Conservation and Evolution in and among SRF- and MEF2-Type MADS Domains and Their Binding Sites

Wenwu Wu,[1,2] Xiaotai Huang,[1,2] Jian Cheng,[3] Zhenggang Li,[3] Stefan de Folter,[4] Zhuoran Huang,[1,2] Xiaoqian Jiang,[1,2] Hongxia Pang,[1,2] and Shiheng Tao*[,1,2]

[1]College of Life Science, Northwest A&F University, Yangling, Shaanxi, China

[2]Bioinformatics Center, Northwest A&F University, Yangling, Shaanxi, China

[3]College of Science, Northwest A&F University, Yangling, Shaanxi, China

[4]Laboratorio Nacional de Genómica para la Biodiversidad (LANGEBIO), Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN), Irapuato, Guanajuato, Mexico

***Corresponding author:** E-mail: shihengt@nwsuaf.edu.cn.

Associate editor: Douglas Crawford

Research article

Abstract

Serum response factor (SRF) and myocyte enhancer factor 2 (MEF2) represent two types of members of the MCM1, AGAMOUS, DEFICIENS, and SRF (MADS)-box transcription factor family present in animals and fungi. Each type has distinct biological functions, which are reflected by the distinct specificities of the proteins bound to their cognate DNA-binding sites and activated by their respective cofactors. However, little is known about the evolution of MADS domains and their DNA-binding sites. Here, we report on the conservation and evolution of the two types of MADS domains with their cognate DNA-binding sites by using phylogenetic analyses. First, there are great similarities between the two types of proteins with amino acid positions highly conserved, which are critical for binding to the DNA sequence and for the maintenance of the 3D structure. Second, in contrast to MEF2-type MADS domains, distinct conserved residues are present at some positions in SRF-type MADS domains, determining specificity and the configuration of the MADS domain bound to DNA sequences. Furthermore, the ancestor sequence of SRF- and MEF2-type MADS domains is more similar to MEF2-type MADS domains than to SRF-type MADS domains. In the case of DNA-binding sites, the MEF2 site has a T-rich core in one DNA sequence and an A-rich core in the reverse sequence as compared with the SRF site, no matter whether where either A or T is present in the two complementary sequences. In addition, comparing SRF sites in the human and the mouse genomes reveals that the evolution rate of CArG-boxes is faster in mouse than in human. Moreover, interestingly, a CArG-like sequence, which is probably functionless, could potentially mutate to a functional CArG-box that can be bound by SRF and vice versa. Together, these results significantly improve our knowledge on the conservation and evolution of the MADS domains and their binding sites to date and provide new insights to investigate the MADS family, which is not only on evolution of MADS factors but also on evolution of their binding sites and even on coevolution of MADS factors with their binding sites.

Key words: SRF, MEF2, MADS, CArG box, conservation, evolution.

Introduction

The MCM1, AGAMOUS, DEFICIENS, and SRF (MADS)-box genes encode for a family of highly conserved transcriptional regulators involved in a diverse range of important biological functions. Among these are embryogenesis and flower and root development in plants (Ng and Yanofsky 2001; Fornara et al. 2003; Thakare et al. 2008; Adamczyk and Fernandez 2009; Zheng et al. 2009), immediate early gene expression and muscle differentiation and maintenance in animals (Treisman and Ammerer 1992; Sun et al. 2006; Potthoff and Olson 2007; Flavell et al. 2008), and cell–type-specific transcription and pheromone response in yeast (Herskowitz 1989; Treisman and Ammerer 1992; Shore and Sharrocks 1995). The family name itself is derived from the abbreviation of its earliest founder members (Schwarz-Sommer et al. 1990), MCM1 from yeast (Passmore et al. 1989), AGAMOUS from *Arabidopsis thaliana* (Yanofsky et al. 1990), DEFICIENS from *Antirrhinum majus* (Sommer et al. 1990), and serum response factor (SRF) from human (Norman et al. 1988), all of which are characterized by the MADS domain (56–60 amino acids) in the N terminal region responsible for DNA binding, protein dimerization, and cofactor interactions (Pellegrini et al. 1995; Shore and Sharrocks 1995; Ng and Yanofsky 2001; Immink et al. 2002; Messenguy and Dubois 2003).

Phylogenetic analysis has shown that the MADS domain family can be divided into two main lineages, called type I and type II (Alvarez-Buylla et al. 2000). Because both the type I and type II genes are present in plants, animals, and fungi, it suggests that a duplication event has occurred in the MADS-box lineage before the divergence of plants from fungi and animals (Alvarez-Buylla et al. 2000). In plants, type II proteins comprised an MADS (M) domain, an intervening (I) domain, a keratin-like (K) domain, and a C-terminal (C) region and, therefore, also called MIKC-type proteins (Munster et al. 1997; Kaufmann et al.

2005). In contrast to type II proteins, type I proteins lack a clear K domain and are sometimes called M-type proteins (Kofuji et al. 2003; Parenicova et al. 2003). In the case of animals and fungi, type I and type II proteins are referred to as SRF-like (e.g., SRF, ARG80, and MCM1) and MEF2-like proteins, respectively. In addition, except for the MADS domain, they also contain an additional conserved region, the SRF, ARG80, and MCM1 (SAM) domain, and the MEF2 domain, respectively (Shore and Sharrocks 1995; Alvarez-Buylla et al. 2000). Although these MADS domain proteins bind to A/T-rich DNA sequences in common (de Folter and Angenent 2006), however, there are distinct consensus sequences: SRF binds as a homodimer specifically to a 10-bp consensus CC(A/T)$_6$GG, called the CArG-box or SRF site (Pollock and Treisman 1990), whereas MEF2 proteins bind as homo- and heterodimer to another 10-bp consensus CTA(A/T)$_4$TAG, called MEF2 site (Pollock and Treisman 1991; Andres et al. 1995). Furthermore, when SRF binds to a CArG-box, it causes a dramatic DNA bending (~72°), whereas when MEF2 binds to a MEF2 site, it induces a minimal DNA bending (~17°) (Pellegrini et al. 1995; West et al. 1997; Santelli and Richmond 2000). Little is known, however, about the evolution of the MADS domains of MEF2 and SRF with their cognate DNA-binding sites.

Here, based on phylogenetic analyses, we report results on the conservation and evolution of the MADS domains of SRF and MEF2 factors, and SRF-binding sites in gene regulatory regions, particularly the binding positions in MADS domains contacting with their cognate DNA-binding positions. In addition, we infer the ancestor of SRF- and MEF2-type MADS domains to try to reveal the process of MADS domain evolution. Finally, we describe the generation and disruption of functional CArG-boxes bound by SRF. Together, these results provide to date the most comprehensive view of the evolution of MADS domains and their binding sites in animals and fungi.

Materials and Methods

Data Collection

Orthologous and paralogous proteins of SRF- and MEF2-type factors were retrieved from The National Center for Biotechnology Information (NCBI) (Wheeler et al. 2007). A Perl script was written to capture the MADS domain regions of these transcription factors (58 amino acids as MADS domain length was used) (see supplementary material S1, Supplementary Material online). On the other hand, hundreds of SRF-binding sites (see supplementary material S2, Supplementary Material online) including positions with respect to transcription start sites (TSSs) of the target genes from the mouse and human genomes were collected from the literature (Miano 2003; Philippar et al. 2004; Selvaraj and Prywes 2004; Zhang et al. 2005; Balza and Misra 2006; Sun et al. 2006; Cooper et al. 2007; Petit et al. 2008). These binding sites are all verified in vivo by experimental approaches to regulate the corresponding SRF target gene transcription and expression. For example, Sun et al. (2006) employed at least two of several validations including luciferase reporter, gel shift, chromatin immunoprecipitation (ChIP), and mRNA expression following RNAi knockdown of SRF to identify 60 CArG boxes exerting influence on mouse gene regulation. Another example is that Cooper et al. (2007) employed ChIP-on-chip in combination with quantitative polymerase chain reaction to identify 146 SRF-binding sites in the human genome to regulate gene expression. Subsequently, the promoters of these genes containing SRF-binding sites were retrieved using transcript IDs (SRF target genes) against the Database of Transcription Start Sites (DBTSS, version 6.0) (Wakaguri et al. 2008) and then a Perl script was written to capture 80 nucleotides on each side of SRF-binding sites using CArG sequences and their corresponding positions to TSS. Finally, we employed these 170 nucleotide length sequences including SRF-binding site in the centre to retrieve homologous sites from other three mammals, that is, dog, cow, and human (when the searched sequence was from mouse) or mouse (when the searched sequence was from human) against the corresponding reference genomic sequences (refseq_genomic) by BLAST searches. The *E*-value cutoff for the searches was initially set at 1×10^{-10}, where hits with higher *E*-values were evaluated manually whether these sequences were homologous. Among these retrieved homologous sites, most are of the same CArG nucleotide composition in mouse and human, however, we just collected the homologous SRF sites that had one or more substitutions in mouse or in human (see table 2).In the case of MEF2-binding sites, we collected over 200 binding sites (see supplementary material S3, Supplementary Material online) in vivo from the literature (Andres et al. 1995). In their research, Andres et al. selected its optimal DNA targets from a library of degenerate oligonucleotides using anti-MEF2A, anti-MEF2C, and anti-MEF2D antibodies combined with Electrophoretic Mobility Shift Assay to identify DNA-binding sites in skeletal muscle, heart, and brain. These binding sites were then used as the representation of MEF2 sites.

Logos of SRF-Type, MEF2-Type MADS Factors, and their DNA-Binding Sites

Logos of the MADS domains of SRF- and MEF2-type factors were generated using WebLogo (using default parameters) online (Crooks et al. 2004). Logos of their DNA-binding sites were generated using enoLOGOS online (Workman et al. 2005). Bits in *Y* axis (see figs. 1 and 4) represent the amount of information content (IC) at each position of the sequences. Given the strong effect of background frequencies of nucleotides on IC (Stormo 2000; Workman et al. 2005; Haeseleer 2006), guanine-cytosine content of the corresponding genomes was used as background frequency to calculate IC at each position of these DNA-binding sites as below:

$$\mathrm{IC}(i) = \sum_{b=\mathrm{A}}^{\mathrm{A,T,G,C}} p(b,i) \ln \frac{p(b,i)}{p_{\mathrm{genome}}(b)} \,(\mathit{bits\ per\ position}),$$

where $p(b,i)$ is the frequency of base b at position i in the DNA-binding sites, whereas $p_{\mathrm{genome}}(b)$ is the frequency of base b in the whole genome.

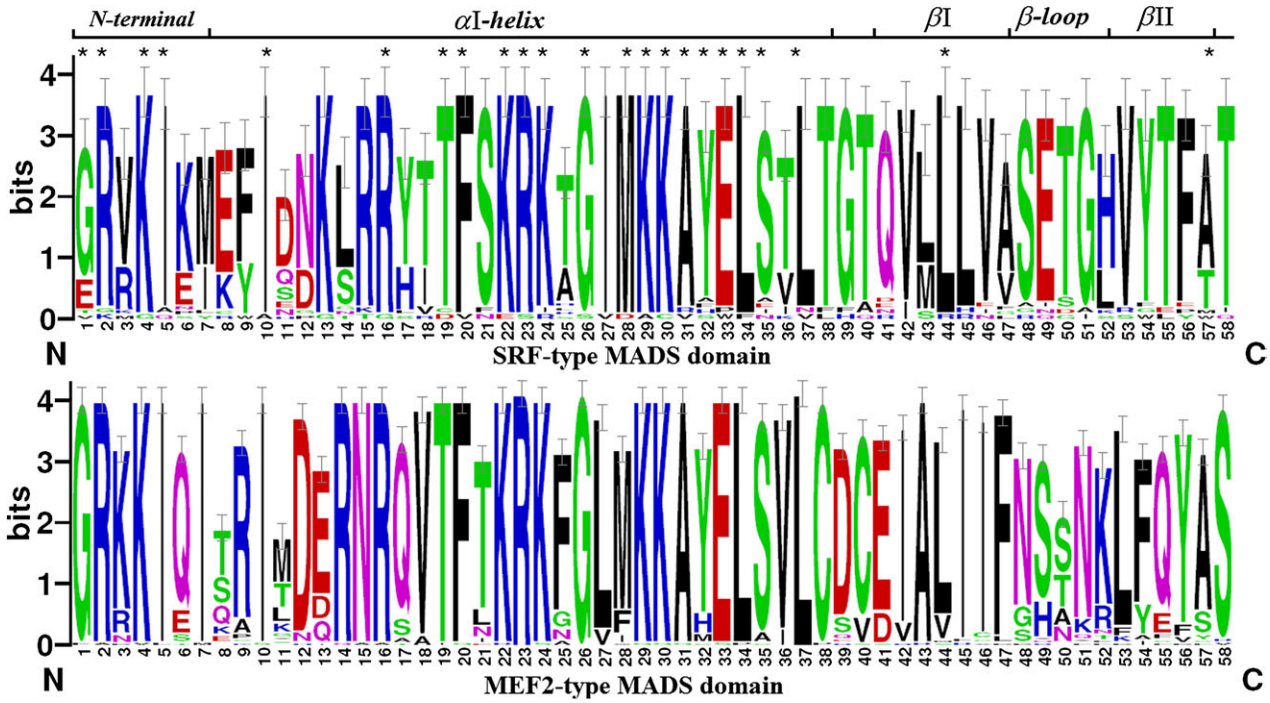

FIG. 1 Logos of the two types of MADS domains for DNA binding. The MADS domain is mainly responsible for DNA binding and protein dimerization, and the secondary protein structure is shown on the top of the two Logos. Sequence Logos of the MEF2- and SRF-type MADS domains show not only high similarities but also clear differences in preferred amino acid composition across the MADS domain. Invariant and highly conserved amino acid residues between the two type MADS domains are indicated by asterisk. Amino acids are colored according to their chemical properties: Polar amino acids (G, S, T, Y, C, Q, N) are green, basic (K, R, H) blue, acidic (D, E) red, and hydrophobic (A, V, L, I, P, W, F, M) amino acids are black. The two Logos were generated from 65 MEF2-type MADS domains compared with 33 SRF-type MADS domains.

Types of CArG-Boxes

Functional CArG-boxes were collected from literature (see Data Collection) and were classified based on the nucleotide composition across the 10-bp nucleotides, and the same nucleotide composition in a CArG-box was assigned to the same type of CArG-box. In other words, a CArG type comprises the same nucleotide composition across the 10-bp CArG sequences. For example, CCAAATTTGG and CCAAATTTGG were assigned to the same CArG type, whereas CCAAAAATGG was assigned to another CArG type. To test the difference of CArG-boxes between human and mouse, Z-test was used as following: $Z = \frac{P_1 - P_2}{\sqrt{P_0(1-P_0)(1/N_1 + 1/N_2)}}$, where P_1 and P_2 represent the ratios of CArG-type number to CArG box number in human and in mouse, respectively; N_1 and N_2 represent the numbers of CArG boxes in human and in mouse, respectively; and P_0 represents the ratio of the total number of CArG types to the total number of CArG boxes in the two species.

Inference of the Ancestor of SRF- and MEF2-Type MADS Domain Regions

When inferring the ancestor of SRF- and MEF2-type MADS domains, we used the software named "ANCESTOR" that is based on the distance-based Bayesian method (Zhang and Nei 1997). Before the inference, we used the phylogenetic relationships from TreeFam (Tree families database) (Li et al. 2006) corrected by the constructed tree of MADS factors in animals and fungi generated using ClustalX2 (Larkin et al. 2007) as the known phylogenetic relationships. Then, we used ANCESTOR (employing default parameters) with this phylogenetic relationships to infer the ancestor of SRF-type MADS domains and the ancestor of MEF2-type MADS domains in eukaryotes, respectively. Last, combined with the MADS domains (here including MADS-like domains) of protista, bacteria, and phages, we inferred the common ancestor of SRF- and MEF2-type MADS domains. TreeView X (Page 1996) was used to open the files of constructed trees and TreeGraph 2 (Muller J and Muller K 2004) was used to make some modifications.

Results and Discussion

More than 100 SRF- and MEF2-type MADS factors from animals, fungi, and protista were collected from NCBI and their MADS domains were extracted using a Perl script. From the literature (Miano 2003; Philippar et al. 2004; Selvaraj and Prywes 2004; Zhang et al. 2005; Balza and Misra 2006; Sun et al. 2006; Cooper et al. 2007; Petit et al. 2008), over 400 SRF-binding sites, which were experimentally verified functional in vivo, were found in the mouse and human genomes, in addition, over 200 MEF2-binding sites were found in the mouse (Andres et al. 1995).

Conservation and Evolution of SRF- and MEF2-Type MADS Domains

Despite the absence of their natural C-terminal, the MADS domains of SRF and MEF2 remain sufficient to mediate the correct DNA-binding specificity (Nurrish and Treisman 1995; West et al. 1997; Santelli and Richmond 2000). This leads us to investigate the conservation and evolution of SRF- and MEF2-type MADS domains compared with their respective consensus binding sites.

We generated Sequence Logos (fig. 1) of the two types of MADS domains using WebLogo (Crooks et al. 2004) to examine the conservation and evolution in and between SRF- and MEF2-type MADS domains. There are 17 positions, that is, 2, 4, 5, 10, 16, 19, 20, 22, 23, 24, 26, 29, 30, 31, 33, 34, and 37 (for convenience, the first residue of the MADS domain is designated position 1), almost totally conserved and some other positions are highly conserved, such as 1, 28, 32, 35, 44, and 57 in both the SRF- and MEF2-type MADS domains. In contrast, there are still some positions with profoundly different amino acid properties in the two types, such as 13, 14, 15, 41, 49, 55, and 56. The remaining positions are partly conserved or variable but with similar amino acid properties between the two types of MADS domains. In addition, the analysis of mutual information (see supplementary material S4, Supplementary Material online) on MADS domains generated from enoLOGOS (Workman et al. 2005) revealed that there are covariations between positions 3, 6, 8, 11, 13, 17, and 25, among which, position 13 is very important in determining the bending of the DNA upon binding of the MADS domain proteins (West et al. 1997). In summary, positions from 1 to 37, that is, N-terminal and αI-helix, have great similar conserved amino acids between SRF- and MEF2-type MADS domains, whereas positions from 38 to 58, that is, β-strands, have different conserved amino acids between SRF- and MEF2-type MADS domains. The explanation for the similar conserved positions (1–37) is that all protein-DNA contacts of the SRF- and MEF2-type MADS domains are common confined to the N-terminal extension and αI-helix except for β-loop of SRF to DNA (Pellegrini et al. 1995; Santelli and Richmond 2000). Whereas the explanation for different conserved positions (38–58) is that SRF and MEF2 proteins have different dimerization patterns (West and Sharrocks 1999; Huang et al. 2000; Santelli and Richmond 2000), which need the specific contribution of β-strands.

Although they share strong sequence similarities and bind DNA sites commonly via αI-helix and N-terminal, the MEF2-type MADS domain induces minimal DNA bending (about 17°) as compared with the dramatic DNA bending (about 72°) induced by the SRF-type MADS domain (Pellegrini et al. 1995; West et al. 1997; Santelli and Richmond 2000). Each MADS domain (see fig. 1) contains a long αI-helix and an 11 amino acid N-terminal random coil (N-extension) comprising the DNA contacting layer, and two β-strands connected by a β-turn forming the middle layer playing a major role in the dimerization (Pellegrini et al. 1995; Tan and Richmond 1998; Santelli and Richmond 2000). The highly conserved residues are R2, K4, and I5 in the N-terminal, and K22, R23, K24, G26, K29, K30, and E33 residues in the αI-helix, all of them are indispensable for the DNA binding. For example (Santelli and Richmond 2000), in N-terminal region, I5 interacts with the phosphodiester backbone to push R2 into the DNA groove and to promote that K4 making a salt link with the phosphate group of A + 1, another example (Santelli and Richmond 2000) in αI-helix is K22 and K29 hydrogen bond to DNA from the major groove side as compared with R23 and K30 hydrogen bonding to DNA from the minor groove side, and R23 and K29 are linked together through hydrogen bonds with E33 forming a seemingly rigid constellation of side chains important for specifying the local DNA conformation. Other conserved residues are probably critical for the proper positioning of the highly conserved residues to bind DNA, for protein dimerizing with other MADS factors, and for maintaining the folded conformation of the protein. In contrast to the similarities, the primary differences between the two types are at positions 2–3 in the N-terminal, position 13 in the αI-helix, and positions 48–53 in the β-loop. In the process of the MADS domain binding to DNA sites (Pellegrini et al. 1995; Santelli and Richmond 2000), the N-terminal penetrates the minor groove formed by the A/T-rich nucleotides, and R2 side chain (SRF) lies in the minor groove with the opposite orientation as compared with the homologous R2 (MEF2); the basic residues K13 and K24 (SRF) cause DNA bending through their interaction with DNA phosphate groups and then facilitate T50 and H52 binding to flanking DNA outside the CArG box in the major groove, somewhat explaining the presence of conserved positions in the SRF-CArG-binding context as described previously by us (Wu et al. 2010), whereas the acid residue E13 (MEF2) cannot directly bind to DNA phosphate groups and hence not facilitate the β-loop binding to the DNA. This scenario explains well positions 48–52 are highly conserved in SRF-type MADS domains compared with relatively less conserved in MEF2-type MADS domains. In addition to K13 promoting that T50 and H52 bind to DNA phosphates, Y17 in SRF can form hydrogen bonds with T50 and influence the orientation of β-loop to the underlying αI-helix, whereas Q17 in MEF2 has no such function (Huang et al. 2000). The main amino acid positions that induce different DNA bending between SRF- and MEF2-type MADS domains (Molkentin et al. 1996; West et al. 1997; West and Sharrocks 1999; Santelli and Richmond 2000) were listed in table 1.

Furthermore, the MADS domain usually does more than just bind to DNA. An example is that the C-terminal half of the MADS domain forms part of dimerization to reinforce the efficient dimerization by an additional domain (i.e., SAM domain in SRF-type proteins and MEF2-specific domain in MEF2-type proteins) adjacent to the MADS domain (West and Sharrocks 1999; Huang et al. 2000; Santelli and Richmond 2000). Hence, we suggest that different dimerization patterns between SRF- and MEF2-type MADS domains is the main reason for the presence of

Table 1. Main Amino Acid Positions Inducing Different DNA Bending Between SRF- and MEF2-Type MADS Domains.

Position	3	13	50	52
Residue in SRF	V	K	T	H
Residue in MEF2	K	E	S	K
Function	Mutation of V3K into SRF reduces DNA bending	Mutation of K13E into SRF severely reduces DNA bending	Mutation of T50 in SRF severely attenuates DNA bending	The additional mutation H52 to K13 into SRF reduces a further DNA bending

different conserved amino acids in the C-terminal halves of SRF-type and MEF2-type MADS domains.

Ancestor of SRF- and MEF2-Type MADS Domains

We inferred the ancestor of SRF- and MEF2-type MADS domains (see fig. 2) to unravel the origin of the divergence of SRF- and MEF2-type MADS domains. First, we inferred the ancestor of MEF2A, MEF2B, MEF2C, and MEF2D mainly from vertebrates and then inferred their common ancestor with other MEF2-type MADS domains from animals and fungi. Second, we inferred the ancestor of SRF-type MADS domains from animals and fungi. Finally, we used the ancestor of MEF2, the ancestor of SRF, and other MADS domains from protista and bacteria to infer the ancestor of SRF- and MEF2-type MADS domains. For convenience, we designated the ancestor of SRF- and MEF2-type MADS domains as co-SRF-MEF2.

The ancestor sequence of MEF2-type MADS domains (referred to 109 in fig. 2) is "GRKKIQIERISDDRNRQVTFTKRKNGLMKKAYELSVLCDCEIALIIFNSNNKLFQYSS," the ancestor sequence of SRF-type MADS domains (referred to 173 in fig. 2) is "GRRKIKIEFIDDKSRRHITFSKRKAGIMKKAYELSTLTGTQVLLLVASETGHVYTFAT," and the ancestor sequence of MEF2-, SRF-type MADS domains from eukaryotes, and other MADS domains from bacteria and phage (referred to 108 in fig. 2) is "GRRKIKIERIDDERNRQVTFTKRKAGLMKKAYELSVLCDCQIALIIANSNGKLFQYSS," which we designated as co-SRF-MEF2 in the following. Sequence alignment showed that the co-SRF-MEF2 is more similar to MEF2-type MADS domains than to SRF-type MADS domains. In addition, the analysis of MADS domains in bacteria and phage showed that these MADS domains are more similar to MEF2-type MADS domains than to SRF-type MADS domains. Especially, in that position 13 also shows preference for acidic residues as found in the MEF2-type MADS domain, which is the major determinant causing DNA bending. Furthermore, a recent study (Gramzow et al. 2010) proposed that the MADS domain originated from a region of topoisomerases IIA subunit A. From Gramzow's literature (2010), we collected MADS domains and partial sequences of TOPOIIA subunit A and aligned these sequences. Although the result showed position 13 in TOPOIIA subunit A is generally an Ile or Leu residue different from both MEF2- and SRF-type MADS domains, the adjacent position 12 in TOPOIIA showed a preference for Asp or Glu similar to that found in MEF2-type MADS domains at position 13, arbitrarily confirming that the ancestor sequence is more similar to MEF2-like MADS domain than to SRF-like MADS domain. These results led us to conclude that before the divergence of plants, fungi, and animals, a duplication event has occurred duplicating a MEF2-type MADS gene, giving rise to a SRF-type MADS gene.

MADS Binding Sites

In mammals, there is one SRF and four members of the MEF2 subfamily, in addition with their greater diversity by alternative splicing variants, belonging to the MADS domain family (Theissen et al. 1996). The binding sites for these factors are similar to each other containing A/T-rich DNA sequences. However, there are still many different properties between SRF- and MEF2-binding sites in line with the differences between the SRF- and MEF2-type MADS domains.

We generated Sequence Logos (fig. 3) of the binding sites for the two types of MADS domains using WebLogo (Crooks et al. 2004) to investigate similarities and differences between SRF sites and MEF2 sites. By comparing the Logos, we can see that, besides similarities, there are some important differences between the Logos of the two types of DNA-binding sites. In the middle (-3 to $+3$), there is a common preference for A/T-rich sequence. However, at one strand of MEF2 sites, it appears to be only A at some positions, such as -1, $+1$, and $+2$, and only T at position $+3$, and at the complementary strand of MEF2 sites, it appears to be T at positions -1, $+1$, and $+2$, and A at $+3$. In contrast to the preference in MEF2 sites, it appears no matter A or T at the positions or at the complementary positions in SRF sites. In the case of the two sides (-4, -5, $+4$, and $+5$), there is preference for C and G at -5 and $+5$ with subtle base substitutions for T and A roughly in common. However, there is preference for C and G in SRF sites with subtle base substitutions for T and A at -4 and $+4$, respectively; on the contrary, in the case of MEF2 sites, T and A account for the same occupancy as C and G at -4 and $+4$ in SRF sites.

X-ray crystal structures (Pellegrini et al. 1995; Santelli and Richmond 2000) showed that the DNA is bent overall to only 17° for MEF2 binding as compared with 72° for SRF binding. In addition, the A/T-rich region of the MEF2 site forms the minor groove of 3.7 Å and the major groove of 12.4 Å (Santelli and Richmond 2000) as compared with the minor groove of 8 Å and the major groove of 19.5 Å (Pellegrini et al. 1995) in the A/T-rich region of the SRF site. The two different conformational states of the DNA is not

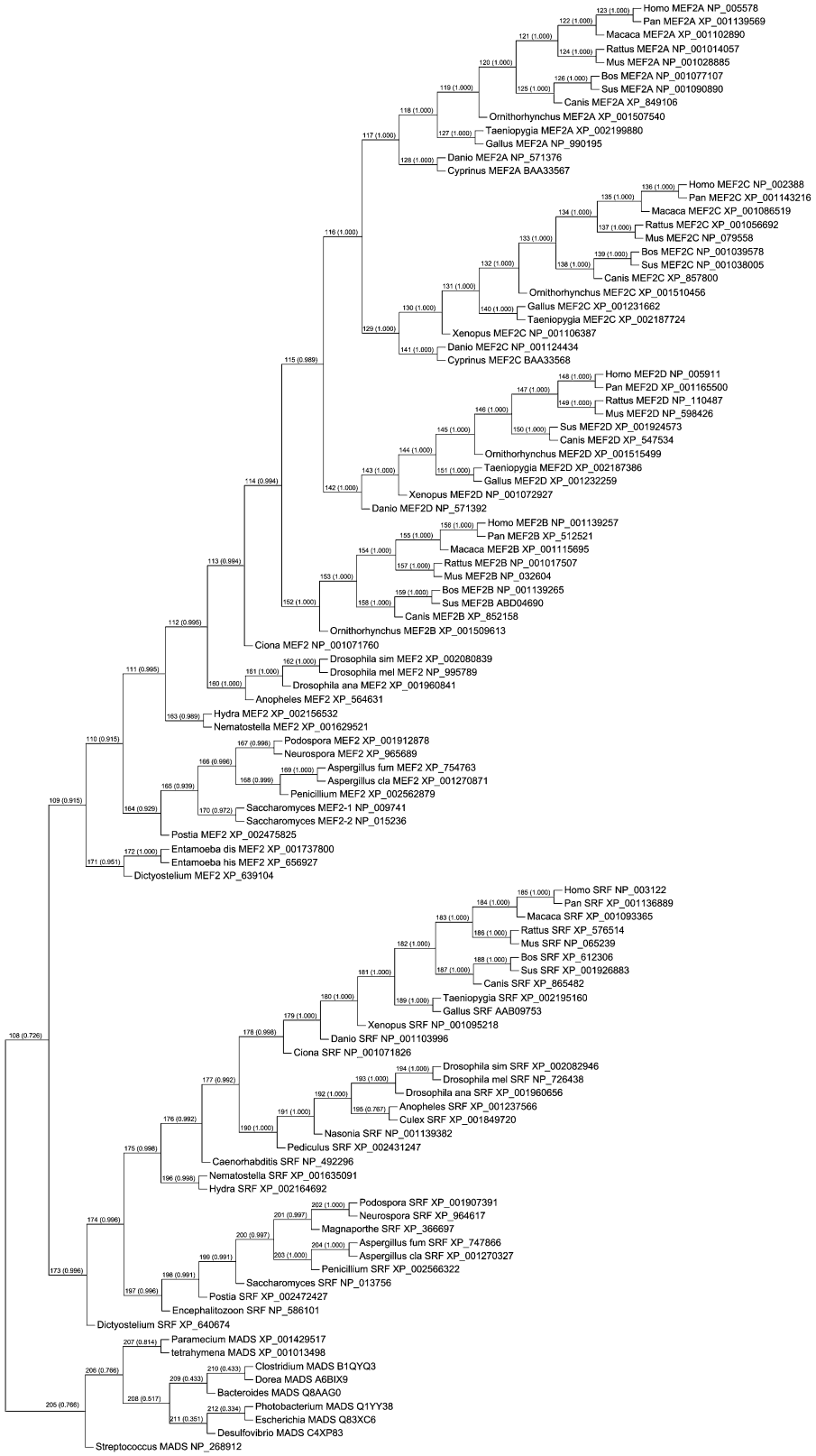

Fig. 2 The phylogeny of MADS domains from different species ranging from animals and fungi to protista, bacteria, and phage. Leaves comprise species name, MADS type, and protein accession number. On branches, the numbers referred to the ancestor nodes of the right subtrees are shown left of brackets, and the accuracy values for inferring the ancestor nodes are shown within brackets. The corresponding sequences of ancestor nodes are listed in supplementary material S6, Supplementary Material online.

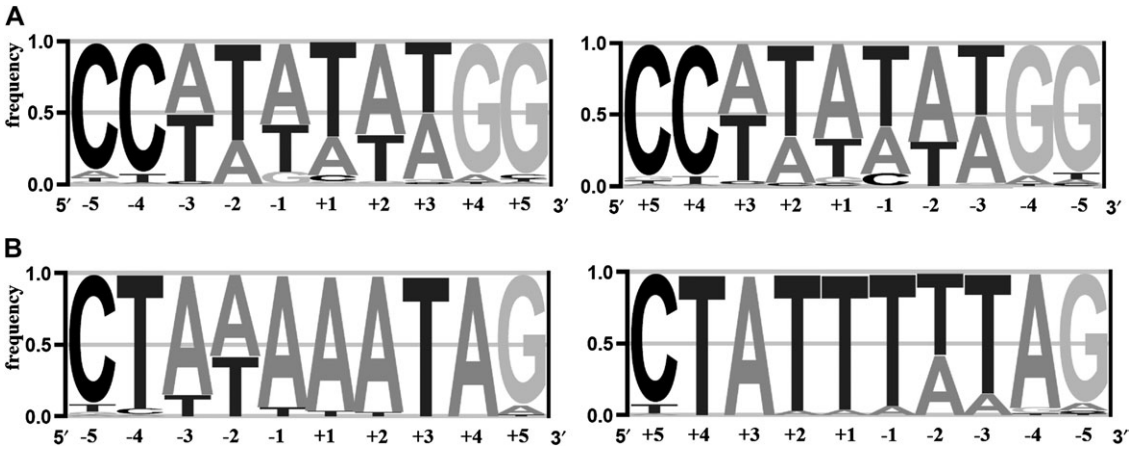

Fig. 3 Logos of DNA-binding sites for SRF- and MEF2-type MADS domains in the mouse genome. (*A*) 223 SRF-binding sites were used to generate Sequence Logo of SRF sites. (*B*) 241 MEF2-binding sites were used to generate the Sequence Logo of MEF2 sites. The two Sequence Logos on the right side were generated from the complementary sequences of the SRF- and MEF2-binding sites to a comprehensive understanding of positive and negative binding sites for SRF and MEF2. For convenience, we designated the ten positions from −5 to +5. Longitudinal coordinates represent the frequencies of the occupancies corresponding to the four nucleotides (A, T, G, and C).

only induced by their cognate transcription factors but also determined by the intrinsic features of different A/T-rich compositions in the MEF2 site compared with the SRF site. The helical twist in the centre of the MEF2 site induced by ApA (TpT) dinucleotides forms a narrower minor groove than that formed by the SRF site, and this narrower minor groove is in line with the short N-terminal extension of MEF2-type MADS factors and allows the DNA to bend a little when the MEF2 site is bound by MEF2. However, the helical twist, which is partly compensated and broadened by the ApT dinucleotides in the middle of the SRF site, forms the minor groove >2 times wider than in the MEF2 site (Pellegrini et al. 1995; Santelli and Richmond 2000), and this wider minor groove suits the long N-terminal extension of SRF-type MADS factors well, allowing the DNA to bend more by binding to SRF. In addition, different nucleotides at two-sides ($C_{-5}C_{-4}$ and $G_{+4}G_{+5}$ in SRF sites and $C_{-5}T_{-4}$ and $A_{+4}G_{+5}$ in MEF2 sites) also contribute to DNA bending (West and Sharrocks 1999; Stepanek et al. 2007). For example, mutation of GG in SRF sites severely reduces DNA bending (West and Sharrocks 1999).

Intriguingly, the frequency of G and C at positions −1 and +1 is very low in the SRF site (fig. 3*A*, SRF site), and the substitution of consensus CArG for a degenerate CArG would lower the binding capacity of SRF in vitro. For example, there is a degenerate CArG box (G substitution at position −1) located within 70-bp upstream region of the SM α-actin TSS and in vitro SRF binds to this region weakly. However, in vivo, during normal development, the substitution has no effect on SRF binding to this degenerate CArG box, but in injury-induced tissues, the substitution attenuates SRF binding to this degenerate CArG box (Hendrix et al. 2005). These results indicate that sometimes there is adaptive selection in spite of commonly strong negative selection on the mutation of A (T) to G (C) at positions −1 and +1.

Conservation and Evolution of SRF Sites in the Mouse and Human Genomes

Over 200 functional CArG-boxes (see Materials and Methods) were extracted from the mouse and human genomes to test the conservation level of SRF sites. Figure 4 presents the Logos of the CArG-boxes and reveals the similarities with the consensus CArG-box $CC(A/T)_6GG$ but also some differences between the binding sites of the human and mouse genomes. In detail, the outer 6 positions are most alike, although minor differences are observed (such as position +3). However, the inner 4 positions have different properties: The CArG boxes in human favor more $T_{-2}A_{-1}T_{+1}A_{+2}$ than in mouse, demonstrating that the functional constraints on CArG in human are stricter than in mouse. Furthermore, different IC generated from enoLOGOS (Workman et al. 2005) (see supplementary material S5, Supplementary Material online) of the CArG boxes implicates a different conservation level between the two species. CArG boxes with 12.256 bites in the human genome show that CArG occurs once expectedly every ~4,900 bp ($2^{12.256}$), whereas CArG boxes with 10.598 bites in the mouse genome show that CArG occurs once expectedly every ~1,550 bp ($2^{10.598}$). Interestingly, this data suggest that CArG boxes are more conserved in the human genome than in the mouse genome, which will be investigated more below.

In addition, we classified the functional CArG boxes into CArG types according to the nucleotide composition across the box from the mouse and the human genomes, respectively, and the results are depicted in fig. 5. It can be seen that the number of CArG types is much greater in mouse than in human. Besides, *Z*-test (see Materials and Methods) also demonstrated that diversity of CArG boxes between mouse and human is significantly different (Z value $= -4.8709$ and P value < 0.0001 (5.55×10^{-7}). These results illustrate that functional CArG boxes are more conserved in human than in mouse and the diversity of CArG boxes in human is much lower than in mouse.

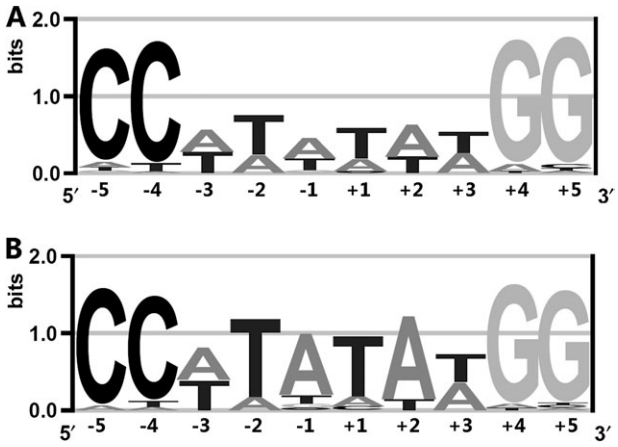

FIG. 4 Sequence Logos of CArG boxes from the mouse (*A*) and human (*B*) genomes. The two Logos show a high level of similarities in preferred base composition across the CArG boxes. However, the two Logos exhibit a different conservation level across the core tetranucleotides in CArG boxes between the mouse and human genomes. The logos were generated from 223 and 216 CArG boxes in the mouse (*A*) and human (*B*) genomes, respectively.

The data presented in figs. 4 and 5 suggest that the evolution rate of CArG boxes in the mouse genome is faster than that in the human genome. To further investigate this, we retrieved orthologous CArG boxes (see table 2) from both human and mouse since divergence from an outgroup species that of dog and/or cow (Madsen et al. 2001; Murphy et al. 2001; Reyes et al. 2004). Two points can be observed from date in table 2. First, the evolution rate of CArG boxes in the mouse genome is indeed faster than that in the human genome. Second, interestingly, a CArG-like sequence that is probably functionless could potentially mutate into a functional CArG box, which can be bound by SRF (e.g., AGPT1 **CC**TATTAAGG of mouse), and conversely, a functional CArG box could mutate into a functionless CArG-like sequence that lose its function bound by SRF (e.g., STX12 C**T**ATATTTG**A** of mouse). Although no direct evidence is yet available to support such a theory, some available indirect evidence (Tilly et al. 1998; Mack and Owens 1999; Strobeck et al. 2001; Freddie et al. 2007; Chen et al. 2010) that mutations of CArG boxes abolish the function in regulating gene expression lends support to this theory. These results give us some clues on the origin and disruption of functional CArG boxes. However, it also leads to one question: What factors drive functionless CArG-like sequences mutating into functional CArG boxes and vice versa? To try to answer this question, we searched the sequences flanking mutated functionless CArG-like sequences to examine whether there are other CArG boxes nearby, which might act as the bench CArG instead of the mutated one. Although, we indeed found a few CArG boxes flanking mutated boxes, it is not sufficient to demonstrate that one of these boxes can be bound by SRF. In addition, the adjacent positions (Wu et al. 2010) of CArG box may also play roles on CArG mutation process. We hope that further experiments will be carried out to identify whether these flanking CArG boxes might be bound by SRF instead of the mutated ones to regulate gene transcription.

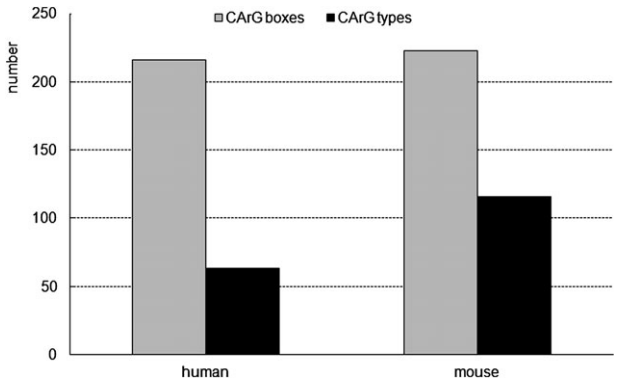

FIG. 5 Numbers of CArG boxes and CArG types in the human and mouse genomes. The number of CArG boxes indicated in gray in human is roughly equal to that in mouse. However, the number of CArG types indicated by black in human is nearly 2-fold less than that in mouse.

Conclusion

In summary, we mainly used phylogenetic analyses to study the conservation and evolution in and among MADS domains and their DNA-binding sites. The analyses reveal that the highly conserved positions in common in SRF- and MEF2-type domains are critical for maintaining the configuration of the DNA contacting layer and for protein dimerization. Except for the highly conserved positions in both type MADS domains, there are also some striking differences between SRF- and MEF2-type MADS domains. One is amino acid position 13 of SRF-type MADS domains, which has a highly conserved basic residue (K), whereas the corresponding position of MEF2-type MADS domains has a conserved acidic amino acid, which cannot interact with the DNA phosphate, as the basic residue (K) of SRF does. This direct interaction of SRF with DNA allows the β-loop to make connections with the DNA in the major groove, giving an explanation why the β-loop of SRF-type MADS domains is more conserved than that of MEF2-type MADS domains. In line with the similarities and differences between the two types of MADS domains, there are also some similarities and differences between the two types of MADS-binding sites. The MADS-binding sites all have an A/T-rich region in the middle; however, MEF2 sites have a distinct A/T composition from SRF sites. The inference of co-SRF-MEF2 showed that the ancestor sequence is more similar to MEF2-type MADS domains than to SRF-type MADS domains, arbitrarily concluding that co-SRF-MEF2-binding site is more similar to the MEF2 site than to the SRF site. Furthermore, we investigated the conservation, generation, and disruption of SRF sites in mammalian genomes. The results show that the evolution rate of CArG boxes is faster in the mouse genome than in the human genome, and a functional CArG box could mutate to become functionless and vice versa. A major task that lies ahead will be to elucidate what factors drive and constrain the evolution of CArG box, especially the generation and disruption of CArG box. In addition, different DNA-binding specificities occur between SRF- and MEF2-type

Table 2. Evolution of SRF Sites in Mammalian Genomes.

Gene	Mouse CArG Box	Human Ortholog	Cow Ortholog	Dog Ortholog
EGR1	***CCTTATATGG***	***CCTTATTTGG***	CCTTATTTGG	CCTTGTTTGG
RPL24	CCCTATAAGG	***CCTTATAAGG***	CCTTATAAGG	GCTTATAAGG
CKM	***CCATGTAAGG***	CCTTGTAAGG	CCTTGTAAGG	CCTTGTAAGG
MAPKBP1	CCGTAAAAGG	***CCATAAAAGG***	CCATAAAAGG	CCTTAAAAGG
TRIP6	***CCAAAATTGG***	CCAAAATTAG	CCAAAATTAG	CCAAAATTAG
GPC4	***CCATTCATGG***	CCTTTCATGG	CCTTTCATGG	CCTTTCATGG
LZF	***CCTTTTATGG***	CCTTTTAAGG	TCTTAAAAGG	CCTTATA-GG
SHKBP1	***CCAAATATGG***	CCAAATACGG	CCAAATACAG	CCAAAAGCGG
EEF1B2	CCTTGTAAGG	***CCTTCTAAGG***	CCTTCTCAAG	CCTTCTCAAG
STX12	CTATATTTGA	***CCATATTTGG***	CCATATTTGG	CCATATATGG
AGPT1	***CCTATTAAGG***	TGTATTAAGG	TGTATTAAGG	TGTATTAAGG
PRM1	***CCATATTTGG***	CCATATATGG	CCATATATGG	CCATATATGG
CSRP2	***CCATTAATAG***	CAATTAGTAG	CAATCAGTAG	CAATTAGTAG
PTGS2	CCATACCTGG	***CCATAAATGG***	CCACAAATGG	CCACAAATCG
LPP	***CCTAATATGG***	CCATATATGG	CCATATATGG	CCATATATGG
	CCATATAAGG	CCATATATGG	CCATATATGG	CCATATATGG
MYH4	***CCAAAAATGG***	CCAAAAAAGG	CCAAAAAAGG	CCAAAAAAGG
PTGS1	***CCATAAATGG***	CTATAAATGG	CTACAAGTGG	CTATAAATGG
ANKRD1	***CCATTTTAGC***	CCTATTTAGC	CCTTTTTAGC	CCTTTTTAGT
	CCTTATAAGG	***CCATATAAGG***	***CCATATAAGG***	CCATATAAGG
PTMS	CCTTAAAAGC	***CCTTATAAGC***	No	CTTTATAAGC
NFYB	***CCTTTAAAGG***	CCTTAAAAGG	No	CCTTAAAAGG
RRAD	***CCTTTTTAGG***	CCTTCTTAGG	No	CCCTCTCAGA
ACTC1	***CCCTATATGG***	CCCTATTTGG	CCCTATTTGG	No
	CCAAGAATGG	CCATGAATGG	CCATGAATGG	No
ANKRD1	***CCTTTATGGG***	CTTTTATGGG	CTTTCATGGA	No
ITGB1BP2	***CCATGTTTGG***	CCATGTTAGG	CCATGTTAGG	No
PTN	***CCAAGAATGG***	CCAAAAAAGG	CCAGAAAAAG	No
DTR	***CCAAAAAATG***	CCAAAAATTG	CCAAAATTTG	No
IER2	TCTTATATGG	***CCTAATATGG***	CCTTATTAGG	CCTTATATGG
TUFT1	***CCATATAAGT***	***CTATATAAGG***	CCATATAAGG	CCATATAAGG
CTGF	***CCTTATAGGG***	***CCATATACGG***	CCTTATACGG	ACTTATATGG
RG9MTD2	CCTGATACGG	***CCTCATATGG***	CCTGATATGG	CCTGATATAG
RFX5	GTATACAAGG	***GCATATATGG***	GTATATATGG	GTAGATGTGG
ELF5	***CCATAAAAGG***	CCAGAAAAGG	CCATAGAAGG	TCATAGAAGG
DUSP6	***CCTTGTATGG***	CCTTGTAAGG	CCTTGTATGA	CCTTGTGTGA
AKAP12	***CCTAATATGG***	***CCAAATATGG***	CCTAATATGG	CCTAATATGG
MUS81	CCTTGTCTGG	***CCTTATCTGG***	CCTTGTCTGG	CCTTGTCTGG
RAI2	***CTAAATATGG***	***CCAAATATGG***	CTAAATATGG	CTAAATATGG
UROD	***CCTAATTAGG***	CCTAATTAGA	CCTAATTAGG	CCTAATTAGG
CPT1B	***CCAATTTGGG***	CCGATTTGGG	CAAATTTGGG	CAAATTTGGG
MYH6	***CCTTTCATGG***	CCTTTCTTGG	CCTTTCATGG	CCTTTCATGG
NKX2-5	***CTTTAAAAGG***	CTTTAAAAGT	CTTTAAAAGG	CTTTAAAAGG
HEY2	***CCAATTTAGG***	CCAATTTCGG	No	CCAATTTAGG
CSRP2	***CCAAATTTGG***	CCAAATTTGA	No	ACAAATTTGG

NOTE.—The genes were reported as experimentally verified as SRF targets in mouse or in human. The CArG boxes, which have been verified bound by SRF (reported data), are indicated by bold-faced, italicized type. The corresponding orthologous sites from other species were retrieved by BLAST (see Materials and Methods) and the substitutions across the sites in human and/or in mouse are indicated by underline using Cow and/or Dog as outgroups.

MADS domains and even within the two types, and hence a deeper characterization of the evolution of MADS domains and their cognate DNA-binding sites may reveal a more comprehensive view of the functions of these very important and essential transcription factors in organisms during evolution.

Supplementary Material

Supplementary materials S1–S6 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We are grateful to the members at the Bioinformatics Center of Northwest A&F University for many useful suggestions. We are deeply appreciative to the two anonymous reviewers for their insightful suggestions and criticisms to improve the manuscript. Research in S.d.F. lab is financed by the Mexican Science Council (CONACyT 82826). This work is supported by National High-Tech R & D Program of China (863 Program) (grant no. 009AA02Z308).

References

Adamczyk BJ, Fernandez DE. 2009. MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in Arabidopsis. *Plant Physiol.* 149:1713–1723.

Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, Ribas DPL, Martinez-Castilla L, Yanofsky MF. 2000. An ancestral MADS-box gene duplication occurred before the

divergence of plants and animals. *Proc Natl Acad Sci U S A*. 97:5328–5333.
Andres V, Cervera M, Mahdavi V. 1995. Determination of the consensus binding site for MEF2 expressed in muscle and brain reveals tissue-specific sequence constraints. *J Biol Chem*. 270:23246–23249.
Balza RO Jr., Misra RP. 2006. Role of the serum response factor in regulating contractile apparatus gene expression and sarcomeric integrity in cardiomyocytes. *J Biol Chem*. 281:6498–651.
Chen CH, Wu ML, Lee YC, Layne MD, Yet SF. 2010. Intronic CArG box regulates cysteine-rich protein 2 expression in the adult but not in developing vasculature. *Arterioscler Thromb Vasc Biol*. 4:835–842.
Cooper SJ, Trinklein ND, Nguyen L, Myers RM. 2007. Serum response factor binding sites differ in three human cell types. *Genome Res*. 17:136–144.
Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res*. 14:1188–1190.
de Folter S, Angenent GC. 2006. trans meets cis in MADS science. *Trends Plant Sci*. 11:224–231.
Flavell SW, Kim TK, Gray JM, Harmin DA, Hemberg M, Hong EJ, Markenscoff-Papadimitriou E, Bear DM, Greenberg ME. 2008. Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron* 60:1022–1038.
Fornara F, Marziani G, Mizzi L, Kater M, Colombo L. 2003. MADS-box genes controlling flower development in rice. *Plant Biol*. 5:16–22.
Freddie CT, Ji Z, Marais A, Sharrocks AD. 2007. Functional interactions between the Forkhead transcription factor FOXK1 and the MADS-box protein SRF. *Nucleic Acids Res*. 15:5203–5212.
Gramzow L, Ritz MS, Theissen G. 2010. On the origin of MADS-domain transcription factors. *Trends Genet*. 26:149–153.
Haeseleer PD. 2006. What are DNA sequence motifs? *Nat Biotechnol*. 24:423–425.
Hendrix JA, Wamhoff BR, McDonald OG, Sinha S, Yoshida T, Owens GK. 2005. 5′ CArG degeneracy in smooth muscle alpha-actin is required for injury-induced gene suppression in vivo. *J Clin Invest*. 115:418–427.
Herskowitz I. 1989. A regulatory hierarchy for cell specialization in yeast. *Nature* 342:749–757.
Huang K, Louis JM, Donaldson L, Lim FL, Sharrocks AD, Clore GM. 2000. Solution structure of the MEF2A–DNA complex: structural basis for the modulation of DNA bending and specificity by MADS-box transcription factors. *EMBO J*. 19:2615.
Immink RG, Gadella TJ, Ferrario S, Busscher M, Angenent GC. 2002. Analysis of MADS box protein-protein interactions in living plant cells. *Proc Natl Acad Sci U S A*. 99:2416–2421.
Kaufmann K, Melzer R, Theissen G. 2005. MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* 347:183–198.
Kofuji R, Sumikawa N, Yamasaki M, Kondo K, Ueda K, Ito M, Hasebe M. 2003. Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses. *Mol Biol Evol*. 20:1963–1977.
Larkin MA, Blackshields G, Brown NP, et al. (13 coauthors). 2007. ClustalW and ClustalX version 2.0. *Bioinformatics* 23:2947–2948.
Li H, Coghlan A, Ruan J, et al. (15 coauthors). 2006. TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res*. 34:D572–D580.
Mack CP, Owens GK. 1999. Regulation of smooth muscle alpha-actin expression in vivo is dependent on CArG elements within the 5′ and first intron promoter regions. *Circ Res*. 84:852–861.
Madsen O, Scally M, Douady CJ, et al. (10 coauthors). 2001. Parallel adaptive radiations in two major clades of placental mammals. *Nature* 409:610–614.
Messenguy F, Dubois E. 2003. Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. *Gene* 316:1–21.
Miano JM. 2003. Serum response factor: toggling between disparate programs of gene expression. *J Mol Cell Cardiol*. 35:577–593.
Molkentin JD, Black BL, Martin JF, Olson EN. 1996. Mutational analysis of the DNA binding, dimerization, and transcriptional activation domains of MEF2C. *Mol Cell Biol*. 16:2627–2636.
Muller J, Muller K. 2004. TreeGraph: automated drawing of complex tree figures using an extensible tree description format. *Mol Ecol Notes*. 4:786–788.
Munster T, Pahnke J, Di Rosa A, Kim JT, Martin W, Saedler H, Theissen G. 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 U S A*. 94:2415–2420.
Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ. 2001. Molecular phylogenetics and the origins of placental mammals. *Nature* 409:614–618.
Ng M, Yanofsky MF. 2001. Function and evolution of the plant MADS-box gene family. *Nat Rev Genet*. 2:186–195.
Norman C, Runswick M, Pollock R, Treisman R. 1988. Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* 55:989–1003.
Nurrish SJ, Treisman R. 1995. DNA binding specificity determinants in MADS-box transcription factors. *Mol Cell Biol*. 15:4076–4085.
Page RD. 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci*. 12:357–358.
Parenicova L, de Folter S, Kieffer M, et al. (12 coauthors). 2003. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. *Plant Cell*. 15:1538–1551.
Passmore S, Elble R, Tye BK. 1989. A protein involved in minichromosome maintenance in yeast binds a transcriptional enhancer conserved in eukaryotes. *Genes Dev*. 3:921–935.
Pellegrini L, Tan S, Richmond TJ. 1995. Structure of serum response factor core bound to DNA. *Nature* 376:490–498.
Petit MM, Lindskog H, Larsson E, et al. (12 coauthors). 2008. Smooth muscle expression of lipoma preferred partner is mediated by an alternative intronic promoter that is regulated by serum response factor/myocardin. *Circ Res*. 103:61–69.
Philippar U, Schratt G, Dieterich C, et al. (11 coauthors). 2004. The SRF target gene Fhl2 antagonizes RhoA/MAL-dependent activation of SRF. *Mol Cell*. 16:867–880.
Pollock R, Treisman R. 1990. A sensitive method for the determination of protein-DNA binding specificities. *Nucleic Acids Res*. 18:6197–6204.
Pollock R, Treisman R. 1991. Human SRF-related proteins: DNA-binding properties and potential regulatory targets. *Genes Dev*. 5:2327–2341.
Potthoff MJ, Olson EN. 2007. MEF2: a central regulator of diverse developmental programs. *Development* 134:4131–4140.
Reyes A, Gissi C, Catzeflis F, Nevo E, Pesole G, Saccone C. 2004. Congruent mammalian trees from mitochondrial and nuclear genes using Bayesian methods. *Mol Biol Evol*. 21:397–403.
Santelli E, Richmond TJ. 2000. Crystal structure of MEF2A core bound to DNA at 1.5 A resolution. *J Mol Biol*. 297:437–449.
Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H, Sommer H. 1990. Genetic control of flower development by homeotic genes in Antirrhinum majus. *Science* 250:931–936.
Selvaraj A, Prywes R. 2004. Expression profiling of serum inducible genes identifies a subset of SRF target genes that are MKL dependent. *BMC Mol Biol*. 5:13.

Shore P, Sharrocks AD. 1995. The MADS-box family of transcription factors. *Eur J Biochem* 229:1–13.

Sommer H, Beltran JP, Huijser P, Pape H, Lonnig WE, Saedler H, Schwarz-Sommer Z. 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.

Stepanek J, Vincent M, Turpin PY, Paulin D, Fermandjian S, Alpert B, Zentz C. 2007. C → G base mutations in the CArG box of c-fos serum response element alter its bending flexibility. Consequences for core-SRF recognition. *FEBS J*. 274:2333–2348.

Stormo GD. 2000. DNA binding sites: representation and discovery. *Bioinformatics* 16:16–23.

Strobeck M, Kim S, Zhang J, Clendenin C, Du KL, Parmacek MS. 2001. Binding of serum response factor to CArG box sequences is necessary but not sufficient to restrict gene expression to arterial smooth muscle cells. *J Biol Chem*. 276:16418.

Sun Q, Chen G, Streb JW, Long X, Yang Y, Stoeckert CJ, Miano JM. 2006. Defining the mammalian CArGome. *Genome Res*. 16:197–207.

Tan S, Richmond TJ. 1998. Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex. *Nature* 391: 660–666.

Thakare D, Tang W, Hill K, Perry SE. 2008. The MADS-domain transcriptional regulator AGAMOUS-LIKE15 promotes somatic embryo development in Arabidopsis and soybean. *Plant Physiol*. 146:1663–1672.

Theissen G, Kim JT, Saedler H. 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.

Tilly JJ, Allen DW, Jack T. 1998. The CArG boxes in the promoter of the Arabidopsis floral organ identity gene APETALA3 mediate diverse regulatory effects. *Development* 125:1647.

Treisman R, Ammerer G. 1992. The SRF and MCM1 transcription factors. *Curr Opin Genet Dev*. 2:221–226.

Wakaguri H, Yamashita R, Suzuki Y, Sugano S, Nakai K. 2008. DBTSS: database of transcription start sites, progress report 2008. *Nucleic Acids Res*. 36:D97–D101.

West AG, Sharrocks AD. 1999. MADS-box transcription factors adopt alternative mechanisms for bending DNA. *J Mol Biol*. 286:1311–1323.

West AG, Shore P, Sharrocks AD. 1997. DNA binding by MADS-box transcription factors: a molecular mechanism for differential DNA bending. *Mol Cell Biol*. 17:2876–2887.

Wheeler DL, Barrett T, Benson DA, et al. (30 coauthors). 2007. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 35:D5–D12.

Workman CT, Yin Y, Corcoran DL, Ideker T, Stormo GD, Benos PV. 2005. enoLOGOS: a versatile web tool for energy normalized sequence logos. *Nucleic Acids Res*. 33:W389–W392.

Wu WW, Shen X, Tao SH. 2010. Characteristics of the CArG-SRF binding context in mammalian genomes. *Mamm Genome*. 21:104–113.

Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM. 1990. The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. *Nature* 346:35–39.

Zhang J, Nei M. 1997. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. *J Mol Evol*. 44:S139–S146.

Zhang SX, Garcia-Gras E, Wycuff DR, et al. (10 coauthors). 2005. Identification of direct serum-response factor gene targets during Me2SO-induced P19 cardiac cell differentiation. *J Biol Chem*. 280:19115–19126.

Zheng Y, Ren N, Wang H, Stromberg AJ, Perry SE. 2009. Global identification of targets of the Arabidopsis MADS domain protein AGAMOUS-Like15. *Plant Cell*. 21:2563–2577.