Comparative Analysis of JmjC Domain-containing Proteins Reveals the Potential Histone Demethylases in *Arabidopsis* and Rice

Falong Lu$^{1,2}\dagger$, Guanglin Li$^{1}\dagger$, Xia Cui$^{1}\dagger$, Chunyan Liu$^1$, Xiu-Jie Wang$^1$ and Xiaofeng Cao$^1\ast$

$^1$State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, the Chinese Academy of Sciences, Beijing 100101, China;

$^2$Graduate University, the Chinese Academy of Sciences, Beijing 100039, China

Supported by the Hi-Tech Research and Development (863) Program of China (2006AA10A101), the State Key Basic Research and Development Plan of China (2005CB522400), the National Natural Science Foundation of China (30771209 to C.Liu and 30621001 to X.Cao) and the Chinese Academy of Sciences (CXTD-S2005-2) to X.Cao.

† These authors contributed equally to this work.

\* Author for correspondence.

Tel: +86 (0)10 6486 9203;

Fax: +86 (0)10 6487 3428;

E-mail: <xfcao@genetics.ac.cn>.

Running title: JmjC Domain Proteins in *Arabidopsis* and Rice

Received 3 Apr. 2008
Accepted 14 Apr. 2008
Handling editor: Chun-Ming Liu
Abstract

Histone methylation homeostasis is achieved by controlling the balance between methylation and demethylation to maintain chromatin function and developmental regulation. In animals, a conserved JmjC domain was found in a large group of histone demethylases. However, it is still unclear whether plants also contain the JmjC domain-containing active histone demethylases. Here we performed genome-wide screen and phylogenetic analysis of JmjC domain-containing proteins in the dicot plant Arabidopsis and monocot plant rice, and found 21 and 20 JMJC-domain containing proteins, respectively. We also examined the expression of JMJC-domain containing proteins and compared them to human JmjC counterparts for potential enzymatic activity. The spatial expression patterns of the Arabidopsis JmjC domain-containing genes revealed that they are all actively transcribed genes. These active plant JmjC domain-containing genes could possibly function in epigenetic regulation to antagonize the activity of the large number putative SET domain-containing histone methyltransferases activity to dynamically regulate histone methylation homeostasis.

Key words: At JMJ; demethylases; demethylation; histone; JmjC; methylation; Os JMJ.
Introduction

In eukaryotic cells, the N-terminal tails of histones are subjected to various post-translational modifications which play vital roles in chromatin functions, such as transcriptional regulation, chromatin remodeling, DNA repair and cell cycle regulation (Strahl and Allis 2000; Jenuwein and Allis 2001; Kouzarides 2007). Histone methylation is one of the most complex modifications in which one, two or three methyl groups could be added to a lysine. Each of these three types of methylation has different unique biological function and methylation at different sites is also found to play distinct roles. In animals, histone H3K4, K36 and K79 methylation are found to predominantly associate with actively transcribing genes, whereas H3K9, H3K27 and H4K20 methylation are related to silent regions (Martin and Zhang 2005). In addition, arginine residues can also be mono-, asymmetric or symmetric di- methylated by type I or type II protein arginine methyltransferases (Zhang and Reinberg 2001).

Due to the vital roles of methylation in transcription and other cellular function regulations, the methylation status at local chromatins should be tightly controlled. In contrast to histone acetylation which is known to be dynamically regulated by histone acetyltransferases and deacetylases, histone methylation has long been regarded as irreversible process until a breakthrough discovery which demonstrated that lysine demethylase 1 (KDM1), also known as lysine specific demethylase1 (LSD1), confers the enzymatic activity to remove methyl group from mono- and di-methylated but not tri-methylated histone H3K4 (Shi et al. 2004). Compared with the large number of SET domain-containing histone methyltransferases, there are much fewer known genes belong to the same family as KDM1/LSD1, therefore there might exist other types of histone demethylases other than KDM1/LSD1 to keep the balance of methylation homeostasis. Shortly after the identification of KDM1/LSD1, a large family of Jumonji C (JmjC) domain-containing proteins have been discovered as histone demethylases (Tsukada et al. 2006; Yamane et al. 2006). Biochemical analysis revealed that both KDM1/LSD1 and JmjC domain-containing proteins directly reverse histone methylation by oxidative demethylation reaction, however, KDM1/LSD1 requires flavin whereas JmjC domain proteins require Fe(II) and α-ketoglutarate (αKG) as cofactors. Bioinformatic analysis uncovers 30 JmjC domain-containing proteins in human, which could be classified into 7 groups (Klose et al. 2006a). Biochemical evidence from a number of studies indicates that KDMs confer substrate specificity, such as
KDM2A/JHDM1A for H3K36me1/2, KDM5/JARID1 for H3K4me1/2/3, KDM4/JHDM3 for H3K9me2/3 and H3K36me2/3, KDM6 for H3K27me2/3, and KDM3/JHDM2 for H3K9me1/2 (Klose et al. 2006a; Allis et al. 2007). In addition, JMJD6 is shown to be a histone arginine demethylase that targets methylated histone H3R2 and H4R3 (Chang et al. 2007). These histone demethylases have been shown to be involved in transcriptional activation/repression, transcription elongation and genome integrity. The importance of lysine demethylation is illustrated by that impaired KDM3A/JHDM2a, KDM4C/JHDM3C and KDM5C/JARID1C will cause abnormal spermatogenesis, carcinogenesis and X-linked mental retardation, respectively (Cloos et al. 2006; Iwase et al. 2007; Okada et al. 2007).

Except for H3K79 and H4K20 methylation, both lysine and arginine methylation is relatively conserved between plants and animals (Johnson et al. 2004; Zhang et al. 2007). In Arabidopsis and rice, there are 39 and 38 SET domain-containing proteins present in each genome and some of which confer histone lysine methyltransferase activity (Baumbusch et al. 2001; Liang et al. 2003; Ding et al. 2007a; Ding et al. 2007b). In addition, 4 histone arginine methyltransferases and 5 putative histone arginine methyltransferases are found in Arabidopsis (Niu et al. 2007; Pei et al. 2007; Yan et al. 2007). In plants, histone methylation is important for chromatin stability, normal development and cellular memory. For example, SET domain-containing protein KYP/SUVH4 and SDG714, two histone H3K9 methyltransferases in Arabidopsis and rice, respectively, are essential for maintaining DNA methylation at CNG context for chromatin stability (Jackson et al. 2002; Ding et al. 2007b). ATX1, another SET domain-containing histone H3K4 methyltransferase, is essential for floral organ development by activating homeotic genes (Alvarez-Venegas et al. 2003). EFS/SDG8 (early flowering in short days/SET Domain Group 8), a histone lysine methyltransferase, is needed to prevent plant from early flowering by modifying chromatin of Flowering Locus C (FLC), a key flowering time repressor in Arabidopsis (Kim et al. 2005; Zhao et al. 2005; Xu et al. 2007). AtPRMT10 and AtPRMT5, two histone arginine methyltransferases, are involved in FLC dependent flowering time control (Niu et al. 2007; Pei et al. 2007; Wang et al. 2007; Schmitz et al. 2008).

Histone methylation is a conserved mechanism between animals and plants. Lines of evidence suggest that histone demethylases play important roles in plants as their human counterparts in homeostasis of histone methylation. Flowering Locus D (FLD) and its homologs,
homologs of human KDM1/LSD1, are known to repress FLC expression possibly through demethylation of H3K4 (He et al. 2003; Jiang et al. 2007). The emerging importance of JmjC proteins in *Arabidopsis* has been shown by the identification of developmental abnormalities associated with the JmjC loss-of-function mutants. For example, plants impaired of *EarLy Flowering 6* (*At* ELF6/JMJ11) and *Relative of Early Flowering 6* (*At* REF6/JMJ12) are early and late flowering, respectively (Noh et al. 2004); *Maternal Effect Embryo arrest 27* (*At* MEE27/JMJ15) is involved in gametocyte development (Pagnussat et al. 2005) and *Increase in Bonsai Methylation 1* (*At* IBM1/JMJ25), which is related to KDM3/JHDM2 group, is essential to prevent the spreading of silencing from heterochromatin to active genes (Saze et al. 2008). In spite of the important roles of JmjC proteins in developmental regulation in *Arabidopsis*, no biochemical evidence shows that these proteins are active histone demethylases. In addition, rice is one of the most important crops and is a model species for monocots. However, no JmjC proteins have been reported in rice and the importance of lysine or arginine demethylation remains elusive.

Here, we carry out a systematic JmjC domain-containing proteins identification in *Arabidopsis* and rice by phylogenetic analysis. As the structural analysis of KDM4A/JHDM3A shows that the activity and specificity of JmjC proteins are determined by the JmjC domain, we predict whether these proteins are active histone demethylase by focusing on the conserved Fe(II) and αKG binding sites within the JmjC domain. Furthermore, we confirm that these JmjC domain-containing genes are actively transcribed by analyzing spatial expression of them in *Arabidopsis*. We propose that these *Arabidopsis* and rice proteins are conserved and important for histone methylation homeostasis and plant normal development.
Materials and Methods

Sequences of JmjC domain-containing proteins

The amino acid sequences of all JmjC domain-containing proteins from human, *Arabidopsis thaliana* and *Oryza sativa*, were downloaded from UniProt (http://www.ebi.ac.uk/uniprot/), InterPro (http://www.ebi.ac.uk/interpro), SMART (http://smart.embl-heidelberg.de) and TIGR (http://www.tigr.org) databases. Unreported JmjC domain-containing proteins were identified by comparing known JmjC domain-containing proteins with the Refseq protein sequences of above species using BLASTP program. After removing redundant and variant sequences, we included 30 human proteins, 21 *Arabidopsis thaliana* proteins, and 20 *Oryza sativa* proteins in our JmjC domain-containing protein database. The denominations of JmjC domain-containing proteins used in this study were according to the criterions in ChromDB (http://chromdb.org/).

Protein domain analysis

All the domains in the 71 JmjC-domain-containing proteins were analyzed using HMM (Hidden Markov model) algorithm in SMART and Pfam (http://pfam.wustl.edu). After removing the redundant and/or low score (e >1.0E-10) domains, we obtained a full domain architecture of all the JmjC domain-containing proteins by using Perl scripts.

Sequence alignment and phylogenetic tree construction

A multiple sequence alignment of all the JmjC domains was performed using ClustalW (Thompson et al. 1994) with default parameters, then phylogenetic tree of all the JmjC domain-containing proteins were constructed and visualized by software MEGA (version 3.1) using neighbor-joining method (Kumar et al. 2004). The aligned sequences were visualized using GeneDoc software.

Expression of JmjC domain-containing proteins

*Arabidopsis thaliana* wild type Columbia (Col) plants were grown at 23°C under long day condition (16 h light, 8 h dark). Rosette leaf, stem, inflorescence and silique tissues were collected from 6 week-old plants grown in soil. Root was collected from 13-d-old plants planted in MS
plate containing 3% sucrose. Total RNAs were extracted using Trizol reagent and were reverse transcribed with MMLV reverse transcriptase according to the manufacture’s instruction. Semi-quantitative PCR was performed by amplifying these samples for appropriate cycles ranging from 24 to 32 before reached the stationary phase. Primers used for RT-PCR are as follows:

AT3G48430 (At REF6/JMJ12): cx1812 TGTGCCATCCAGGTGGCATTCCA; AT5G04240 (At ELF6/JMJ11): cx1814 CTTTGTCTTCAGACTGGAAGTTGTCCGACATCA, cx1815
TTTCTGAGATCTCTGCTCTCTGACTTC; AT1G09060 (At JMJ24): cx1822 TATCCAGACTGATTTTGTGTCACGCC, cx1823
GGTGGTTGCTCTGCTCTAGTGTGAGA; AT1g63490 (At JMJ17): cx1842 GCAGCAGTAAAAAGGGCAACACCTTGA, cx1843 GGCTCTGTGCTGCTGCGGGA; AT5G06550 (At JMJ22); cx1844 TTGAGAGAAAGTTCCGGTTTTTGGA, cx1845 GAAATTTCAACATTCAACAAATTACTCCTCTGCTA; AT4G21430 (At JMJ28): cx1892 AGCCTGATTCTGGAAGAATTAGGTGTTTGGA, cx1893
TCATGGTTCAACGTCAAATCTTCTCTCTTA; AT4G20400 (At JMJ14): cx1894 CTGGGCAAGAGTTGGATGGTGTTTGA, cx1895
TGATCAAGAGCCCTACAACTTATTACCA; AT3G20810 (At JMJ30): cx1896 TTTTGTTGGAAGCAAGAGCTTGTGTA, cx1897
CGATATTGTCTAGATCAACCTGACTAGATGTGCA; AT5G63080 (At JMJ20): cx1898 GGTTGGGAAAGGTACTTTGACACCTCTA, cx1899 CATTTAATTCATTCTTCTTCTGTGACAGCA; AT5G19840 (At JMJ31): cx1900 TGATAGACAGGAATGAGCCCTGTGTTAGTCTGCA, cx1901
CGAAAGAACCCTGCTTCTCCTCCCTCA; AT4G00990 (At JMJ27): cx1902 CCGGCTCAGGAAGCTGAGAAGATCTAATTA, cx1903
CCACCTTATGCAAGACTGTCATATTCCTC; AT1G11950 (At JMJ26): cx1904 GTAACACTAATGAGCATTCAGTTGTACCTCGA, cx1905
CCTTGTGCACTGACTTGAGATTCGCA; AT1G78280 At JMJ21: cx1906 ATTAACCCTAGGATACTCTCTGCA, cx1907
CCAAACAGTTTTGCTCAACTTGTCCTC; AT3G07610 (At IBM1/JMJ25): cx1908 GTAGCGTTGACTTCTCTCACAAGTGA;
AT2G34880 (At MEE27/JMJ15): cx1910 AGATATGTGAGGATGCATCCATCATGGA, cx1911 CCTCAGTGGCCTGAACAATAAACCGGA; AT1G62310 (At JMJ29): cx1912 TCAGACGCGGTGAATATTCTAAGCGCA, cx1913 GACTTTGTGACACGACTTTAGATTCCGA; AT1G30810 (At JMJ18): cx1914 GAACCTTGTTGTATCGGTCAGCGCTA, cx1915 TTCAATGCGCTGAACTATAAGGGTGAG; AT2G38950 (At JMJ19): cx1916 CCAAAAGAAAGTCATAATCCTGTGATGATGA, cx1917 CAGTTTGATCCGATCTTTAGATTCCGA; AT5G46910 (At JMJ13): cx1918 CGAAAGAAGACGGATATACACCGTATTCCA, cx1919 TCAGTCTCTTTAAGGTCCGTGATGATGTCAGA; at1g08620 At JMJ16: cx1920 ACAGTCCATTGTGCATAGTTACTTGGAGA, cx1921 AAACCTCTTTAAGGTGCTTTAGTTACTTCCA; At3g45880 (At JMJ32): cx1922 CCGAGATACTGATGCCTCTCTCAAGCTTG, cx1923 AACTTGCTTTCCCTCTGCATGACCTGGA.
Results

Domain architecture and evolution of all JmjC domain-containing proteins in plants

The JmjC domain-containing histone demethylases share a highly conserved JmjC domain and comprise a large gene family in plants. After intensive bioinformatics analysis, we obtained a non-redundant set of 71 JmjC domain-containing proteins, including 21 proteins from *Arabidopsis thaliana*, 20 from *Oryza sativa* and 30 from Human. Basing on the JmjC domain sequences, we constructed a phylogenetic tree of human, *Arabidopsis thaliana* and *Oryza sativa* JmjC domain-containing proteins (Figure 1). By merging information between the phylogenetic tree and the domain architecture of each protein, we classified the JmjC domain-containing proteins into eight groups, namely KDM6/JMJD3 group, KDM5/JARID1 group, KDM4/JHDM3 group, KDM3/JHDM2 group, KDM2/JHDM1 group, PHF group, JMJD6 group, and JmjC domain only group. The JmjC domain only group includes proteins with only JmjC domain but not other domains and not belonging to any other groups. It is worth to note that proteins belong to the KDM6/JMJD3 group, KDM2/JHDM1 group and PHF group are missing in both *Arabidopsis* and rice.

To further analyze the JmjC domain-containing proteins in *Arabidopsis* and rice, we focused on domain character and amino acid conservation within the predicted Fe(II) and αKG binding sites so that further information for potential active demethylase activities can be provided. In the following sections we will discuss potential enzymatic activity, target specificity on histone methylation and biological function of each JmjC domain-containing group, in order to provide information for investigating JmjC domain functions both in *Arabidopsis* and rice.

The KDM5/JARID1 group

Human KDM5/JARID1 group proteins are found to be active histone demethylases using H3K4me3, H3K4me2 and H3K4me1 as substrates (Christensen et al. 2007; Iwase et al. 2007; Klose et al. 2007; Lee et al. 2007; Xiang et al. 2007b). In higher plants, 6 *Arabidopsis* and 3 rice proteins were identified as members of this group by phylogenetic analysis. Of these 9 members, two subgroups are divided according to their domain architecture. At JM17 and Os JMJ708 share the same domain architecture with their human counterparts, including the following domains,
JmjN, ARID, JmjC, zf-C5HC2 and PHD. The other seven proteins show extensive similarity to human KDM5/JARID1 family proteins in their JmjC domain. Instead of zf-C5HC2 and PHD domains, six of these proteins have FYRN and FYRC domains at the C terminal. It has been shown that the PHD domain in human KDM5/JARID1C preferentially binds to H3K9me3. Interestingly, FYRN and FYRC domains are normally found in Trithorax and its homologs, a group of histone H3K4 methyltransferases (Finn et al. 2006). It is possible that FYRN and FYRC domains harbor chromatin binding activity and target these enzymes to some specific chromatin regions other than the PHD targeting site H3K9me3. Note that At JMJ19 is a relative short protein that lacks the C terminal part of FYRN and FYRC domains comparing with other proteins.

Crystal structure of the KDM4A/JHDM3A catalytic core sequence shows that Fe(II) is chelated by three absolutely conserved residues (His188, Glu190 and His276) within the JmjC domain. Two additional residues (Thr185 and Lys206) are required for αKG binding (Chen et al. 2006; Klose et al. 2006a). Substitutions of the Glu that binds to Fe(II) to Asp, the Thr that binds to αKG to Tyr or Phe and the Lys that binds to αKG to Arg are compatible with histone demethylation activity (Klose et al. 2006a; Agger et al. 2007; De Santa et al. 2007; Hong et al. 2007; Lan et al. 2007; Xiang et al. 2007a). Successful prediction of demethylation activity of JmjC domain-containing proteins is achieved based on these features (Klose et al. 2006a). So we set out to predict whether Arabidopsis and rice JmjC domain-containing proteins have the demethylation activity according to the conservation of these Fe(II) and αKG binding amino acids. All the proteins but At JMJ19 and Os JMJ708 in KDM5/JARID1 group have conserved amino acids of His(H), Glu(E) and His(H) for Fe(II) and Phe(F) and Lys(K) for αKG binding within the cofactor binding sites (Figure 2), suggesting that they are likely to be active histone demethylases. Based on the highest similarity and domain structure, At JMJ17 shall demethylate histone H3K4 as its human counterparts. Further, we propose that Os JMJ703, At JMJ16, Os JMJ704, At JMJ14, At JMJ18 and At MEE27/JMJ15 are potential active histone H3K4 demethylase that target a different subset of targets. Further investigation of these proteins may extend our knowledge about roles of histone H3K4 methylation in plant development.

**The KDM4/JHDM3 group**

The KDM4/JHDM3 group contains four members in human, three members in Arabidopsis and
five members in rice. These proteins share the same features of JmjN and JmjC domains at their N termini but have different C terminal domains. Human KDM4A/JHDM3A, KDM4B/JHDM3B and KDM4C/JHDM3C have PHD and Tudor domains at the C termini. Tudor domain of KDM4A/JHDM3A has been shown to bind H3K4me3, H3K9me3, H4K20me3 and H3K20me2 (Kim et al. 2006). Proteins of this group have two kinds of C terminal domains in *Arabidopsis* and rice. One is C5HC2 type zinc finger (zf-C5HC2) in At JMJ13, Os JMJ706 and Os JMJ707. The other is 4 tandem repeats of C2H2 type zinc finger (4 x zf-C2H2) in At REF6/JMJ12, Os JMJ705, Os JMJ702, At ELF6/JMJ11 and Os JMJ701. Zf-C5HC2 is a putative DNA binding domain. The 4 x zf-C2H2 domain in At REF6/JMJ12 and At ELF6/JMJ11 could be well modeled to 5 x zf-C2H2 found in GLIoma-associated oncogene homolog 1 (GLI1) protein which has been demonstrated to have a specific DNA binding activity (Pavletich and Pabo 1993). So plant proteins in this group might bind directly to specific DNA sequence, whereas human KDM4/JHDM3 proteins might bind to specific chromatin region by interacting with some specific combination of modified histones. This indicates that these At JMJ and Os JMJ proteins may define a new type of transcriptional repressor or activator but not co-repressor or co-activator.

Human KDM4/JHDM3 family proteins are shown to actively demethylate histone H3K9me2/3 and H3K36me2/3 (Cloos et al. 2006; Klose et al. 2006b; Whetstine et al. 2006). All the *Arabidopsis* and rice proteins of this group have conserved Fe(II) and αKG binding amino acids within the cofactor binding site (Figure 3), suggesting that these proteins are likely to be active histone demethylases, and are likely to target tri- and di- methylated H3K9 and H3K36. Although there is a high sequence similarity between At ELF6/JMJ11 and At REF6/JMJ12, they were found to participate in different genetic pathways in flowering time regulation. At REF6/JMJ12 was shown to control flowering time through regulating acetylation status of *FLC* in the autonomous pathway, whereas At ELF6/JMJ11 functioned in photoperiod pathway (Noh et al. 2004). In particular, in *at ref6/jmj12* mutant, late flowering phenotype was shown to be associated with increased histone acetylation level in *FLC* locus. It is possible that the change of histone acetylation level in *FLC* locus is caused by deregulation of histone methylation. The roles that At REF6/JMJ12 play in histone demethylation are waiting to be elucidated. Further, this may give us new insight into chromatin regulation of *FLC* locus that determines flowering time.
The KDM3/JHDM2 group

The KDM3/JHDM2 group contains four members in human, six members in Arabidopsis and five members in rice, which have a JmjC domain at the C terminal. Some of the members have a putative Ring type zinc finger (zf-Ring) domain ahead of it. The zf-Ring domain is shown to be essential for the demethylation activity of KDM3A/JHDM2A (Yamane et al. 2006). KDM3A/JHDM2A can demethylate H3K9me2 and H3K9me1 but not H3K9me3. When the JmjC domain sequence of these groups were aligned together, we found that At JMJ29, At JMJ26, At IBM1/JMJ25, Os JMJ719, Os JMJ718 and At JMJ27 have the conserved Fe(II) and αKG binding amino acids within cofactor binding site (Figure 4), but At JMJ24, Os JMJ716, At JMJ28, Os JMJ715 and Os JMJ720 do not. It is likely that At JMJ29, At JMJ26, At IBM1/JMJ25, Os JMJ719, Os JMJ718 and At JMJ27 are active histone demethylase targeting H3K9me2 and H3K9me1. Recent report showed that mutation in At IBM1/JMJ25 resulted in DNA methylation spreading from a transposon LINE to nearby coding gene Bonsai and showed pleiotropic developmental phenotypes (Saze et al. 2008). Moreover, the pleiotropic phenotypes can be suppressed by mutation in a histone H3K9 histone methyltransferase KYP/SUVH4. Taken together, these clues strongly suggest that At IBM1/JMJ25 is a histone H3K9 demethylase. However, direct biochemical evidence supporting this hypothesis is still to be determined. It is worth to note that mutation in At IBM1/JMJ25 does not affect DNA methylation at global level. And this also suggests that other members within this group should participate in controlling H3K9 methylation and DNA methylation.

The JMJD6 group

The JMJD6 group contains one protein in human, two in Arabidopsis and two in rice. Human JMJD6 has no other obvious protein domain besides the JmjC domain and is shown to be histone arginine demethylase that can demethylate H3R2me2 and H4R3me2 (Chang et al. 2007). Arabidopsis and rice homologs of JMJD6 have an additional F-Box domain in the N terminal. F-Box domain defines some ubiquitin E3 ligase activity, suggesting a link between protein demethylation and ubiquitination. The F-Box domain is also found in human JHDM1A and JHDM1B. However, the function of links between these two distinct enzymatic activities remains to be defined.
The JmjC domain of *Arabidopsis* and rice proteins of this group differs from that of human JMJD6 because the Thr residue of the first αKG binding site is substituted with Ala or Ser (See arrowhead in Figure 5). Whether such substitution affects their potential αKG binding activity is still unknown. However, in all JmjC domain proteins analyzed so far, there is a bias for Thr (Klose et al. 2006a).

**The JmjC domain only group**

There are eight JmjC domain only proteins in human, four in *Arabidopsis* and five in rice without other recognizable protein domain except for JmjC and JmjN domains. All *Arabidopsis* and rice JmjC domain only proteins but At JMJ31 and Os JMJ713 have conserved Fe(II) and αKG binding amino acids within the cofactor binding site (Figure 6), indicating that these proteins might be active demethylase. However, the substrate preference is unknown. Further investigation on these enzymes might reveal that these proteins have distinct enzymatic activity.

**Spatial expression of *Arabidopsis* JmjC domain-containing proteins**

To determine whether these genes are active in *Arabidopsis*, we used RT-PCR to test their spatial expression in different tissues, including cultured cell, rosette leaf, stem, inflorescence, silique and root. Since at least one primer of each primer pair used in the RT-PCR spans two exons, only spliced cDNA but not genomic DNA can be amplified. The results showed that all these 21 genes were expressed under normal growth condition, indicating that all these genes are active expressing genes (Figure 7). Most of these genes are ubiquitously expressed with a slightly higher expression in rosette leaves and inflorescences. At MEE27/JMJ15 is mainly expressed in siliques and At JMJ16 is mainly expressed in inflorescences, indicating they may confer specialized function at reproductive developmental stage.

**Discussion**

JmjC domain proteins represent a large family of histone demethylases in both animals and plants. Here we found 21 JmjC domain-containing proteins in *Arabidopsis* and 20 in rice. We analyzed the domain structure of these 41 proteins and compared them with human counterparts. Also by analyzing potential Fe(II) and αKG binding site, we predict that 15 *Arabidopsis* and 13 rice JmjC
domain-containing proteins are very likely to confer histone demethylase activity. In addition we demonstrated that the *Arabidopsis* JmjC domain-containing genes are active genes by analyzing their spatial expression patterns.

However, plants lack some groups of proteins found in human including the KDM6/JMJD3, KDM2/JHDM1, and PHF groups. KDM2/JHDM1 is the first identified histone H3K36me2 and H3K36me1 demethylase (Tsukada et al. 2006). Given the fact that histone H3K36 methylation is found to be highly abundant and plays important role in flowering time regulation (Johnson et al. 2004; Zhao et al. 2005; Zhang et al. 2007), we believe that the H3K36 methylation level should be tightly controlled by making a good balancing between methylation and demethylation. The KDM4/JHDM3 group proteins might account for part of demethylation of methylated H3K36. However, KDM4/JHDM3 group proteins in human predominantly demethylate H3K36me3 and weakly demethylate H3K36me2. One possibility is that *Arabidopsis* and rice KDM4/JHDM3 family proteins are also capable of demethylating H3K36me1. The other possibility is that some other JmjC domain proteins demethylate H3K36me1 in *Arabidopsis* and rice. The molecular function of PHF group is still unknown. We shall notice that they showed great similarity to KDM2/JHDM1 group protein in their JmjC domain. Probably they have similar enzymatic activity but target different chromatin contents. The KDM6/JMJD3 group contains three members in human, among which Hs KDM6A/UTX and Hs KDM6B/JMJD3 have been shown to demethylate H3K27me3 and H3K27me2 (Swigut and Wysocka 2007). Histone H3K27 methylation is highly conserved between animals and plants and is required for silencing of homeobox transcription factors that are important for proper patterning during development. H3K27 methylation level is developmentally regulated. For example, *AGAMOUS (AG)*, a floral homeotic gene required for stamen and carpel identity, is silenced in vegetative tissue by histone H3K27 methylation and is activated in the center of floral primordia prior to the initiation of stamen and carpel primordia (Schubert et al. 2006; Ito et al. 2007). As H3K27 demethylation occurs during cell differentiation, there should be active H3K27 demethylases in plants. And possibly they are some At JMJs.

Histone methylation, which is catalyzed by a large group of SET domain proteins or protein arginine methyltransferases, is known to play a great role in plant development (Ng et al. 2007). Roles of JmjC domain-containing proteins in plant development are just emerging. Till now, we
know that At ELF6/JMJ11 and At REF6/JMJ12 are required for proper flowering timing (Noh et al. 2004), At MEE27/JMJ15 is involved in gametocyte development (Pagnussat et al. 2005) and mutation in At IBM1/JMJ25 causes pleiotropic phenotypes (Saze et al. 2008). However, how these proteins control histone methylation and how histone methylation regulated by these proteins controls development is still unknown. Further study of roles these proteins play in development and the biochemical mechanism underlining these in conjugation with studying histone methyltransferases in development shall further extend our knowledge about how histone methylation programs plant development.

References


Baumbusch LO, Thorstensen T, Krauß V, Fischer A, Naumann K, Assalkhou R et al. (2001). The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. Nucleic Acids Res. 29, 4319-4333.


Figure Legend

**Figure 1. Phylogenetic tree of 71 JmjC-domain-containing proteins based on JmjC domain sequences.**

The neighbour-joining phylogenetic tree was constructed by using Poisson correction distance for amino acid sequences in MEGA package. Percentage bootstrap values of 1000 replications (larger than 50%) are shown on internal branches. Proteins from Human, *Arabidopsis thaliana* and *Oryza sativa* are indicated with Hs, At and Os prefix, respectively. Groups in red indicate that they are conserved among human and *Arabidopsis* and rice, groups in green indicate those are not conserved and group in black is JMJC domain only proteins not classified in the above groups and not classified further. The domain architecture for each protein is represented by band chart. Different colors are used for different domains and are indicated at the bottom right of the figure.

Protein database IDs for At JMJ and Os JMJ proteins are: At ELF6/JMJ11 (Q6BDA0), At REF6/JMJ12 (Q9STM3), At JMJ13 (Q9FJS0), At JMJ14 (Q8GUL6), At MEE27/JMJ15 (O64752), At JMJ16 (Q9FRS3), At JMJ17 (Q9SH34), At JMJ18 (Q8W4M0), At JMJ19 (Q8L7T6), At JMJ20 (Q67ZB6), At JMJ21 (Q9M9E8), At JMJ22 (Q67XX3,Q9FG15), At JMJ24 (Q27GN5), At IBM1/JMJ25 (Q9SSE9), At JMJ26 (O65384), At JMJ27 (Q8VYB9), At JMJ28 (Q8H1S7), At
Figure 2. The KDM5/JARID1 group proteins contain potential H3K4 demethylases in *Arabidopsis* and *rice*.

Cofactor binding sites within JmjC domain of KDM5/JARID1 group proteins that from human, *Arabidopsis* and *rice* were aligned. The conserved residues compatible with the demethylation activity within the Fe(II) binding site are highlighted in red and those in the αKG binding site are indicated in blue. Proteins from *Homo sapiens*, *Arabidopsis thaliana* and *Oryza sativa* are indicated with Hs, At and Os prefix, respectively.

Figure 3. The KDM4/JHDM3 group proteins are potential active histone demethylases in *Arabidopsis* and *rice*.

Cofactor binding sites within JmjC domain of KDM4/JHDM3 group proteins from human, *Arabidopsis* and *rice* were aligned. The conserved residues compatible with the demethylation activity within the Fe(II) binding site are highlighted in red and those in the αKG binding site are highlighted in blue. Proteins from *Homo sapiens*, *Arabidopsis thaliana* and *Oryza sativa* are indicated with Hs, At and Os prefix, respectively.

Figure 4. The KDM3/JHDM2 group proteins contain potential H3K9 demethylases in *Arabidopsis* and *rice*.

Cofactor binding sites within the JmjC domain of KDM3/JHDM2 group proteins from human,
Arabidopsis and rice were aligned. The conserved residues compatible with the demethylation activity within the Fe(II) binding site are highlighted in red and those in the αKG binding site are indicated in blue. Proteins from Homo sapiens, Arabidopsis thaliana and Oryza sativa are indicated with Hs, At and Os prefix, respectively.

Figure 5. The Arabidopsis and rice JMJD6 group proteins contain amino acid substitution in the first αKG binding site.

Cofactor binding sites within the JmjC domain of JMJD6 group proteins from human, Arabidopsis and rice were aligned. The conserved residues compatible with the demethylation activity within the Fe(II) binding site are highlighted in red and those in the αKG binding site are indicated in blue. Proteins from Homo sapiens, Arabidopsis thaliana and Oryza sativa are indicated with Hs, At and Os prefix, respectively. The first conserved αKG binding residue is substituted with Ala or Ser in Arabidopsis and rice proteins within this group and is indicated by arrowhead.

Figure 6. The Arabidopsis and rice JmjC domain only proteins contain some potential active histone demethylases.

Cofactor binding sites within the JmjC domain of JmjC domain only group proteins from human, Arabidopsis and rice were aligned. The conserved residues compatible with the demethylation activity within the Fe(II) binding site are highlighted in red and those in the αKG binding site are indicated in blue. Proteins from Homo sapiens, Arabidopsis thaliana and Oryza sativa are indicated with Hs, At and Os prefix, respectively.

Figure 7. Spatial expression pattern of JmjC domain-containing genes in Arabidopsis.

“C” is short for callus. “RL” is short for rosette leaf. “St” is short for stem. “In” is short for inflorescence. “SI” is short for silique. “Rt” is short for root. The genes are arranged as the groups we discussed above and this is also indicated in the right. Actin is used as an internal control that indicates equal amount cDNA is used.
Figure 1.
Figure 5.