.d.	n.d.	AJ430694	This work	pBRACE9-20
ZMM28	SQUA	n.d.	n.d.	AJ430695	This work	pJC17
ZMM29	GLO	mpik41	8L079	AJ292961	MÜNSTER *et al.*, 2001	pSW425
ZMM31	AGL2	n.d.	n.d.	AJ430640	This work	pWFH1243
ZmMADS2	AGL17	n.d.	n.d.	AF112149	HEUER *et al.*, 2000	pWFI096
ZmMADS3	SQUA	n.p.	7S000	AF112150	HEUER *et al.*, 2001	pSW159

[a] Gene names are given in alphabetical order. Only MIKC-type genes for which sequence information is available are listed here. The non-MIKC-type transposon-like *TMZ1* elements (Transposed MADS-box elements of *Zea* No. 1; also termed *ZEM* genes; *ZAG4* is also a representative of this class of sequence elements), described (FISCHER *et al.* 1995; MENA *et al.* 1995; MONTAG *et al.* 1995, 1996) and reviewed (THEIßEN *et al.* 2000) previously, and also *ZMM22* and *WFE096*, are not considered here. *ZmMADS1* (HEUER *et al.*, 2001) is not listed here, since based on sequence comparisons it is obviously identical to *ZMM5* (partial sequence described already by THEIßEN *et al.*, 1996). *ZMM30* is not listed here, because it is identical to *ZmMADS2*, and we adopt this name here for priority reasons.

[b] Nomenclature of subfamilies according to the rules outlined elsewhere (THEIßEN *et al.*, 1996).

[c] The marker designations used during BNL mapping are given (n.p., not published; n.d., non determined).

[d] Approximate position on a recent edition of the Recombinant Inbred Maize Genome Database at BNL is given (http://burr.bio.bnl.gov/ace-maz.html).

[e] Accession numbers for cDNA sequences are given. Additional accession numbers for some protein or genomic DNA sequences are available at the MADS homepage (http://www.mpiz-koeln.mpg.de/mads).

[f] References are given for the first publication of a given gene, cDNA or protein sequence.

[g] Defining clone means the cDNA clone in our lab which has been most extensively characterized and is used as a reference ("/", no clone isolated in our lab).

RESULTS AND DISCUSSION

Isolation and structural evaluation of MADS-box gene cDNA clones from maize

By screening a maize (*Zea mays* ssp. *mays*) line T232 cDNA library made from immature ears (CACHARRÓN *et al.*, 1999) several times with different MADS-box gene probes, and by RACE (Rapid Amplification of cDNA Ends) on cDNA pools representing different maize tissues and organs (see Materials and Methods), more than 1400 cDNA clones of MADS-box genes were isolated and partially sequenced. Sequence comparisons and Southern hybridization studies suggested that these clones rep-

resent 31 different genes (data not shown). Conceptual translation of the cDNAs yielded proteins which show the typical domain structures of MIKC-type MADS-domain proteins, except for two genes, *ZMM22* and *WFE096*, which are not considered here further (data not shown). Extensive data base searches and sequence comparisons revealed that of the remaining 29 putative genes, 14 had been published before, while 15 were completely novel. Of the novel genes, four have been presented recently elsewhere (MÜNSTER *et al.*, 2001; BECKER *et al.*, 2002), while the other 11 are reported here (Table 1). Together with the *SILKY1*, *ZAG3* and *ZAG5* genes, for which a cDNA was not picked up in our screen, for 32 different MIKC-type MADS-box genes from maize extensive sequence information is now available (Table 1).

Phylogeny reconstructions I: the monocot - eudicot dichotomy

To clarify the evolutionary relationships between the maize genes and all other MIKC-type MADS-box genes known, phylogeny reconstructions were carried out. This way, putative orthologs could be identified at different taxonomic levels, thus allowing minimal estimates for the number of MIKC-type genes in the most recent common ancestors of monocots and eudicots, and of rice and maize. Initially, the phylogeny reconstructions involved all available MIKC-type MADS-box genes (data not shown). For simplicity, smaller data sets were used later, when the ommited sequences provided no relevant information. A representative tree based on conceptual protein sequences is shown in Fig. 1.

It appears that the MADS-box genes isolated from maize are members of 10 different subfamilies already known from eudicots. These subfamilies, probably constituting monophyletic gene clades, comprise the *SQUA*-, *AGL6*- and *AGL2*-like genes, together also constituting a clade (THEIßEN *et al.*, 1996; MÜNSTER *et al.*, 1997), the *TM3*-, *AG*-, *DEF*- and *GLO*-like genes (THEIßEN *et al.*, 1996), the *STMADS11*-like genes (BECKER *et al.*, 2000), the B_{sister} (B_s) genes (the putative sister genes of the *DEF*- and *GLO*-like genes and their gymnosperm orthologs; BECKER *et al.*, 2002), and the *AGL17*-like genes (Fig. 1). For the *Arabidopsis* gene *AGL12* a putative ortholog from rice (*OSMADS26*) has been identified, suggesting that an ortholog from maize may await isolation, or has been lost in the lineage that led to maize, but after the split of the maize and rice lineages. Taken together, 11 different clades have been identified containing both monocot and eudicot genes. We conclude that at least 11 different MIKC-type MADS-box genes were already present in the last common ancestor of monocots and eudicots, which, according to molecular estimates (WOLFE *et al.*, 1989; SAVARD *et al.*, 1994), existed about 200 MYA. These ancestral genes were early representatives of the 11 different ancient gene clades mentioned above. As explained generally elsewhere (BECKER *et al.*, 2000) this is only a minimal estimate, and the real number is possibly higher. For example, it could well be that some of the subfamilies described here comprise two or more ancient clades with each having both monocot and eudicot representatives.

Phylogeny reconstructions II: the maize - rice split

Employing the gene tree shown in Fig. 1, minimal clades containing genes from both maize and rice could be identified. This reveals putative orthologs of MIKC-type genes in maize and rice, and provides a minimal estimate for the number of MIKC-type genes in the most recent common ancestor of rice and maize.

A number of gene subfamilies shared by eudicots and monocots contains just one minimal clade including both rice and maize MIKC-type genes. Also putative orthologs from other monocots (mostly cereals) can often be identified this way. These subfamilies comprise the *DEF*-like genes (1.), with *SILKY1* from maize and its putative ortholog from rice, *OSMADS16*; *TAMADS51* and *LRDEF* are putative *DEF* orthologs from wheat and lily (*Lilium regale*), respectively (Fig. 1); the *TM3*-like genes (2.), with *ZMM5* (*ZmMADS1*) from maize and *FDRMADS8* from rice; the *AGL6*-like genes (3.), containing *ZAG3* and *ZAG5* from maize, *OSMADS6* from rice, *TAMADS12* from wheat and *LRAGL6* from lily (*Lilium regale*). The fact that there is a second rice gene, *OSMADS17*, sister to this clade suggests that there may be more *AGL17*-like genes in the other monocot genomes as well.

The B_s (4.), *AGL12*- (5.) and *AGL17*-like genes (6.) are special in that either only a rice (5.), or only a maize gene (4., 6.) has been published for these clades so far. However, since they comprise also eudicot genes, ancient representatives were very likely present in the most recent common ancestor of eudicots and monocots, and hence also in that of all monocots. Thus orthologs from the other cereal species may be expected, although, in principle, genes could also have been lost in the one or the

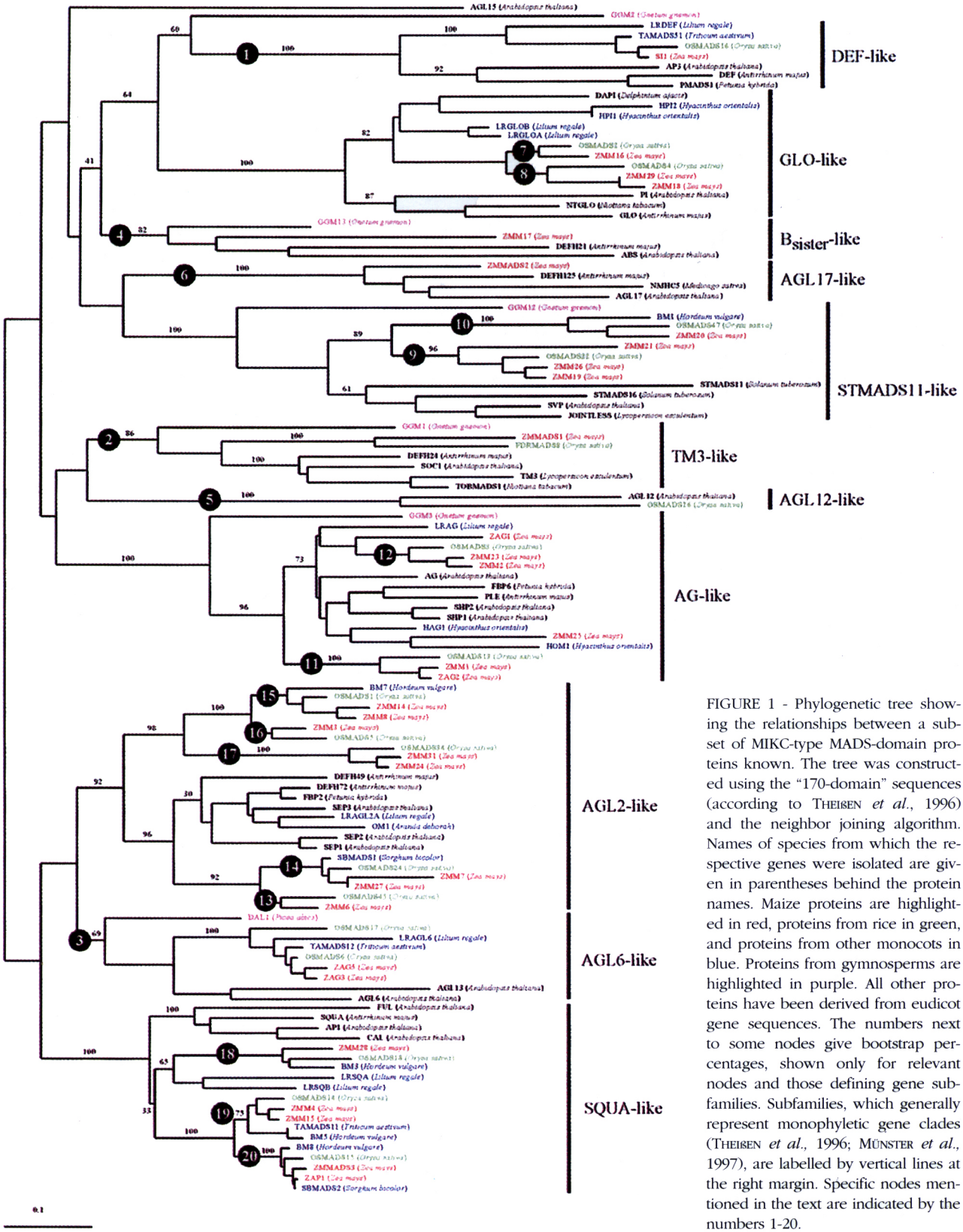

FIGURE 1 - Phylogenetic tree showing the relationships between a subset of MIKC-type MADS-domain proteins known. The tree was constructed using the "170-domain" sequences (according to THEIßEN *et al.*, 1996) and the neighbor joining algorithm. Names of species from which the respective genes were isolated are given in parentheses behind the protein names. Maize proteins are highlighted in red, proteins from rice in green, and proteins from other monocots in blue. Proteins from gymnosperms are highlighted in purple. All other proteins have been derived from eudicot gene sequences. The numbers next to some nodes give bootstrap percentages, shown only for relevant nodes and those defining gene subfamilies. Subfamilies, which generally represent monophyletic gene clades (THEIßEN *et al.*, 1996; MÜNSTER *et al.*, 1997), are labelled by vertical lines at the right margin. Specific nodes mentioned in the text are indicated by the numbers 1-20.

other lineages during evolution. Indeed, employing data provided by the rice genome projects (GOFF *et al.*, 2002; YU *et al.*, 2002), putative orthologs of the maize B_S and *AGL17*-like genes could recently be identified from rice (A. BECKER, personal communication; G.T., unpublished observations). In contrast, whether there is an *AGL12*-like gene in maize still remains to be seen.

The other gene subfamilies contain (sometimes only potentially) more than one clade including maize as well as rice genes. In case of the *GLO*-like genes, there are two clades, one comprising *OSMADS2* and *ZMM16* (7.), and the other one *OSMADS4*, *ZMM18* and *ZMM29* (8.), indicating that the most recent maize-rice ancestor contained already at least two *GLO*-like genes (Fig. 1; MÜNSTER *et al.*, 2001). In case of the *STMADS11*-like genes, we have one clade (9.) comprising *ZMM19* and *ZMM26* from maize, and *OSMADS22* from rice; and another one (10.) comprising *ZMM20* from maize, *OSMADS47* from rice, and *BM1* from barley (Fig. 1). The maize gene *ZMM21*, being sister to the clade containing *ZMM19* (Fig. 1), may indicate that there were even more than two *STMADS11*-like genes in the most recent common ancestor of maize and rice.

Within the subfamily of *AG*-like genes, we have one clade (11.) containing *ZAG2* and *ZMM1* from maize and *OSMADS13* from rice. There is another clade (12.) containing *ZMM2* and *ZMM23* from maize and *OSMADS3* from rice. *ZAG1* from maize (Schmidt *et al.*, 1993) is sister to this clade, and there is yet another clade comprising *ZMM25* from maize and *HOM1* and *HAG1* from *Hyancinthus orientalis*, which is also a monocot, pointing to even more *AG*-like genes in the most recent common ancestor of maize and rice. All in all, there exist at least 6 different *AG*-like genes in the maize genome.

The *AGL2*-like genes gave rise to quite a number of clades comprising genes from both maize and rice (Fig. 1). There is one clade (13.) which is defined by *ZMM6* from maize and its putative rice ortholog, *OSMADS45* (also known as *OSMADS7*; GRECO *et al.*, 1997; KANG and AN, 1997). Another clade (14.) comprises the maize genes *ZMM7* and *ZMM27*, the rice gene *OSMADS24* (also known as *OSMADS8*; GRECO *et al.*, 1997; KANG and AN, 1997), and the sorghum gene *SBMADS1*. In other clades there are *ZMM8* and *ZMM14* from maize, *OSMADS1* from rice, and *BM7* from barley (15); (16.) *ZMM3* from maize and *OSMADS5* from rice (16.); *ZMM24* and *ZMM31* from maize and *OSMADS34* from rice (17). These data suggest that the subfamily of *AGL2*-like MADS-box genes strongly diversified during monocot evolution. While there is no hard evidence for more than one *AGL2*-like gene in the most recent common ancestor of monocots and eudicots, there are five different clusters of putatively orthologous genes from maize and rice (nos. 13.-17.), suggesting that the most recent common ancestor of these two species contained already at least five different *AGL2*-like genes. These may have been generated, and maintained, after the monocot-eudicot split, but before the maize-rice split. Mutant analysis has shown already that the *OSMADS1* (also known as *LEAFY HULL STERILE1*) gene from rice functionally resembles to a certain extent the *SEPALLATA* genes from *Arabidopsis* (JEON *et al.*, 2000). However, functional studies on more monocot *AGL2*-like genes from different species are required to determine the extent of conservation, neofunctionalization and subfunctionalization which has been involved in the evolution of the diversity of *AGL2*-like genes in grasses.

Within the subfamily of *SQUA*-like genes, there is one clade (18.) comprising *ZMM28* from maize, *OSMADS18* from rice and *BM3* from barley; another clade (19.) contains the gene pair *ZMM4* and *ZMM15* from maize, the rice gene *OSMADS14*, the barley gene *BM5* and the wheat gene *TAMADS11*; a third clade (20.) of *SQUA*-like genes comprises *ZAP1* and *ZMMADS3* from maize, *OSMADS15* from rice and *SBMADS2* from sorghum (Fig. 1).

Those genes in these 20 different clades (according to the numbering in this paragraph), which are from different species, are excellent candidates for being orthologs. (For pairs of closely related genes from the same species, maize, see below). In total, we can postulate that there were already at least 20 different MIKC-type genes present in the most recent common ancestor of maize and rice about 50 – 70 MYA (GOFF *et al.*, 2002), with some evidence that there were even more.

Phylogeny reconstructions III: the ancient tetraploidy of the Zea *genome*

In quite a number of cases the closest known relative of a MADS-box gene from maize is just another gene from the same species (Fig. 1). These young paralogs were obviously generated some time after the maize-rice split. A reasonable hypothesis suggests that these pairs of closely related maize genes were generated during the segmental allotetraploid event that produced the maize genome (GAUT and DOEBLEY, 1997). If so, the last common ancestor of these genes existed probably

either about 21 MYA (when the diploid species that later hybridized to a tetraploid maize progenitor separated) or 11 MYA (when the tetraploid maize precursor switched from tetrasomic to disomic inheritance) (GAUT and DOEBLEY, 1997).

In case of *ZAG2* and *ZMM1* (THEISEN *et al.*, 1995), *ZMM8* and *ZMM14* (CACHARRÓN *et al.*, 1999) and *ZAG3* and *ZAG5* (MENA *et al.*, 1995), the segmental allotetraploid hypothesis of gene pair origin has already been corroborated by demonstrating that the respective genes are located in duplicate regions of the maize genome. *ZMM18* and *ZMM29* (8.), however, are not in duplicate regions of the maize genome, and the respective loci originated probably in a gene duplication more recent than the tetraploidization event (MÜNSTER *et al.*, 2001).

Since the chromosomal map locations on the BNL map (http://burr.bio.bnl.gov/acemaz.html) were determined for most of the genes reported here (Table 1; Fig. 2), the segmental allotetraploid origin hypothesis could be tested for additional pairs of MIKC-type genes.

The *STMADS11*-like genes *ZMM19* and *ZMM26* are in duplicate regions in the maize genome (Fig. 3). However, the molecular markers in the respective regions are not in perfect colinear arrangement, suggesting that at least one inversion of a chromosomal segment has followed the allotetraploid event (Fig. 3). Stronger support for the allotetraploid origin of *ZMM19* and *ZMM26* may be obtained by chromosomal mapping at higher resolution.

More complex is the phylogenetic relationship in the case of the maize *AG*-like genes *ZAG1*, *ZMM2* and *ZMM23*. Although *ZAG1* and *ZMM2* (70% similarity, 61.1% identity on amino acid level) are also in duplicate regions of the maize genome (THEISEN *et al.*, 1995), these genes separated already roughly about 60 MYA and are thus much more ancient than the segmental allotetraploid event (GAUT and DOEBLEY, 1997; THEISEN *et al.*, 2000). This view is strongly supported by our phylogeny reconstructions (Fig. 1) which show that *ZAG1* is sister to a cluster of the putatively orthologous genes *OSMADS3* from rice and the *ZMM2/ZMM23* gene pair from maize (89.6% similarity, 88.4% identity on amino acid level), which could be identified in this analysis. Thus the separation of the most recent common ancestor of these genes from the *ZAG1* ancestor occurred probably before the maize-rice split, which was well before the segmental allotetraploid event which generated the maize genome.

Taking together, chromosomal mapping supports meanwhile a segmental allotetraploid origin for four pairs of MIKC-type MADS-box genes, but not for all maize genes which cluster together within phylogenetic trees.

Increase in the number of MIKC genes during evolution

By phylogeny reconstructions and analysis of exon-intron structures, MIKC-type genes can be subdivided into MIKC^c-type and MIKC*-type genes (HENSCHEL *et al.*, 2002). While MIKC^c-type genes are probably present in all green plants, MIKC*-type genes have been found so far only in a moss (*Physcomitrella*) and a lycophyte (*Lycopodium*) species, suggesting that these genes existed already in the most recent common ancestor of mosses and vascular plants (including lycophytes) about 450 MYA, but may have been lost during the evolution of higher vascular plants (tracheophytes except lycophytes) (HENSCHEL *et al.*, 2002).

The minimal estimate of 20 different MIKC^c-type MADS-box genes postualted for the most recent common ancestor of maize and rice (50–70 MYA) contrasts with the minimal estimate (11 different MIKC^c-type genes) for the most recent common ancestor of monocots and eudicots (about 200 MYA) and with at least 7 different MIKC^c-type genes reconstructed for the most recent common ancestor of extant gymnosperms and angiosperms about 300 MYA (BECKER *et al.*, 2000). There is an even stronger apparent contrast to the at least 2 different MIKC^c-type genes postulated for the most recent common ancestor of extant ferns and seed plants about 400 MYA (MÜNSTER *et al.*, 1997) and to the at least one MIKC^c-type gene (and one MIKC*-type gene) in the last common ancestor of vascular plants (lycophytes + ferns and their allies + seed plants) and mosses about 450 MYA (HENSCHEL *et al.*, 2002). This apparent increase in gene number may have several reasons, ranging from artefacts such as sampling bias for some lineages (e.g., limited sampling in non-flowering plants) to differential gene loss in some lineages, our inability to properly identify very ancient gene relationships, and a true increase in gene number by gene duplication, sequence divergence and fixation. A true increase in gene number is a quite likely scenario at least within some gene subfamilies, such as the *AGL2*- and *SQUA*-like genes of monocots compared to eudicots (Fig. 1), since most of the possibilities for errors mentioned above can almost be excluded here.

The observation that some MADS-box gene sub-

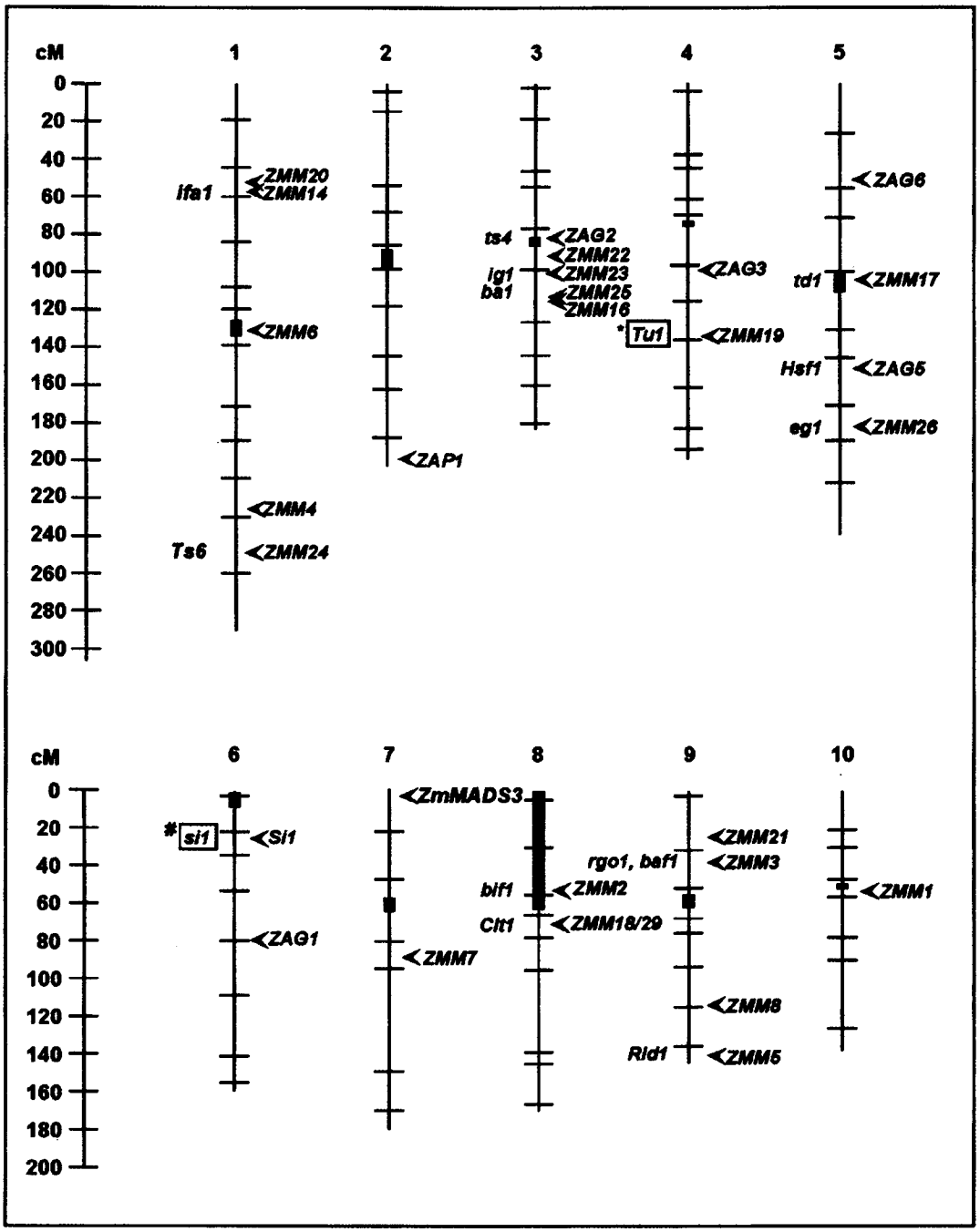

FIGURE 2 - Distribution of MADS-box genes in the maize genome. Numbers above vertical lines denote chromosome numbers. Arrow heads and names on the right of each chromosome indicate the map positions of MADS-box genes. On the left of each chromosome, the approximate map positions of some mutant candidate loci are indicated. Boxed mutants have been shown to correspond to the co-localized MADS-box genes (# *Si1*: AMBROSE *et al.*, 2000; **Tu1*: T.M., L.W., W.F., W.D., H.S. and G.T., unpublished data). The MADS-box genes have been mapped using the Brookhaven National Laboratory ressources (http://burr.bio.bnl.gov/acemaz.html), and for *ZmMADS3* by HEUER *et al.* (2001). Map positions of candidate mutants have been obtained from NEUFFER *et al.* (1997), except for *ifa1*, which was mapped by LAUDENCIA-CHINGCUANCO and HAKE (2002), and for *rgo1*, which is shown here according to the BIN map at MaizeDB (http://www.agron.missouri.edu). Abbreviations: *ba1= barren stalk1, baf1= barren stalk fastigiate1, bif1= barren inflorescence1, Clt1= Clumped tassel1,* cM= centimorgan, *eg1= expanded glumes1, Hsf1= Hairy sheath frayed1, ifa1= indeterminate floral apex1, ig1= indeterminate gametophyte1, rgo1= reverse germ orientation1, Rld1= Rolled leaf1, si1= silky1, td1= thick tassel dwarf1, ts4= tasselseed4, Ts6= Tasselseed6, Tu1= Tunicate1, ZAG= Zea mays AGAMOUS, ZAP= Zea mays APETALA, ZmMADS= Zea mays MADS, ZMM= Zea mays MADS.*

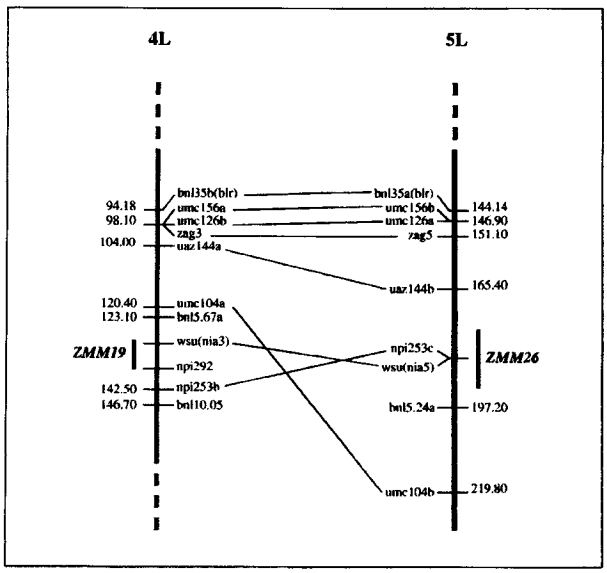

FIGURE 3 - Map locations of *ZMM19* and *ZMM26* on the long arms of maize chromosomes 4 and 5. The figure is based on the BNL96 map (http://www.agron.missouri.edu:80/cgi-bin/sybgw_mdb/mdb3/ Map/128615 and 128616, respectively). Only fragments of the chromosomes are shown. Numbers at some loci denote map units in centimorgans. Thin connecting lines denote paralogous (duplicate) molecular marker loci.

families, and possibly also the MADS-box gene family as a whole, probably significantly increased during plant evolution is in sharp contrast to observations made on many 'housekeeping genes', such as *ADH* (encoding alcohol dehydrogenase) or *rbc*S (encoding the small subunit of ribulose-1,5-bisphosphate carboxylase), although these genes are also nuclear encoded (CLEGG *et al.*, 1997). Since duplicate genes that do not confer an adaptive advantage are expected to rapidly erode into pseudogenes owing to the accumulation of mutations, gene duplicates which are retained over longer periods of evolutionary time must be positively selected based on some adaptive advantage conferred on the plant species (OBERHOLZER *et al.*, 2000). It is intruiguing to speculate, therefore, that a higher number of certain MADS-domain transcription factors was of selective advantage during the evolutionary process that ultimately led to extant flowering plants such as maize. The same may be true for some other genes encoding transcription factors, such as the Myb genes (OBERHOLZER *et al.*, 2000), but not for housekeeping genes providing basic enzymatic functions. Indeed, if there is more than one copy of e.g. *rbc*S in the genome of a plant, these are usually young rather than ancient paralogs without individually corresponding orthologs in more distantly related species, and there is no evidence for functional diversification of the different copies (CLEGG *et al.*, 1997).

On the other hand, especially developmental control genes with complex spatio-temporal expression patterns confered by a number of different cis-regulatory elements may be more often subject to subfunctionalization (with an partitioning of the ancestral function) rather than neofunctionalization (where at least one of the duplicates adopts a novel function) (FORCE *et al.*, 1999). So to what extent were only ancestral functions partitioned to several genes, and to what extent were genuinely novel functions established during grass and maize evolution? Detailed studies on the expression and function of the genes reported here will be required to answer these questions, and thus to assess the importance of the likely expansion of the MADS-box gene family during monocot evolution.

Identification of candidate genes

Determination of the map positions of the majority of the maize MIKC-type genes has also been used to find out whether they coincide with known mutant loci (Fig. 2). Up to now, only one MADS-box gene, *SILKY1* (a *DEF*-like gene with a class B floral homeotic function) has been published to be located at a classical maize mutant locus (AMBROSE *et al.*, 2000). Ironically, *SILKY1* was initially identified as a MADS-box gene by classical transposon tagging, not by a chromosomal mapping of a cDNA.

Quite a number of interesting mutant candidates for MADS-box genes have been discussed before (VEIT *et al.*, 1993), but none of them has been corroborated so far. This contrasts to species like *Arabidopsis*, *Antirrhinum* and even tomato, where several classical morphogenetic mutants have been shown to be affected in MADS-box genes (SOMMER *et al.*, 1990; YANOFSKY *et al.*, 1990; HUIJSER *et al.*, 1992; JACK *et al.*, 1992; MANDEL *et al.*, 1992; TRÖBNER *et al.*, 1992; BRADLEY *et al.*, 1993; GOTO and MEYEROWITZ, 1994; MAO *et al.*, 2000; VREBALOV *et al.*, 2002).

However, the considerable number of genes reported here (Table 1; Fig. 1) provides new chances for the molecular cloning of interesting mutant loci of maize. Some interesting candidates are depicted in Fig. 2. For example, the mapping data suggest that the *SlMADS11*-like gene *ZMM19* is located close to the *Tunicate1* (*Tu1*) locus. Meanwhile, several independent lines of evidence strongly support the view that *ZMM19* indeed represents the *TU1* gene (T.M., L.W., W.F., W.D., H.S. and G.T., unpub-

lished data). It will be interesting to see whether there are more classical genes represented by our collection of MADS-box gene cDNAs.

Maize MADS-box gene diversity beyond MIKC

This study focuses on the isolation and initial characterization of MIKC-type MADS-box genes from maize. However, there are certainly also other classes of MADS-box genes present in the maize genome, adding further to the complexity and diversity of the MADS-box gene family in maize. For example, there are transposon-like elements in the maize genome containing *AG*-like MADS-boxes that have not been found so far in any other plant species except maize relatives (FISCHER *et al.* 1995; MONTAG *et al.*, 1995, 1996). In contrast, *ZMM22* is a non-MIKC-type MADS-box gene which appeared in our screen for which a putative ortholog in rice could be identified (T.M. and W.F., unpublished data). This gene type thus exists at least since the maize – rice split about 50-70 MYA. One may also expect to find Type I (*SRF*-like) (ALVAREZ-BUYLLA *et al.*, 2000) MADS-box genes in the maize genome. However, the analysis of such 'exotic' MADS-box genes was beyond the scope of the endeavour outlined here.

Perspectives

The MIKC-type MADS-box genes reported here represent a cornucopia for further studies. For most genes, initial studies on the expression patterns have been carried out already by Northern or *in situ* hybridizations (L.W., J.C., H.S. and G.T., unpublished data). These revealed a diversity of patterns that has largely been conserved since the time when the eudicot and monocot lineages separated. However, some more recent changes in expression patterns (and thus very likely also function) became obvious. The respective genes may be interesting study objects to better understand the differential evolution of developmental genes after gene duplications, and the sub- and neofunctionalization of such genes (FORCE *et al.*, 1999). Some MADS-box genes have very defined expression patterns in space and time, e.g. in reproductive organs such as carpels and stamens, often with subtle differences for closely related genes. Such genes are ideal candidates for promoter studies aiming at figuring out the evolution of cis-regulatory elements responsible for tissue-specific expression patterns.

Further information about the functions of the genes reported here will be obtained by the analysis of classical mutants such as *Tu1*, transgenic studies and reverse genetics. For *ZMM6* and *ZMM8*, for example, changes in gene expression in transgenic maize plants resulted in specific morphological phenotypes within tassel and ears that were predicted from gene expression patterns (W.D., H.S. and G.T., unpublished data).

Finally, the genes reported here may have a considerable agronomic potential, as is becoming more and more obvious for the MADS-box gene family from other crop species as well (MEYEROWITZ, 1994; THEIßEN, 2000, 2001a, 2002). Changes in maize MADS-box gene functions, brought about e.g. by transgenic technology, or mutagenesis and marker assisted breeding, may be employed to optimize reproductive or vegetative traits of maize. Many maize MADS-box genes have very defined and diverse expression patterns, e.g. in the different plant organs. The promoters of these genes may be used to drive the expression of arbitrary genes of interest in almost any specific maize organ, including only carpels (e.g., *ZAG2*, *ZMM1*; SCHMIDT *et al.*, 1993; THEIßEN *et al.*, 1995); or in all reproductive organs, i.e. carpels and stamens (*ZAG1*, *ZMM2*; SCHMIDT *et al.*, 1993; CACHARRÓN *et al.*, 1999); or only in the nonreproductive organs of each spikelet, i.e. glumes, palea, lemma and lodicules (*ZMM15*; CACHARRÓN, 1998); or only in the upper, but not the lower floret of each spikelet (CACHARRÓN *et al.*, 1999). Note that within a spikelet, the expression patterns of *ZMM2* and *ZMM15* are perfectly complementary.

Hence we expect that the genes listed in Table 1 will keep a number of maize students in several labs busy for some time to come.

ACKNOWLEDGEMENTS - We are grateful to Ben Burr for his help with the chromosomal mapping, and to the ADIS core facility of the MPIZ for sequencing work. We thank two anonymous reviewers for helpful suggestions on an earlier draft of the manuscript. This work was supported in part by BMBF grant "Entwicklungskontrollgene zum gezielten Design von Nutzpflanzen" (Förderkennzeichen 0311378).

The authors dedicate this article to Prof. Francesco Salamini. One of us (H.S.) especially thanks Prof. Salamini for many fruitful and interesting discussions and his long standing friendship.

REFERENCES

ALVAREZ-BUYLLA E.R., S. PELAZ, S.J. LILJEGREN, S.E. GOLD, C. BURGEFF, G.S. DITTA, L.R. DE POUPLANA, L. MARTINEZ-CASTILLA, M.F. YANOFSKY, 2000 An ancestral MADS-box gene duplication occured before the divergence of plants and animals. Proc. Natl. Acad. Sci. USA **97:** 5328-5333.

AMBROSE B.A., D.R. LERNER, P. CICERI, C.M. PADILLA, M.F. YANOFSKY, R.J. SCHMIDT, 2000 Molecular and genetic analyses of the

Silky1 gene reveal conservation in floral organ specification between eudicots and monocots. Mol. Cell **5:** 569-579.

ANGENENT G.C., L. COLOMBO, 1996 Molecular control of ovule development. Trends Plant Sci. **1:** 228-232.

BECKER A., K.-U. WINTER, B. MEYER, H. SAEDLER, G. THEIßEN, 2000 MADS-box gene diversity in seed plants 300 million years ago. Mol. Biol. Evol. **17:** 1425-1434.

BECKER A., K. KAUFMANN, A. FREIALDENHOVEN, C. VINCENT, M.-A. LI, H. SAEDLER, G. THEIßEN, 2002 A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. Mol. Genet. Genomics **266:** 942-950.

BRADLEY D., R. CARPENTER, H. SOMMER, N. HARTLEY, E. COEN, 1993 Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the *plena* locus of Antirrhinum. Cell **72:** 85-95.

BURR B., F.A. BURR, K.H. THOMPSON, M.C. ALBERTSON, C.W. STUBER, 1988 Gene mapping with recombinant inbreds in maize. Genetics **118:** 519-526.

BURR B., F.A. BURR, 1991 Recombinant inbreds for molecular mapping in maize: theoretical and practical considerations. Trends Genet. **7:** 55-60.

CACHARRÓN J., 1998 MADS-Box-Gene in *Zea mays*: Vergleichende Expressionsuntersuchungen an Modellen paraloger und orthologer Genpaare. Ph.D. thesis, Mathematisch-Naturwissenschaftliche Fakultät der Universität zu Köln, Germany.

CACHARRÓN J., H. SAEDLER, G. THEIßEN, 1999 Expression of the MADS-box genes *ZMM8* and *ZMM14* during inflorescence development of *Zea mays* discriminates between the upper and the lower floret of each spikelet. Dev. Genes Evol. **209:** 411-420.

CHO S., S. JANG, S. CHAE, K.M. CHUNG, Y.-H. MOON, G. AN, S.K. JANG, 1999 Analysis of the C-terminal region of *Arabidopsis thaliana* APETALA1 as a transcription activation domain. Plant Mol. Biol. **40:** 419-429.

CLEGG M.T., M.P. CUMMINGS, M.L. DURBIN, 1997 The evolution of plant nuclear genes. Proc. Natl. Acad. Sci. USA **94:** 7791-7798.

CRANE P.R., E.M. FRIIS, K.R. PEDERSEN, 1995 The origin and early diversification of angiosperms. Nature **374:** 27-33.

DAYHOFF M.O., 1979 Atlas of protein sequences and structure. vol. 5, supplement 3, 1978. National Biomedical Research Foundation, Washington. DC.

DOYLE J.J., 1994 Evolution of a plant homeotic multigene family: towards connecting molecular systematics and molecular developmental genetics. Syst. Biol. **43:** 307-328.

EDWARDS M.D., C.W. STUBER, J.F. WENDEL, 1987 Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics **116:** 113-125.

EGEA-CORTINES M., H. SAEDLER, H. SOMMER, 1999 Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrinum majus*. EMBO J. **18:** 5370-5379.

FELSENSTEIN J., 1993 PHYLIP Phylogeny Inference Package 3.5. Department of Genetics, The University of Washington.

FISCHER A., N. BAUM, H. SAEDLER, G. THEIßEN, 1995 Chromosomal mapping of the MADS-box multigene family in *Zea mays* reveals dispersed distribution of allelic genes as well as transposed copies. Nucl. Acids Res. **23:** 1901-1911.

FORCE A., M. LYNCH, F.B. PICKETT, A. AMORES, Y.-I. YAN, J. POSTLETHWAIT, 1999 Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151:** 1531-1545.

FROHMAN M.A., M.K. DUSH, G.R. MARTIN, 1988 Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85:** 8998-9002.

GAUT B.S., J.F. DOEBLEY, 1997 DNA sequence evidence for the segmental allotetraploid origin of maize. Proc. Natl. Acad. Sci. USA **94:** 6809-6814.

GOFF S.A., *et al.*, 2002 A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science **296:** 92-100.

GOTO K., E.M. MEYEROWITZ, 1994. Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. Genes Dev. **8:** 1548-1560.

GRECO R., L. STAGI, L. COLOMBO, G.C. ANGENENT, M. SARI-GORLA, M.E. PÈ, 1997 MADS box genes expressed in developing inflorescences of rice and sorghum. Mol. Gen. Genet. **253:** 615-623.

GU Q., C. FERRÁNDIZ, M.F. YANOFSKY, R. MARTIENSSEN 1998 The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. Development **125:** 1509-1517.

HARTMANN U., S. HÖHMANN, K. NETTESHEIM, E. WISMAN, H. SAEDLER, P. HUIJSER, 2000 Molecular cloning of *SVP*: a negative regulator of the floral transition in *Arabidopsis*. Plant J. **21:** 351-360.

HASEBE M., J.A. BANKS, 1997 Evolution of MADS gene family in plants. pp. 179-197. *In:* Iwatsuki K., P.H. Raven (Eds.), Evolution and diversification of land plants, Chapter 8, Springer-Verlag Tokyo, Japan.

HENSCHEL K., R. KOFUJI, , M. HASEBE, , H. SAEDLER, , T. MÜNSTER, G. THEIßEN, 2002 Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens*. Mol. Biol. Evol. **19:** 801-814.

HEUER S., H. LÖRZ, T. DRESSELHAUS, 2000 The MADS box gene *ZmMADS2* is specifically expressed in maize pollen and during maize pollen tube growth. Sex. Plant Reprod. **13:** 21-27.

HEUER S., S. HANSEN, J. BANTIN, R. BRETTSCHNEIDER, E. KRANZ, H. LÖRZ, T. DRESSELHAUS, 2001 The maize MADS box gene *ZmMADS3* affects node number and spikelet development and is co-expressed with *ZmMADS1* during flower development, in egg cells, and early embryogenesis. Plant Physiol. **127:** 33-45.

HONMA T., K. GOTO, 2001 Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature **409:** 525-529.

HUANG H., M. TUDOR, C.A. WEISS, Y. HU, H. MA, 1995 The *Arabidopsis* MADS-box gene *AGL3* is widely expressed and encodes a sequence-specific DNA-binding protein. Plant. Mol. Biol. **28:** 549-567.

HUIJSER P., J. KLEIN, W.-E. LÖNNIG, H. MEIJER, H. SAEDLER, H. SOMMER, 1992 Bracteomania, an inflorescence anomaly, is

caused by the loss of function of the MADS-box gene *squamosa* in *Antirrhinum majus*. EMBO J. **11:** 1239-1249.

JACK T., L.L. BROCKMAN, E.M. MEYEROWITZ, 1992 The homeotic gene *APETALA3* of Arabidopsis thaliana encodes a MADS box and is expressed in petals and stamens. Cell **68:** 683-697.

JEON J.-S., S. JANG, S. LEE, J. NAM, C. KIM, S.-H. LEE, Y.-Y. CHUNG, S.-R. KIM, Y.H. LEE, Y.-G. CHO, G. AN, 2000 *leafy hull sterile1* is a homeotic mutation in a rice MADS box gene affecting rice flower development. Plant Cell **12:** 871-884.

KANG H.-G., G. AN, 1997 Isolation and characterization of a rice MADS box gene belonging to the *AGL2* gene family. Mol. Cells **7:** 45-51.

KANG H.-G., J.-S. JEON, S. LEE, G. AN, 1998 Identification of class B and class C floral organ identity genes from rice. Plant Mol. Biol. **38:** 1021-1029.

KROGAN N.T., N.W. ASHTON, 2000 Ancestry of plant MADS-box genes revealed by bryophyte (*Physcomitrella patens*) homologues. New Phytol. **147:** 505-517.

LAUDENCIA-CHINGCUANCO D., S. HAKE, 2002 The *indeterminate floral apex1* gene regulates meristem determinacy and identity in the maize inflorescence. Development **129:** 2629-2638.

LAWTON-RAUH A.L., E.R. ALVAREZ-BUYLLA, M.D. PURUGGANAN, 2000 Molecular evolution of flower development. Trends Ecol. Evol. **15:** 144-149.

LEE H., S.-S. SUH, E. PARK, E. CHO, J.H. AHN, S.-G. KIM, J.S. LEE, Y.M. KWON, I. LEE, 2000 The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in Arabidopsis. Genes Dev. **14:** 2366-2376.

LILJEGREN S.J., G.S. DITTA, Y. ESHED, B. SAVIDGE, J.L. BOWMAN, M.F. YANOFSKY, 2000 *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. Nature **404:** 766-770.

MA H., M.F. YANOFSKY, E.M. MEYEROWITZ, 1991 *AGL1-AGL6*, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. Genes Dev. **5:** 484-495.

MA H., C. DEPAMPHILIS, 2000 The ABCs of floral evolution. Cell **101:** 5-8.

MANDEL M.A., C. GUSTAFSON-BROWN, B. SAVIDGE, M.F. YANOFSKY, 1992 Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. Nature **360:** 273-277.

MAO L., D. BEGUM, H.-W. CHUANG, M.A. BUDIMAN, E.J. SZYMKOWIAK, E.E. IRISH, R.A. WING, 2000 *JOINTLESS* is a MADS-box gene controlling tomato flower abscission zone development. Nature **406:** 910-913.

MENA M., B.A. AMBROSE, R.B. MEELEY, S.P. BRIGGS, M.F. YANOFSKY, R.J. SCHMIDT, 1996 Diversification of C-function activity in maize flower development. Science **274:** 1537-1540.

MENA M., M.A. MANDEL, D.R. LERNER, M.F. YANOFSKY, R.J. SCHMIDT, 1995 A characterization of the MADS-box gene family in maize. Plant J. **8:** 845-854.

MEYEROWITZ E.M., 1994 The genetics of flower development. Sci. Am. **271(5):** 40-47.

MICHAELS S.D., R.M. AMASINO, 1999 *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell **11:** 949-956.

MONTAG K., F. SALAMINI, R.D. THOMPSON, 1995 *ZEMa*, a member of a novel group of MADS box genes, is alternatively spliced in maize endosperm. Nucl. Acids Res. **23:** 2168-2177.

MONTAG K., F. SALAMINI, R.D. THOMPSON, 1996 The ZEM2 family of maize MADS box genes possess features of transposable elements. Maydica **41:** 241-254.

MÜNSTER T., J. PAHNKE, A. DI ROSA, J.T. KIM, W. MARTIN, H. SAEDLER, G. THEIßEN, 1997 Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. Proc. Natl. Acad. Sci. USA **94:** 2415-2420.

MÜNSTER T., L.U. WINGEN, W. FAIGL, S. WERTH, H. SAEDLER, G. THEIßEN, 2001 Characterization of three *GLOBOSA*-like MADS-box genes from maize: evidence for ancient paralogy in one class of floral homeotic B-function genes of grasses. Gene **262:** 1-13.

NEUFFER M.G., E.H. COE, S.R. WESSLER, 1997 Mutants of maize. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

NIKLAS K.J., 2000 The evolution of plant body plans - a biomechanical perspective. Annals of Botany **85:** 411-438.

NG M., M. YANOFSKY, 2001 Function and evolution of the plant MADS-box gene family. Nature Rev. Gen. **2:** 186-195.

OBERHOLZER V., M.L. DURBIN, M.T. CLEGG, 2000 Comparative genomics of chalcone synthase and *Myb* genes in the grass family. Genes Genet. Syst. **75:** 1-16.

PELAZ S., G.S. DITTA, E. BAUMANN, E. WISMAN, M.F. YANOFSKY, 2000 B and C organ identity functions require *SEPALLATA* MADS-box genes. Nature **405:** 200-203.

PURUGGANAN M.D., S.D. ROUNSLEY, R.J. SCHMIDT, M.F. YANOFSKY, 1995 Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. Genetics **140:** 345-356.

RIECHMANN J.L., E.M. MEYEROWITZ, 1997 MADS domain proteins in plant development. Biol. Chem. **378:** 1079-1101.

RIECHMANN J.L., J. HEARD, G. MARTIN, L. REUBER, C.-Z. JIANG, J. KEDDIE, L. ADAM, O. PINEDA, O.J. RATCLIFFE, R.R. SAMAHA, R. CREELMAN, M. PILGRIM, P. BROUN, J.Z. ZHANG, D. GHANDEHARI, B.K. SHERMAN, G.-L. YU, 2000 *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science **290:** 2105-2110.

ROUNSLEY S.D., G.S. DITTA, M.F. YANOFSKY, 1995 Diverse roles for MADS box genes in Arabidopsis development. Plant Cell **7:** 1259-1269.

SAITOU N., M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406-425.

SAMBROOK J., E.F. FRITSCH, T. MANIATIS, 1989 Molecular cloning: A laboratory manual. Second ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

SAVARD L., P. LI, S.H. STRAUSS, M.W. CHASE, M. MICHAUD, J. BOUSQUET, 1994 Chloroplast and nuclear gene sequences indicate Late Pennsylvanian time for the last common ancestor of extant seed plants. Proc. Natl. Acad. Sci. USA **91:** 5163-5167.

SCHMIDT R.J., B. VEIT, M.A. MANDEL, M. MENA, S. HAKE, M.F. YANOFSKY, 1993 Identification and molecular characterization of

ZAG1, the maize homolog of the Arabidopsis floral homeotic gene *AGAMOUS*. Plant Cell **5:** 729-737.

SCHWARZ-SOMMER Z., P. HUIJSER, W. NACKEN, H. SAEDLER, H. SOMMER, 1990 Genetic control of flower development by homeotic genes in *Antirrhinum majus*. Science **250:** 931-936.

SHELDON C.C., D.T. ROUSE, E.J. FINNEGAN, W.J. PEACOCK, E.S. DENNIS, 2000 The molecular basis of vernalization: the central role of *FLOWERING LOCUS C* (*FLC*). Proc. Natl. Acad. Sci. USA **97:** 3753-3758.

SHORE P., A.D. SHARROCKS, 1995 The MADS-box family of transcription factors. Eur. J. Biochem. **229:** 1-13.

SOMMER H., J.-P. BELTRÁN, P. HUIJSER, H. PAPE, W.-E. LÖNNIG, H. SAEDLER, 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.

TANDRE K., V.A. ALBERT, A. SUNDAS, P. ENGSTRÖM, 1995 Conifer homologues to genes that control floral development in angiosperms. Plant Mol. Biol. **27:** 69-78.

THEIßEN G., H. SAEDLER, 1995 MADS-box genes in plant ontogeny and phylogeny: Haeckel's 'biogenetic law' revisited. Curr. Opin. Genet. Dev. **5:** 628-639.

THEIßEN G., T. STRATER, A. FISCHER, H. SAEDLER, 1995 Structural characterization, chromosomal localization and phylogenetic evaluation of two pairs of *AGAMOUS*-like MADS-box genes from maize. Gene **156:** 155-166.

THEIßEN G., J. KIM, H. SAEDLER, 1996 Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. J. Mol. Evol. **43:** 484-516.

THEIßEN G., H. SAEDLER, 1999 The golden decade of molecular floral development (1990-1999): a cheerful obituary. Dev. Genet. **25:** 181-193.

THEIßEN G., 2000 Shattering developments. Nature **404:** 711-713.

THEIßEN G., A. BECKER, A. DI ROSA, A. KANNO, J.T. KIM, T. MÜNSTER, K.-U. WINTER, H. SAEDLER, 2000 A short history of MADS-box genes in plants. Plant Mol. Biol. **42:** 115-149.

THEIßEN G., 2001a *SHATTERPROOF* oil seed rape: a *FRUITFULL* buisiness? MADS-box genes as tools for crop plant design. Biotech News Intern. **6:** 13-15.

THEIßEN G., 2001b Development of floral organ identity: stories from the MADS house. Curr. Opin. Plant Biol. **4:** 75-85.

THEIßEN G., H. SAEDLER, 2001 Floral quartets. Nature **409:** 469-471.

THEIßEN G., 2002 Key genes of crop domestication: molecular analyses. Progress in Botany 63, Springer-Verlag Berlin Heidelberg, pp. 189-203.

TRÖBNER W., L. RAMIREZ, P. MOTTE, I. HUE, P. HUIJSER, W.E. LÖNNIG, H. SAEDLER, H. SOMMER, Z. SCHWARZ-SOMMER, 1992 *GLOBOSA*: a homeotic gene which interacts with *DEFICIENS* in the control of *Antirrhinum* floral organogenesis. EMBO J. **11:** 4693-4704.

VEIT B., R.J. SCHMIDT, S. HAKE, M.F. YANOFSKY, 1993 Maize floral development: new genes and old mutants. Plant Cell **5:** 1205-1215.

VREBALOV J., D. RUEZINSKY, V. PADMANABHAN, R. WHITE, D. MEDRANO, R. DRAKE, W. SCHUCH, J. GIOVANNONI, 2002 A MADS-box gene necessary for fruit ripening at the tomato *Ripining-Inhibitor* (*Rin*) locus. Science **296:** 343-346.

WEIGEL D., E.M. MEYEROWITZ, 1994 The ABCs of floral homeotic genes. Cell **78:** 203-209.

WINTER K.-U., A. BECKER, T. MÜNSTER, J.T. KIM, H. SAEDLER, G. THEIßEN, 1999 MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. Proc. Natl. Acad. Sci. USA **96:** 7342-7347.

WOLFE K.H., M. GOUY, Y.-W. YANG, P.M. SHARP, W.-H. LI, 1989 Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc. Natl. Acad. Sci. USA **86:** 6201-6205.

YANOFSKY M.F., H. MA, J.L. BOWMAN, G.N. DREWS, K.A. FELDMAN, E.M. MEYEROWITZ, 1990 The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. Nature **346:** 35-39.

YU J., *et al.*, 2002 A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science **296:** 79-91.

ZHANG H., B.G. FORDE, 1998 An Arabidopsis MADS box gene that controls nutrient-induced changes in root architecture. Science **279:** 407-409.

ZHANG J., M. NEI, 1996 Evolution of Antennapedia-class homeobox genes. Genetics **142:** 295-303.