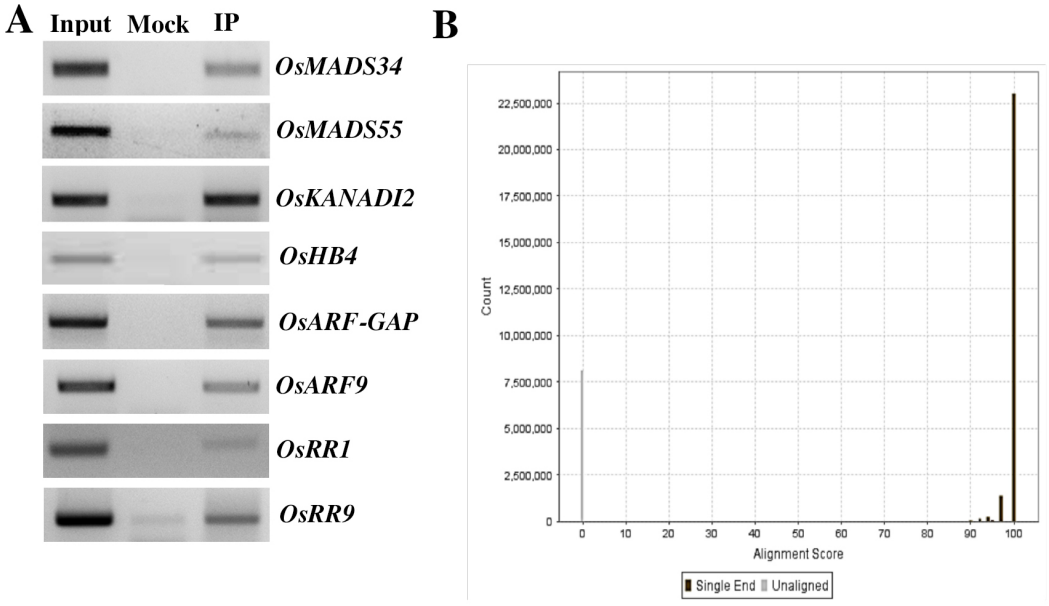
Genome-wide identification of target genes regulated by OsMADS1 during rice floret development reveals its DNA recognition properties

Imtiyaz Khanday[2, 3], Sanjukta Das[2], Grace L Chongloi[2], Manju Bansal, Ueli Grossniklaus and Usha Vijayraghavan*

Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560012, India (I. K., G.L.C., U. V.); b Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India (S. D., M. B.) and Department of Plant and Microbial Biology & Zurich-Basel Plant Science Center, University of Zurich, Zurich 8008, Switzerland (U. G.).



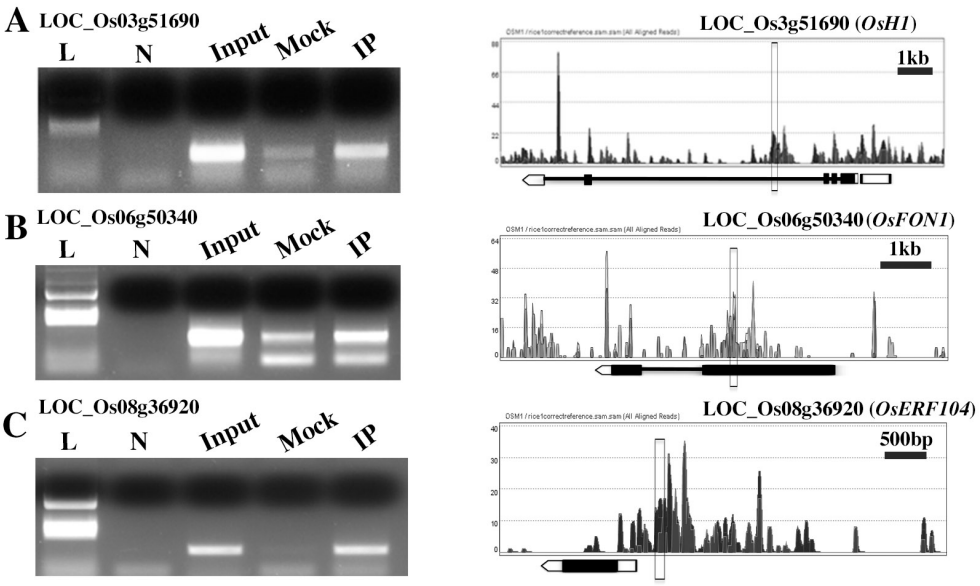
Supplemental Figure S1. Quality assurance of OsMADS1 immuno-precipitated sample and mapping of the ChIP-seq reads. **A**, Chromatin pooled from 8 independent ChIP experiments were demonstrated for efficient OsMADS1 binding on eight loci previously reported by Khanday et al. (2013) to be gene targets of OsMADS1. **B**, Roughly 32 million reads, obtained after the sequencing of immuno-precipitated DNA with anti-OsMADS1 antibody, are mapped to rice reference genome. Y-axis shows the number of reads and X-axis shows the alignment score. About 23 million reads mapped uniquely to rice genome with 100% identity.

A

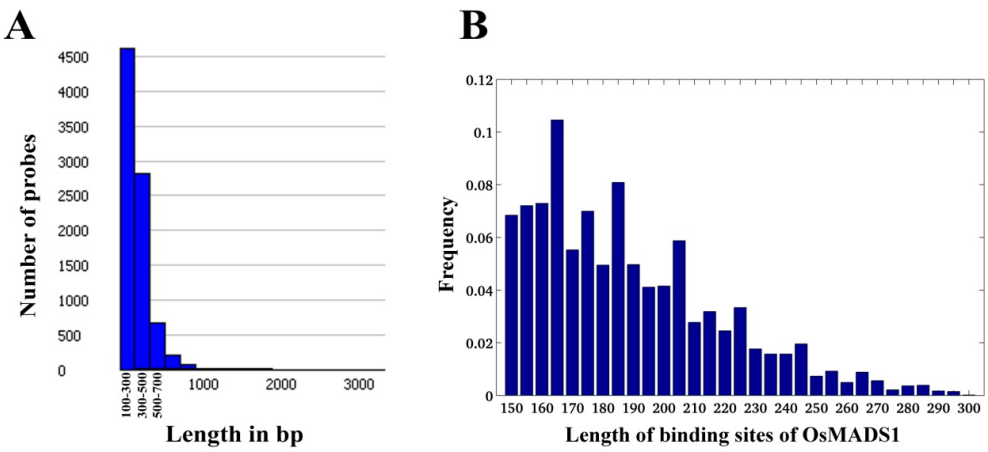
B

Chromosome	Number of peaks	Size of chromosome in Mb	Number of peaks per Mb
Chr01	602	45.0	13.4
Chr02	412	36.8	11.2
Chr03	465	37.3	12.5
Chr04	279	36.1	7.7
Chr05	234	30.1	7.8
Chr06	203	32.1	6.3
Chr07	214	30.4	7.1
Chr08	199	28.5	7.0
Chr09	174	23.9	7.3
Chr10	139	23.7	5.9
Chr11	105	31.2	3.4
Chr12	144	27.7	5.2

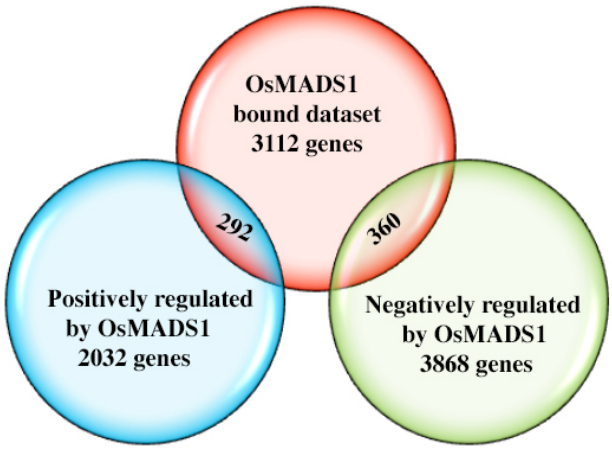
Supplemental Figure S2. Genomic view of OsMADS1 peaks and density of OsMADS1 binding sites. **A,** Genomic view of OsMADS1 peaks. **B**, Density of OsMADS1 binding peaks per Mb of DNA across the twelve chromosomes of rice



Supplemental Figure S3. Semi-quantitative PCR validation for some additional OsMADS1 ChIP-seq binding sites. OsMADS1 binding on 3rd Intron of *OsH1* **(A),** 1st exon of *OsFON1* **(B)** and upstream sequences of *OsERF104* **(C).** On the horizontal (X-axis) gene structure, black boxes are exons while black lines indicate introns. White boxes represent UTRs whereas the white pentagon (3' UTR) indicates orientation of the gene. Vertical (Y-axis) grey unfilled boxes depict the region assessed by ChIP- PCR analyses. L - *Hinf*I restricted pBluescript KS DNA ladder, N- negative control PCR with no template.



Supplemental Figure S4. Length of OsMADS1-bound sequences in the genome. **A,** Histogram showing the length of contigs of genomic sequence reads that are OsMADS1-bound. The majority (57%) of binding sites encompass 100-300 bp of chromosomal sequences while the rest can span from 300-500 bp with a very minor set going up to 1 kb. **B,** A DNA length frequency histogram for binding sites ranging between 150-300 bp. The predominant coverage of OsMADS1 binding sites is between 160 to 190 bp.



Supplemental Figure S5. Overlap of OsMADS1-bound gene loci with genes whose expression is modulated by OsMADS1. Red circle represents gene-associated loci from OsMADS1 ChIP-seq data set. This gene set was compared with gene expression data sets consisting of genes that are up-regulated or down-regulated in young florets of 2 to 20 mm *OsMADS1*RNAi panicles as compared to similarly staged wild-type panicles, Khanday et al. (2013). These genes with altered expression are represented in green and blue circles respectively. A total of 652 genes are directly regulated by OsMADS1, of which we deduce, 292 are directly and positively regulated by OsMADS1 while 360 are directly and negatively regulated by *OsMADS1* in wild-type panicles.

Supplementary Table S1: List of primers used for ChIP-PCR and gene expression studies. The TIGR Locus ID and corresponding gene annotation are also provided. For genes with no TIGR annotation, the closest ortholog annotation is designated.

Gene	Forward Primer (5'-3')	Reverse Primer (5'-3')	Use
LOC_Os01g57240 OsULTRAPETALA (ULT1)	CATCCACCGGTTTTCGGCATC	ACGTCACCTCCACGAA ATCCT	ChIP qPCR
LOC_Os02g46930 OsJMJ707	TGAACAAATGAGGGACATACC CTT	CCA CTT TCT CAA AAT CCG GAG CA	ChIP qPCR
LOC_Os03g56140 OsROUGH SHEATH1 (RS1)	GTAAAAATGGTGGCGATG CACC	GTCGGATTCTTGCTAT ATTCGGTT	ChIP qPCR
LOC_Os03g51690 OsH1	AGTTAGCCAAGGAGATCACCC	CTCCATGACTTGGTAT TCTGTG	ChIP PCR
LOC_Os04g46440 OsERF34	CGCTTGACGTTCAGATTTCCT	CTAATACATGACGCCG TGCAG	ChIP qPCR
LOC_Os04g47870 OsAGO1b	AGTACAAAGCTTGGGGGGAC	ATCTCTTGTCAGAATC GGCG	ChIP qPCR
LOC_Os05g11810 OsGA2 Oxidase 1 (OsGA2ox1)	TCTCGTCCCTCTACGTGCTT	CCGATTACAACGGGAG CTTA	ChIP qPCR
LOC_Os06g06900 OsSPATULA (SPT)	AGCATCCTCAAAATCCAC GCA	CTCGCAGTGGTTTCGT GTATT	ChIP qPCR
LOC_Os06g50340 OsFLORAL ORGAN NUMBER1 (OsFON1)	GGTGGTGTTTGGATCTAGGG	ACGAGACGAATCATGA GCC	ChIP PCR
LOC_Os07g06620 OsYABBY1	CTACGTACAGTGGAATGG TCA CA	ACACTTGCTGCTGACC TGGTT	ChIP qPCR
LOC_Os07g25740 OsHEN4-like	GGAAGCATACCACGAATA ATTGCA	AGCTCTGTGTATGGTG TTCGA G	ChIP qPCR
LOC_Os08g36920 OsERF104	GCATGCACCTTTTACCACTCC	CGGAGCAGTCCATTAA CACG	ChIP qPCR
LOC_Os08g31980 Trehalose 6 phosphate synthase (OsTre-6-ph-synthase)	GTTCCCATTTCTCCCTCA TCGT	TTTCGCGGATGTAACG CGG	ChIP qPCR
LOC_Os09g26999	GCGCCGTACTGATCAACATC	AGCTTGCAATTAAGCA	ChIP qPCR

OsDENSE and ERECT PANICLE1(OsDEP1)		GATTGCC	
LOC_Os09g35790 OsHSF24	GACACTACCTCCTTTTCTAAGTAA	TTCAGCCGGTTCAGAGCAGT	ChIP qPCR
LOC_Os09g35790 OsHSF24 3'UTR	CAGCAAATGGTGTTGTGATCCAC	AGACTCCAATCCTCATGCAGAG	RT-qPCR
LOC_Os04g46440 OsERF34 3'UTR	GCCAAATCTAGCATGCGAGGTT	TATCACACTGTCCATTTGGCCCA	RT-qPCR
LOC_Os06g06900 OsSPATULA 3'UTR	TGTTTAGTGAAGAGCAGCACGA	GGACCACGAAGATTTCTTCAGT	RT-qPCR
LOC_Os05g29810 OsERF61	CATGCATGCAGCAGCAAGATATACA	GTTAAGCATTGCATTGCCAGGT	RT-qPCR
LOC_Os01g52680 OsMADS32/CFO	CGCTTGACGTTCAGATTTCCT	CTAATACATGACGCCGTGCAG	RT-qPCR
LOC_Os04g45330 OsYABBY5	TAGGGTCCTGCATGTCTC	ATACGATCGATGCATGCC	RT-qPCR
OSUBIQUITIN5	ACCACTTCGACCGCCACTACT	ACGCCTAAGCCTGCTGGTT	RT-qPCR

Supplementary Methods

Data set Preparation for Gene Association of OsMADS1 ChIP-seq Peaks:

A total of 4741 genes were queried from Chip-Seq data (PICS) and matched with both MSU (version 6.0) and RAPDB (build5). Common genes found in both data sets were again filtered using FDR cut-off ≤ 0.055. We extended the search for the OsMADS1-bound sites up to -4kb upstream of the TSS and included only one bound site per gene. Finally we arrived at a data set of 3112 genes, each associated with a single OsMADS1-bound site.

Calculations of A-Tract Motifs and Structural Features

Frequency of A-tracts: Three or more successive A:T base pairs devoid of TpA step defines an A-tract in the OsMADS1-bound sequences(Strahs and Schlick, 2000; McConnell and Beveridge, 2001; Stefl et al., 2004). The overall frequency of such A-tracts was calculated in a sliding window size of 6 nucleotides, and this was done for all three bound data sets (intergenic, gene body, and A-tract). The frequencies were compared to those in the respective shuffled sequences.

AFE calculation: AFE is a measure of overall stability of double helical DNA and is the summation of the free energy of its constituent base paired dinucleotides. The free energy values are taken from the unified parameters for the 10 unique dinucleotide steps obtained from melting studies of 108 oligo- and polynucleotide duplexes (SantaLucia, 1998). A 15 nucleotide sliding window (consists of 14 dinucleotide steps) was used for the AFE calculation (Kanhere and Bansal, 2005; Morey et al., 2011).

Bendability prediction: Bendability or flexibility was calculated for OsMADS1-bound sequences using trinucleotide parameters that predict DnaseI sensitivity (Brukner et al., 1995). An overlapping window of 30 nucleotides was used to measure the bendability as reported earlier (Kanhere and Bansal, 2005).

Curvature calculation: DNA can form a curved structure even in the absence of external factors, because of the intrinsic geometric properties of different dinucleotides. Curvature calculations were carried out using the in-house software NUCRADGEN (Bhattacharya and Bansal, 1988; Bansal et al., 1995) and BMHT model wedge angles, as defined based on gel retardation assays, were used (Bolshoy et al., 1991). The geometric measure d/l_{max} was used to calculate curvature, where d is the end-to-end distance and l_{max} is the total contour length. A 75 nucleotide moving window was used to calculate d/l_{max}.

Enrichment of GO Functional Term:

Functional categories of OsMADS1-bound genes (3112) were analysed using MapMan (Thimm et al., 2004). Differential gene expression data from *OsMADS1*RNAi (Khanday et al., 2013) were combined with Chip-Seq data from this study to extract bound and unbound, positively and negatively affected gene lists. Genes common to ChIP-Seq data set and with >3 fold reduced expression in *OsMADS1* knockdown panicles (2-20mm) yielded a gene list of positively regulated by OsMADS1 in these developing florets. These 292 “positively regulated bound” data set were taken for GO categorization. Genes not present in the ChIP-Seq data set but those that are positively regulated by OsMADS1, by differential expression studies, are categorized as “positively regulated unbound” data set (1740 genes) and were also taken for GO analysis. Similar categorization generated the “negatively regulated bound” data set of 360 genes and “negatively regulated unbound” data set of 3508 genes (Supplementary Fig. S5). We used PageMan application of MapMan to

evaluate statistically significant GO terms as per the hyper-geometric test. *P*-value was calculated from the Z-score value and color map was plotted using MATLAB. GO terms with p-value ≤ 0.05 were considered as significant. Child GO terms related to the parent RNA- regulation of transcription and to hormones of different functional categories were also plotted with their significance values. Results from MapMan analysis are presented in Supplemental data set S3.

Semi Quantitative ChIP- PCR

Semi-quantitative PCR was performed to validate OsMADS1 binding on genes predicted by PICS algorithm. 25 ngs each of input chromatin, OsMADS1 immunoprecipitated chromatin and no antibody control (Mock) was used as template for a PCR of 40 cycles using primers listed in Supplementary Table S1. The PCR end products were analyzed on a 2% agarose gel and the ethidium bromide stained gel was imaged in a GelDoc XR System (Bio-Rad).

Literature Cited

Bansal M, Bhattacharyya D, Ravi B (1995) NUPARM and NUCGEN: software for analysis and generation of sequence dependent nucleic acid structures. Comput Appl Biosci **11:** 281-287

Bhattacharya D, Bansal M (1988) A general procedure for generation of curved DNA molecules. J Biomol Struct Dyn **6:** 93-104

Bolshoy A, McNamara P, Harrington RE, Trifonov EN (1991) Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. Proc Natl Acad Sci U S A **88:** 2312-2316

Brukner I, Sanchez R, Suck D, Pongor S (1995) Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides. EMBO J **14:** 1812-1818

Kanhere A, Bansal M (2005) Structural properties of promoters: similarities and differences between prokaryotes and eukaryotes. Nucleic Acids Res **33:** 3165-3175

McConnell KJ, Beveridge DL (2001) Molecular dynamics simulations of B '-DNA: sequence effects on A-tract-induced bending and flexibility. J Mol Biol **314:** 23-40

Morey C, Mookherjee S, Rajasekaran G, Bansal M (2011) DNA free energy-based promoter prediction and comparative analysis of Arabidopsis and rice genomes. Plant Physiol **156:** 1300-1315

SantaLucia J, Jr. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci U S A **95:** 1460-1465

Stefl R, Wu H, Ravindranathan S, Sklenar V, Feigon J (2004) DNA A-tract bending in three dimensions: solving the dA4T4 vs. dT4A4 conundrum. Proc Natl Acad Sci U S A **101:** 1177-1182

Strahs D, Schlick T (2000) A-Tract bending: insights into experimental structures by computational models. J Mol Biol **301:** 643-663

Thimm O, ÒBläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Mùller LA, Rhee SY and Stitt M (2004) MAPMAN: a user-driven tool to display

genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J **37: 914-939**