MAIZE MADS-BOX GENES GALORE


2 Max-Planck-Institut für Züchtungsforschung (MPIZ), Abteilung Molekulare Pflanzenzüchtung, 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 lead 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 allopolyploidy 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 resource 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; Evolution; 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; Thieben 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
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; Reichmann 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 SEPALATA1, SEPALATA2 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 Reichmann 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 (Michaelis 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 (Liegeisson et al., 2000), and FUL is required for the normal pattern of cell division, expansion and differentiation during morphogenesis of the siliques (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, ANRI, 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 (Tandré et al., 1995; Münster et al., 1997; Winter et al., 1999; Krokan and Ashton, 2000; Henkel 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; Sorgothem 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 (Meyrowitz, 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 Meyrowitz, 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 Sæddler, 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 (Erskine et al., 1987), and recombinant inbred (RB) lines TcXM 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 (Cacharron et al., 1995) 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 URGLOA. 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 were used purchased from MWG, Metabion or LifeTech.

Isolation of cDNAs by RACE

Partial cDNAs were isolated by 3’-RACE as generally described (Flurkman 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, cobs, anthers, male and female inflorescences, developing kernels, and pollen. In some cases, upstream sequences overlapping with the 5’-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 TcXM and CoXtX (Burr and Burr, 1988; Burr and Burr, 1991) as generally described previously (Fischler 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://burl.bio.bnl.gov/aeomaz.html).

Sequence Alignments and Construction of Phylogenetic Trees

Accession numbers of all sequences used are available at the MADS homepage (http://www.mipz-koeln.mpg.de/mads). Multiple alignments of conceptual amino acid sequences were generated by using thePILEUP 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 3.5.5c, 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 (Müller et al., 1997).
<table>
<thead>
<tr>
<th>Gene</th>
<th>Subfamily</th>
<th>Marker designation</th>
<th>Map position</th>
<th>Accession No.</th>
<th>Reference</th>
<th>Defining clone</th>
</tr>
</thead>
<tbody>
<tr>
<td>SILK</td>
<td>DEF</td>
<td>n.p.</td>
<td>6020</td>
<td>AF181479</td>
<td>AMBROSE et al., 2000</td>
<td>/</td>
</tr>
<tr>
<td>ZAG2</td>
<td>AG</td>
<td>ucsd78A</td>
<td>6080</td>
<td>L18924</td>
<td>SCHMIDT et al., 1995</td>
<td>pSW205</td>
</tr>
<tr>
<td>ZAG3</td>
<td>AG</td>
<td>ucsd81A</td>
<td>3084</td>
<td>L18925</td>
<td>SCHMIDT et al., 1995</td>
<td>pSW611</td>
</tr>
<tr>
<td>ZAG4</td>
<td>AG</td>
<td>ucsd79C</td>
<td>4089</td>
<td>L46397</td>
<td>MESS et al., 1995</td>
<td>/</td>
</tr>
<tr>
<td>ZAG5</td>
<td>AG</td>
<td>ucsd72E</td>
<td>5115</td>
<td>L46398</td>
<td>MESS et al., 1995</td>
<td>/</td>
</tr>
<tr>
<td>ZAP1</td>
<td>SQUA</td>
<td>ucsd106A</td>
<td>2200</td>
<td>L46400</td>
<td>MESS et al., 1995</td>
<td>pSW95</td>
</tr>
<tr>
<td>ZMMA1</td>
<td>AG</td>
<td>mpi23A</td>
<td>10054</td>
<td>AJ130530</td>
<td>THEISEN et al., 1995</td>
<td>pWF074</td>
</tr>
<tr>
<td>ZMMA2</td>
<td>AG</td>
<td>mpi24</td>
<td>8056</td>
<td>AJ130551</td>
<td>THEISEN et al., 1995</td>
<td>pSW125</td>
</tr>
<tr>
<td>ZMMA3</td>
<td>AG</td>
<td>mpi25</td>
<td>9034</td>
<td>Y93901</td>
<td>FINCHER et al., 1995</td>
<td>pWF4012</td>
</tr>
<tr>
<td>ZMMA4</td>
<td>SQUA</td>
<td>mpi22A</td>
<td>11204</td>
<td>AJ130541</td>
<td>FINCHER et al., 1995</td>
<td>pSW202</td>
</tr>
<tr>
<td>ZMMA5</td>
<td>TM5</td>
<td>kws2</td>
<td>9140</td>
<td>AJ130542</td>
<td>FINCHER et al., 1995</td>
<td>/</td>
</tr>
<tr>
<td>ZMMA6</td>
<td>AG</td>
<td>mpi23A</td>
<td>11131</td>
<td>AJ130562</td>
<td>FINCHER et al., 1995</td>
<td>pSW290</td>
</tr>
<tr>
<td>ZMMA7</td>
<td>AG</td>
<td>mpi27</td>
<td>71095</td>
<td>Y93902</td>
<td>FINCHER et al., 1995</td>
<td>/</td>
</tr>
<tr>
<td>ZMMA8</td>
<td>AG</td>
<td>mpi28</td>
<td>9115</td>
<td>Y93903</td>
<td>CACHARRON et al., 1999</td>
<td>pSW24</td>
</tr>
<tr>
<td>ZMMA14</td>
<td>AG</td>
<td>mpi43</td>
<td>18055</td>
<td>AJ005358</td>
<td>CACHARRON et al., 1999</td>
<td>pWF1552</td>
</tr>
<tr>
<td>ZMMA15</td>
<td>SQUA</td>
<td>n.d.</td>
<td>n.d.</td>
<td>AJ130532</td>
<td>THIS WORK</td>
<td>pSW31</td>
</tr>
<tr>
<td>ZMMA16</td>
<td>GLO</td>
<td>mpi39</td>
<td>31110</td>
<td>AJ292959</td>
<td>MUNSTER et al., 2001</td>
<td>pSW436</td>
</tr>
<tr>
<td>ZMMA17</td>
<td>GGM13</td>
<td>mpi40</td>
<td>55107</td>
<td>A1271208</td>
<td>BÖCKER et al., 2002</td>
<td>pSW450</td>
</tr>
<tr>
<td>ZMMA18</td>
<td>GLO</td>
<td>mpi41</td>
<td>80709</td>
<td>AJ292960</td>
<td>MUNSTER et al., 2001</td>
<td>pSW59</td>
</tr>
<tr>
<td>ZMMA19</td>
<td>SIMADS11</td>
<td>mpi42</td>
<td>41145</td>
<td>AJ130533</td>
<td>THIS WORK</td>
<td>pWF4123</td>
</tr>
<tr>
<td>ZMMA20</td>
<td>SIMADS11</td>
<td>mpi44</td>
<td>8053</td>
<td>AJ130534</td>
<td>THIS WORK</td>
<td>pWF748</td>
</tr>
<tr>
<td>ZMMA21</td>
<td>SIMADS11</td>
<td>mpi45</td>
<td>90023</td>
<td>AJ130535</td>
<td>THIS WORK</td>
<td>pWF4105</td>
</tr>
<tr>
<td>ZMMA22</td>
<td>AG</td>
<td>kws6</td>
<td>31095</td>
<td>AJ130557</td>
<td>THIS WORK</td>
<td>pSW128</td>
</tr>
<tr>
<td>ZMMA24</td>
<td>AG</td>
<td>mpi7</td>
<td>11245</td>
<td>AJ130568</td>
<td>THIS WORK</td>
<td>pWF1557</td>
</tr>
<tr>
<td>ZMMA25</td>
<td>AG</td>
<td>mpi17</td>
<td>31099</td>
<td>AJ130576</td>
<td>THIS WORK</td>
<td>pSW71</td>
</tr>
<tr>
<td>ZMMA26</td>
<td>SIMADS11</td>
<td>mpi48</td>
<td>52115</td>
<td>AJ130593</td>
<td>THIS WORK</td>
<td>pBRACE9-20</td>
</tr>
<tr>
<td>ZMMA27</td>
<td>AG</td>
<td>n.d.</td>
<td>n.d.</td>
<td>AJ130609</td>
<td>THIS WORK</td>
<td>pBRACE9-20</td>
</tr>
<tr>
<td>ZMMA28</td>
<td>SQUA</td>
<td>n.d.</td>
<td>n.d.</td>
<td>AJ130595</td>
<td>THIS WORK</td>
<td>pBRACE9-20</td>
</tr>
<tr>
<td>ZMMA29</td>
<td>GLO</td>
<td>mpi11</td>
<td>80709</td>
<td>AJ292961</td>
<td>MUNSTER et al., 2001</td>
<td>pSW425</td>
</tr>
<tr>
<td>ZMMA31</td>
<td>AG</td>
<td>mpi1</td>
<td>n.d.</td>
<td>AJ130604</td>
<td>THIS WORK</td>
<td>pWF1243</td>
</tr>
<tr>
<td>ZMMA32</td>
<td>AG</td>
<td>mpi7</td>
<td>n.d.</td>
<td>AJ1302149</td>
<td>HEUER et al., 2000</td>
<td>pWF1243</td>
</tr>
<tr>
<td>ZMMA33</td>
<td>SQUA</td>
<td>n.p.</td>
<td>80709</td>
<td>AJ1302150</td>
<td>HEUER et al., 2000</td>
<td>pWF1243</td>
</tr>
</tbody>
</table>

---

**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 (CACHARRON 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 omitted 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 STIMADS11-like genes (Becker et al., 2000), the B_sister (B3) 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; Sward 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_sister (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 MIBK-type MADS-domain proteins known. The tree was constructed using the "170-domain" sequences (according to Thines 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 (Thines et al., 1996; Moseman 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; Yc et al., 2002), putative orthologs of the maize B, 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 OSMAD52 and ZMM16 (7.), and the other one OSMAD54, ZMM18 and ZMM20 (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 OSMAD52 from rice; and another one (10.) comprising ZMM20 from maize, OSMAD47 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 OSMAD53 from rice. There is another clade (12.) containing ZMM2 and ZMM23 from maize and OSMAD3 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 Hyacinthinus 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, OSMAD54 (also known as OSMAD57; Greco et al., 1997; Kang and Ax, 1997). Another clade (14.) comprises the maize genes ZMM7 and ZMM27, the rice gene OSMAD24 (also known as OSMAD8; Greco et al., 1997; Kang and Ax, 1997), and the sorghum gene SBMADS1. In other clades there are ZMM8 and ZMM14 from maize, OSMAD51 from rice, and BM7 from barley (15); ZMM3 from maize and OSMAD5 from rice (16); ZMM24 and ZMM31 from maize and OSMAD34 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 OSMAD51 (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, neo-functionalization and sub-functionalization 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, OSMAD58 from rice and BM3 from barley; another clade (19.) contains the gene pair ZMM4 and ZMM15 from maize, the rice gene OSMAD14, the barley gene BM5 and the wheat gene TAMA11; a third clade (20.) of SQUA-like genes comprises ZAPI and ZMMAD53 from maize, OSMAD55 from rice and SBSMAD52 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 ancestors 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 (Theissen 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/aceamz.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 (Theissen 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; Theissen 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*-type and MIKC-type genes (Henschel et al., 2002). While MIKC*-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*-type MADS-box genes postulated for the most recent common ancestor of maize and rice (50–70 MYA) contrasts with the minimal estimate (11 different MIKC*-type genes) for the most recent common ancestor of monocots and eudicots (about 200 MYA) and with at least 7 different MIKC*-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*-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*-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 (S1: AMBRION 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 resources (http://brr.bio.bnl.gov/maizez.html), and for ZmMADS3 by HEBER et al. (2001). Map positions of candidate mutants have been obtained from NEFFER et al. (1997), except for if31, which was mapped by LAIFENZA-CHENGELIAN and HALE (2002), and for rgof, which is shown here according to the BIN map at MaizeDB (http://www.agron.missouri.edu).

Abbreviations: bst = barren stalk, baf = barren inflorescence, Clt = Clumped tassel, cm = centimorgan, eg1 = expanded glumes1, Hg1 = Hairy sheath, if31 = indeterminate floral apex, ig1 = indeterminate gametophyte1, rgof = reverse germ orientation1, Rld1 = Rolled leaf1, s11 = silky1, tdf = thick tassel dwarf1, tvi = tasselvey1, Tvi6 = Tasselvein6, Tu1 = Tumicat1. ZAG = Zea mays AGAMOUS, ZAP = Zea mays APETALA, ZmMADS = Zea mays MADS, ZMM = Zea mays MADS.
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 rbcS (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 intriguing 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. rbcS in the genome of a plant, these are usually young rather than ancient paralogs without individually corre-

sponding 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 conferred 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 (Vitt 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; Tröner 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 STMADS31-like gene ZMM19 is located close to the Tunicate1 (Tun1) locus. Meanwhile, several independent lines of evidence strongly support the view that ZMM19 indeed represents the Tun 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 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 endeavours 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 *Tnt1*, 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; Theisen, 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; Theisen et al., 1995); or in all reproductive organs, i.e. carpels and stamens (ZAG1, ZMM2; Schmidt et al., 1993; Cacharron et al., 1999); or only in the nonreproductive organs of each spikelet, i.e. glumes, palea, lemma and lodicules (ZMM15; Cacharron, 1998); or only in the upper, but not the lower floret of each spikelet (Cacharron 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 ADS 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 BMF grant “Entwicklungskontrollgene zum gezielten Design von Nutzpflanzen” (Förderkennzeichen 0311789). The authors dedicate this article to Prof. Francesco Salamini. One of us (U.S.) especially thanks Prof. Salamini for many fruitful and interesting discussions and his long standing friendship.

**REFERENCES**


Eisen-Collin D., H. Sather, I. Sommer, 1999 Tertiary complex formation between the MADS-box proteins SQUAMOSA DEFICIENS and GLORIA is involved in the control of floral architecture in Arabidopsis thaliana. EMBO J. 18: 5370-5379.


Huendl P., J. Kain, W-E. Linng, H. Meier, H. Sather, H. Sother, 1992 Retrometam, an inflorescence anomaly, is
caused by the loss of function of the MADS-box gene squamosa in Antirrhinum majus. EMBO J. 11: 1239-1249.


MONTAG K., F. SALAMINE, R.D. THOMPSON, 1995 ZEM1, a member of a novel group of MADS box genes, is alternatively spliced in maize endosperm. Nucl. Acids Res. 23: 2168-2177.


SCHMIDT R.J., M. VON, 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.


