J Mol Evol (2007) 64:354–363
DOI: 10.1007/s00239-006-0131-x

JOURNAL OF MOLECULAR EVOLUTION

Molecular Phylogeny, Evolution, and Functional Divergence of the LSD1-Like Gene Family: Inference from the Rice Genome

Qingpo Liu, Qingzhong Xue

Department of Agronomy, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310029, P. R. China

Received: 4 June 2006 / Accepted: 16 October 2006 [*Reviewing Editor:* Dr. Joshua Plotkin]

Abstract. The identification of LSD1-like genes in parasite, green algae, moss, pine, and monocot and dicot species allowed us to trace the phylogenetic history of this gene family. Computational analysis showed that the diversification of members of this family could be dated back to the early stage of plant evolution. The evolution of plant LSD1-like genes was possibly shaped by two duplication events. These proteins, which contain three copies of the LSD1 zinc finger (zf-LSD1) domain within their entire polypeptides and play crucial roles in modulating disease defense and cell death, resulted from the second duplication. A gain of zf-LSD1 domain model was reasonable for explaining the origination of three-zf-LSD1 domain-containing proteins. The zf-LSD1 domain phylogeny showed that the middle (M) and C-terminal (C) domains originated from a common ancestor; the N-terminal (N) domain might be more ancient than the former two. The divergence of the N, M, and C domains was well before the monocot-dicot split. Coevolution analysis revealed that four intramolecular domain pairs, including the N domain and the interregion between the M and the C domains (INTER2), the M and C domain, the N- and C-terminus, and the M domain and C-terminus, possibly coevolved during the evolution of three-zf-LSD1 domain-containing proteins. The three zf-LSD1 domains are evolutionary conserved. Thus, the differences at the N- and C-terminus would be crucial for functional specificity of LSD1 genes. Strong functional constraints should work on the zf-LSD1 domains, whereas reduced functional constraint was found in the INTER2 region. Functional divergence analysis showed that three-zf-LSD1 domain-containing proteins were significantly functionally divergent from those proteins containing only one zf-LSD1 domain, a result demonstrating that shifted evolutionary rates between the two clusters were significantly different from each other.

Key words: Rice — LSD1-like — Phylogeny — Coevolution — Functional divergence

Introduction

In plants, LSD1-like genes are members of a small family whose typical feature is the existence of a novel zinc finger motif termed the zf-LSD1 domain (Dietrich et al. 1997). In *Arabidopsis*, five genes were identified as members of the LSD1-like gene family (Epple et al. 2003): three of five LSD1-like proteins contain two or three highly related zinc fingers. These types of proteins are plant-specific. The remaining two proteins contain only one zf-LSD1 domain and a peptidas_C14, caspase catalytic subunit p20 domain. These proteins are predicted to be metacaspase. Sequence database searches identified five zf-LSD1 domain-containing proteins in two parasite species. Nonetheless, no LSD1-like genes have been found in higher animals, viruses, bacteria, or fungi to date.

Among LSD1-like proteins, those containing three copies of the zf-LSD1 domain were demonstrated to be significantly related to programmed cell death (PCD; Coupe et al. 2004; Dietrich et al. 1997; Epple

Correspondence to: Q. Xue; *email:* qzhxue@zju.edu.cn

et al. 2003; L. Wang et al. 2005). The *Arabidopsis* LSD1 gene negatively regulates PCD in response to a superoxide-dependent signal from cells undergoing hypersensitive response (HR) cell death (Dietrich et al. 1997). Kliebenstein et al. (1999) suggested that LSD1 may limit the spread of cell death by up-regulating copper/zinc superoxide dismutase (CuZn-SOD), leading to detoxification of accumulating superoxide. Interestingly enough, a LSD1-like zinc finger protein, AtLOL1, actually functions as a positive regulator of plant cell death. Overexpression of this gene and the transcription factor AtMYB30 can accelerate the appearance of HR in response to avirulent pathogens (Epple et al. 2003; Vailleau et al. 2002). The rice LSD1 also acts as a negative regulator of cell death, although OsLSD1 is the rice gene with the best homology to AtLOL1. This functional difference may be attributable to the presence of potential N-glycosylation and the phosphorylation sites in AtLOL1 but their absence from OsLSD1 (L. Wang et al. 2005). Expression pattern analysis of AtLSD1, AtLOL1, and OsLSD1 showed that they were constitutively expressed in all examined tissues (Dietrich et al. 1997; Epple et al. 2003; L. Wang et al. 2005). Thus, it seems that the gene expression pattern cannot be arbitrarily used to assess the potential function of putative orthologous genes.

Evidence showed that three-zf-LSD1 domain-containing proteins also played roles in modulating disease defense responses. *lsd1* plants exhibited a lesion mimic phenotype and were resistant to virulent pathogens (Dietrich et al. 1994). In rice, OsLSD1 antisense plants conferred significant resistance against a virulent isolate of blast fungus, the extent of which was speculated to be positively correlated with the high inducibility of defense-related genes (L. Wang et al. 2005). Moreover, overexpression of OsLSD1 can increase the level of chlorophyll b. Interestingly, the metabolism of chlorophyll plays a role in disease response and cell death by stabilizing apoproteins in the thylakoid membrane (Meehan et al. 1996). Regrettably, except for several three-zf-LSD1 domain-containing proteins, few functional analyses have been done on other LSD1-like proteins so far (Suarez et al. 2004). Further functional annotation of those proteins will be of benefit for detailed investigation of their biochemical roles in plants.

The rice genome sequencing project has been completed (International Rice Genome Sequencing Project 2005), providing for the first time a comprehensive overview of the rice LSD1-like gene family. In this study, eight LSD1-like genes were identified in rice. Because investigation of the evolutionary divergence of LSD1-like genes between monocot and dicot species may help us to understand the gene function imposed by evolutionary modification, we further conducted phylogenetic, evolutionary, and functional divergence analyses of the LSD1-like family.

Materials and Methods

Sequence Database Searches and Domain Detection

A PSI-BLAST search against the nonredundant GenBank database was performed using the zf-LSD1 domain consensus sequence (PF06943) as query. The collected *Arabidopsis* and other plant LSD1-like proteins were used to construct a hidden Markov model (HMM) profile, and then a HMMER search of the rice proteome was done. To obtain a full list of rice LSD1-like genes, TBLASTN searches against the rice genomic sequences deposited in RGP, NCBI, and TIGR were also conducted using as query the zf-LSD1 domain sequence. The FGENESH software (http://www.softberry.com) was used to predict the intron/exon structure of putative rice LSD1-like genes. The INTERPROSCAN and SMART programs were employed to detect conserved domains in the LSD1-like proteins.

Multiple Sequence Alignment and Phylogenetic Tree Construction

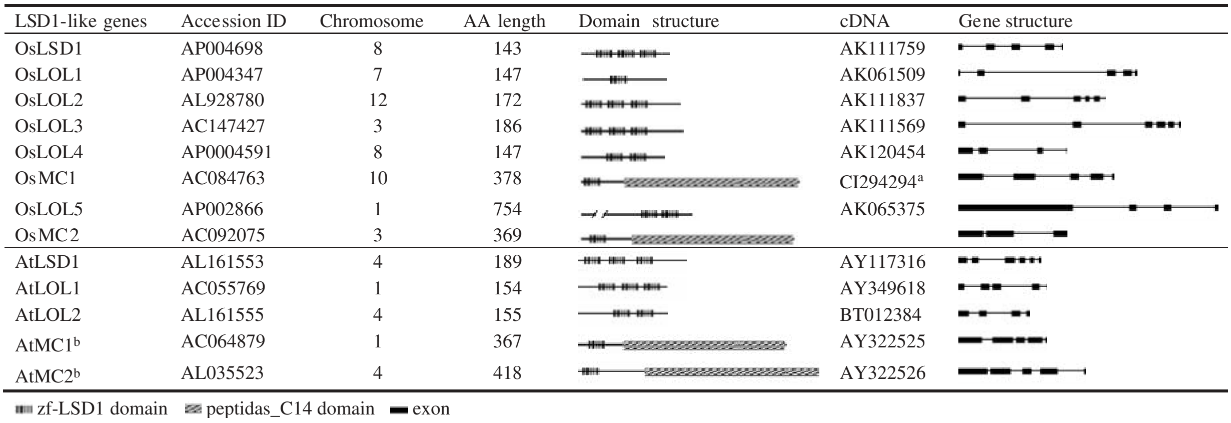
Alignment of amino acid sequences was performed using the CLUSTALW online server with the following parameters: gap opening penalty, 10; gap extension penalty, 1.0; and the PAM series for a protein weight matrix. Phylogenetic trees were constructed with MEGA version 3.1 employing the neighbor-joining (NJ) and minimal evolution (ME) methods, respectively. For both the NJ and the ME methods, the parameters *p*-distance model and pairwise deletion were selected, while two tests of phylogeny, termed the interior branch test and bootstrap test, were performed with 1000 replications each. The NJPLOT program and MEGA v3.1 were used to display phylogenetic trees.

Coevolution Analysis

Goh et al. (2000) chose phosphoglycerate kinase (PGK), a two-domain protein with the enzyme active site formed by the interface between the two domains, as a model system for quantifying coevolution of domains in a single protein. First, a multiple sequence alignment of PGKs from a vast array of species was divided into two independent alignments, one for the N-terminal domain and another for the C-terminal domain. Second, two distance matrices were generated from the multiple alignments using CLUSTALW for each domain. Third, a linear correlation analysis was performed on pairwise evolutionary distances to quantify the coevolution of the two domains. In this study, the full length of three-zf-LSD1 domain-containing proteins was dissected into seven single domains (see below). These domains were independently aligned using CLUSTALW with the default parameters, and pairwise evolutionary distances for every multiple alignments were calculated using MEGA v3.1. Linear Pearson's correlation coefficients ($-1 \leq r \leq 1$) between the distance matrices of all possible interacting domains were calculated accordingly. Positive values of *r* indicate a positive correlation, and *r* values around zero indicate no correlation. Additionally, negative values of *r* indicate anticorrelation.

Calculating of $\mathrm{d_N/d_S}$ *Ratios*

The Yang and Nielsen (2000) method wrapped in the yn00 program of the PAML software package (Yang 1997) was used to calculate the ratio of nonsynonymous to synonymous substitutions



LSD1-like genes	Accession ID	Chromosome	AA length	Domain structure	cDNA	Gene structure
OsLSD1	AP004698	8	143		AK111759	
OsLOL1	AP004347	7	147		AK061509	
OsLOL2	AL928780	12	172		AK111837	
OsLOL3	AC147427	3	186		AK111569	
OsLOL4	AP0004591	8	147		AK120454	
OsMC1	AC084763	10	378		CI294294[a]	
OsLOL5	AP002866	1	754		AK065375	
OsMC2	AC092075	3	369			
AtLSD1	AL161553	4	189		AY117316	
AtLOL1	AC055769	1	154		AY349618	
AtLOL2	AL161555	4	155		BT012384	
AtMC1[b]	AC064879	1	367		AY322525	
AtMC2[b]	AL035523	4	418		AY322526	

▮ zf-LSD1 domain ▨ peptidas_C14 domain ▬ exon

Fig. 1. Chromosomal localization and domain and gene structure of rice and *Arabidopsis* LSD1-like genes. [a]EST sequence. [b]For consistency with the previous nomenclature.

(d_N/d_S). In the three-zf-LSD1 domain-containing proteins, five regions, including the three zf-LSD1 domains and the two intergenic regions among them, were investigated separately. Nucleotide sequences corresponding to amino acid residues of the domains mentioned above were aligned using CLUSTALW. The yn00 program was run to calculate d_N/d_S ratios for each pairwise comparison in a multiple sequence alignment. Thus, an average d_N/d_S ratio for each region was available for comparison. Statistical significance was examined by Student's *t* test.

Estimation of Functional Divergence

DIVERGE, a program developed by Gu and Velden (2002), was used to detect functional divergence between members of a protein family (J. Gu et al. 2002). This software requires the user to input a multiple alignment of amino acid sequences in either FASTA or CLUSTAL format. Next, a phylogenetic tree can be input by the user using the PHYLIP format or generated and rerooted using DIVERGE. Then gene clusters of interest can be selected by simply clicking the internal nodes of the tree. If multiple clusters are selected, DIVERGE performs the statistical analysis (Gu 1999) for all pairs of clusters. The coefficient of type I functional divergence θ and LRT (likelihood ratio statistic) can be quickly calculated. If θ is significantly > 0, it means altered selective constraints of amino acid sites after gene duplication. The estimated θ values for all pairs of clusters can be used to create a matrix of functional distance (d_F) values. Given this matrix, a standard least- squares method can be implemented based on the formula $d_F(A,B) = b_F(A) + b_F(B)$ to estimate b_F for each gene cluster, where $b_F(x)$ is the functional branch length of a given gene cluster x (Gu 2003).

Results

Collection of LSD1-Like Genes

After careful survey of the rice genome, eight LSD1-like genes, including the OsLSD1 gene (L. Wang et al. 2005), were identified (Fig. 1). Besides OsLSD1, the other genes determined in this study were named OsLOL1 to 4, OsMC1, OsLOL5, and OsMC2 according to the PSI-BLAST output order. The reliability of most of the rice LSD1-like genes was supported by six of eight full-length cDNA sequences. BLASTP searches against the GenBank and Swiss-Prot databases also identified five LSD1-like proteins in *Arabidopsis thaliana* (accession numbers CAB79038, AAQ55219, AAS88774, AAP84706, and AAP84707), two in *Brassica oleracea* (AAL50981 and AAL50982), and one in *Brassica rapa* (BAD95789), as well as one partial protein sequence in *Triticum aestivum* (AAP80648) and *Cynodon dactylon* (AAT68199), three in *Trypanosoma cruzi* (EAN92122, EAN93295, and EAN96209), and two in *Trypanosoma brucei* (AAZ11044 and AAZ12483).

The results of domain detection showed that plant LSD1-like proteins may be classified into three major classes according to their zf-LSD1 domain copies. The zf-LSD1 domains are illustrated in Fig. 2. The first class of proteins is those whose entire polypeptide is linearly occupied by three zf-LSD1 domains, such as OsLSD1, OsLOL2, OsLOL3, AtLSD1, and AtLOL1. Proteins belonging to the second class contain two copies (OsLOL4, OsLOL5, and AtLOL2), whereas proteins of the third class contain only one copy of the zf-LSD1 domain (OsLOL1, OsMC1, OsMC2, AtMC1, and AtMC2). Interestingly, with the exception of OsLOL1, proteins belonging to the third class are generally longer than those in the first and second classes, and always contain a peptidas_C14 domain (PF00656; Fig. 1). Thus, it seems that except for OsLOL1, in each class, the rice and *Arabidopsis* LSD1-like proteins may possess similar domain structures. However, *Trypanosoma cruzi* and *Trypanosoma brucei* proteins contain only a zf-LSD1 domain (Fig. 2).

Gene structure comparison showed that the exon number of *Arabidopsis* and rice LSD1-like proteins is from three to six. It is obvious that this difference is apparent among members of the LSD1-like gene family. Notably, except for their differences in exon numbers, the relative exon positions vary significantly also. Compared with *Arabidopsis*, rice LSD1-like genes show clear differences in both intron length and exon numbers (Fig. 1).

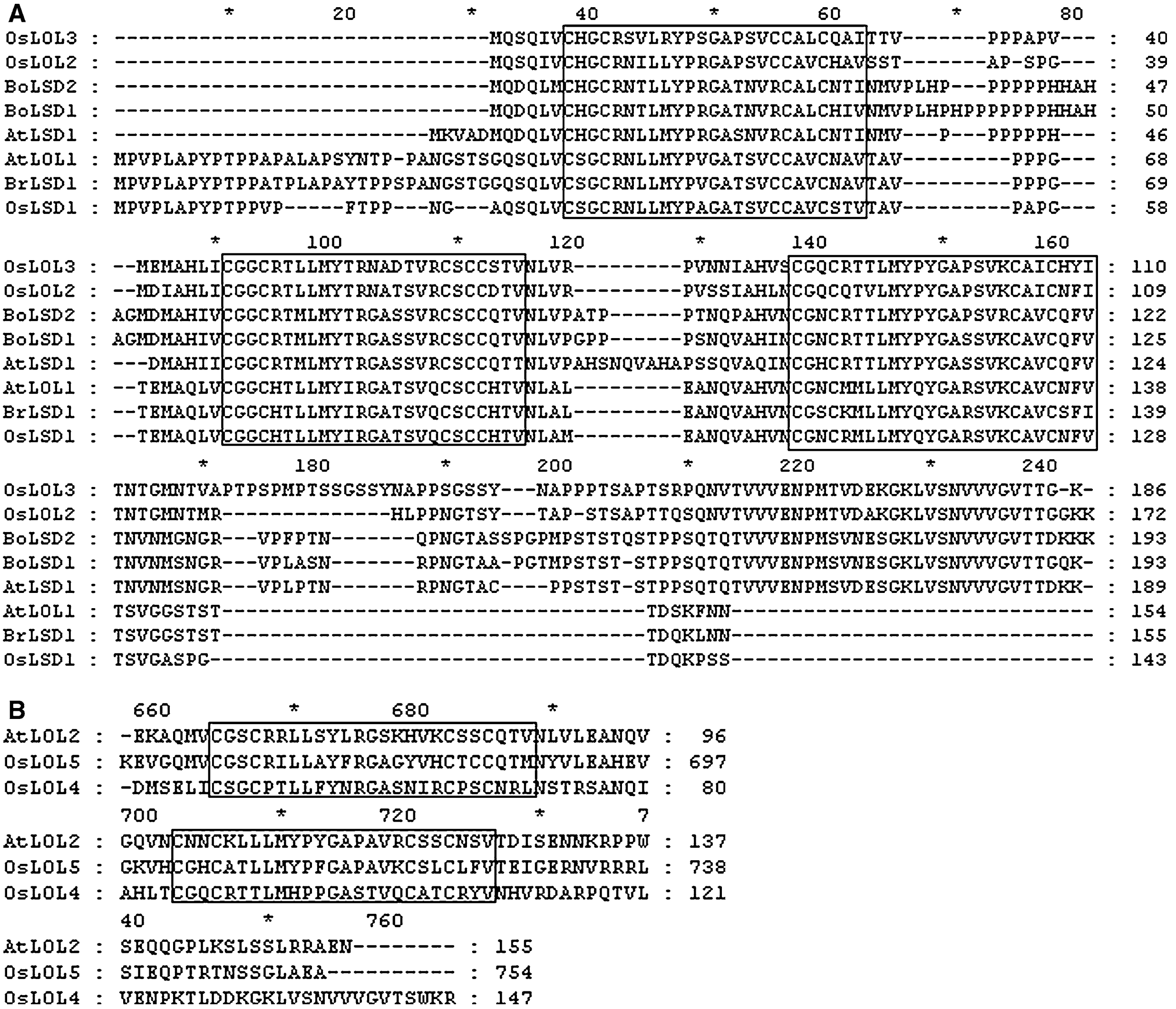
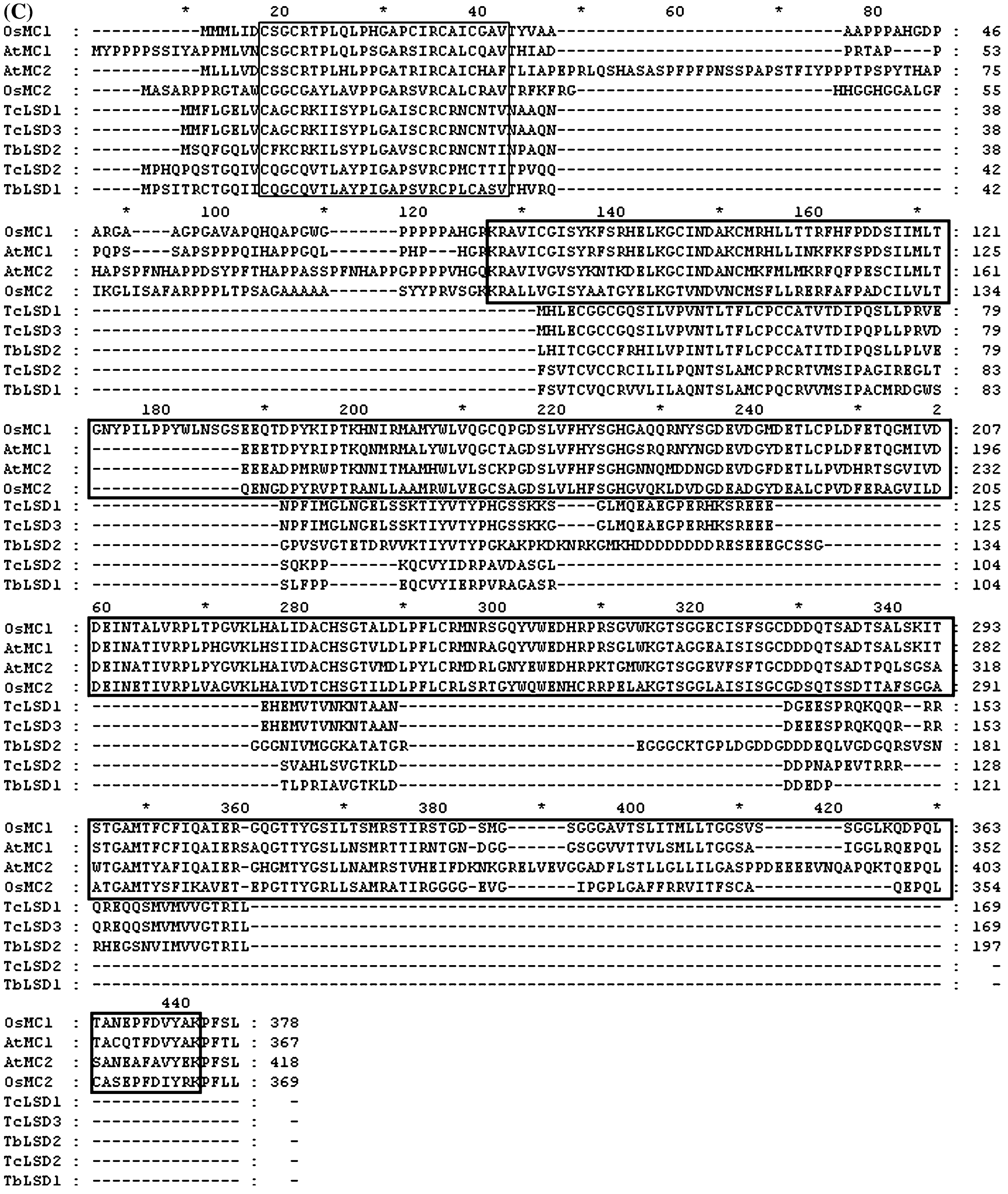

A
* 20 * 40 * 60 * 80
OsLOL3 : -------------------------------MQSQIVCHGCRSVLRYPSGAPSVCCALCQAITTV-------PPPAPV--- : 40
OsLOL2 : -------------------------------MQSQIVCHGCRNILLYPRGAPSVCCAVCHAVSST-------AP-SPG--- : 39
BoLSD2 : -------------------------------MQDQLMCHGCRNTLLYPRGATNVRCALCNTINMVPLHP---PPPPPHHAH : 47
BoLSD1 : -------------------------------MQDQLVCHGCRNTLMYPRGATNVRCALCHIVNMVPLHPHPPPPPPPHHAH : 50
AtLSD1 : --------------------------MKVADMQDQLVCHGCRNLLMYPRGASNVRCALCNTINMV---P---PPPPPH--- : 46
AtLOL1 : MPVPLAPYPTPPAPALAPSYNTP-PANGSTSGQSQLVCSGCRNLLMYPVGATSVCCAVCNAVTAV---------PPPG--- : 68
BrLSD1 : MPVPLAPYPTPPATPLAPAYTPPSPANGSTGGQSQLVCSGCRNLLMYPVGATSVCCAVCNAVTAV---------PPPG--- : 69
OsLSD1 : MPVPLAPYPTPPVP-----FTPP---NG---AQSQLVCSGCRNLLMYPAGATSVCCAVCSTVTAV---------PAPG--- : 58
* 100 * 120 * 140 * 160
OsLOL3 : --MEMAHLICGGCRTLLMYTRNADTVRCSCCSTVNLVR---------PVNNIAHVSCGQCRTTLMYPYGAPSVKCAICHYI : 110
OsLOL2 : --MDIAHLICGGCRTLLMYTRNATSVRCSCCDTVNLVR---------PVSSIAHLNCGQCQTVLMYPYGAPSVKCAICNFI : 109
BoLSD2 : AGMDMAHIVCGGCRTMLMYTRGASSVRCSCCQTVNLVPATP------PTNQPAHVNCGNCRTTLMYPYGAPSVRCAVCQFV : 122
BoLSD1 : AGMDMAHIVCGGCRTMLMYTRGASSVRCSCCQTVNLVPGPP------PSNQVAHINCGNCRTTLMYPYGASSVKCAVCQFV : 125
AtLSD1 : ---DMAHIICGGCRTMLMYTRGASSVRCSCCQTTNLVPAHSNQVAHAPSSQVAQINCGHCRTTLMYPYGASSVKCAVCQFV : 124
AtLOL1 : --TEMAQLVCGGCHTLLMYIRGATSVQCSCCHTVNLAL---------EANQVAHVNCGNCMMLLMYQYGARSVKCAVCNFV : 138
BrLSD1 : --TEMAQLVCGGCHTLLMYIRGATSVQCSCCHTVNLAL---------EANQVAHVNCGSCKMLLMYQYGARSVKCAVCSFI : 139
OsLSD1 : --TEMAQLVCGGCHTLLMYIRGATSVQCSCCHTVNLAM---------EANQVAHVNCGNCRMLLMYQYGARSVKCAVCNFV : 128
* 180 * 200 * 220 * 240
OsLOL3 : TNTGMNTVAPTPSPMPTSSGSSYNAPPSGSSY---NAPPPTSAPTSRPQNVTVVVENPMTVDEKGKLVSNVVVGVTTG-K- : 186
OsLOL2 : TNTGMNTMR--------------HLPPNGTSY---TAP-STSAPTTQSQNVTVVVENPMTVDAKGKLVSNVVVGVTTGGKK : 172
BoLSD2 : TNVNMGNGR---VPFPTN-------QPNGTASSPGPMPSTSTQSTPPSQTQTVVVENPMSVNESGKLVSNVVVGVTTDKKK : 193
BoLSD1 : TNVNMSNGR---VPLASN-------RPNGTAA-PGTMPSTST-STPPSQTQTVVVENPMSVNESGKLVSNVVVGVTTGQK- : 193
AtLSD1 : TNVNMSNGR---VPLPTN-------RPNGTAC----PPSTST-STPPSQTQTVVVENPMSVDESGKLVSNVVVGVTTDKK- : 189
AtLOL1 : TSVGGSTST-----------------------------------TDSKFNN------------------------------ : 154
BrLSD1 : TSVGGSTST-----------------------------------TDQKLNN------------------------------ : 155
OsLSD1 : TSVGASPG------------------------------------TDQKPSS------------------------------ : 143
B
660 * 680 *
AtLOL2 : -EKAQMVCGSCRRLLSYLRGSKHVKCSSCQTVNLVLEANQV : 96
OsLOL5 : KEVGQMVCGSCRILLAYFRGAGYVHCTCCQTMNYVLEAHEV : 697
OsLOL4 : -DMSELICSGCPTLLFYNRGASNIRCPSCNRLNSTRSANQI : 80
700 * 720 * 7
AtLOL2 : GQVNCNNCKLLLMYPYGAPAVRCSSCNSVTDISENNKRPPW : 137
OsLOL5 : GKVHCGHCATLLMYPFGAPAVKCSLCLFVTEIGERNVRRRL : 738
OsLOL4 : AHLTCGQCRTTLMHPPGASTVQCATCRYVNHVRDARPQTVL : 121
40 * 760
AtLOL2 : SEQQGPLKSLSSLRRAEN-------- : 155
OsLOL5 : SIEQPTRTNSSGLAEA---------- : 754
OsLOL4 : VENPKTLDDKGKLVSNVVVGVTSWKR : 147

Fig. 2. Multiple sequence alignments of LSD1-like proteins. (**A**) Three-zf-LSD1 domain-containing proteins; (**B**) two-zf-LSD1 domain-containing proteins; (**C**) one-zf-LSD1 domain-containing proteins. The zf-LSD1 domains are framed by a black box. The peptidas_C14 domain is framed by an overstriking black box.

Phylogenetic Analysis of LSD1-Like Proteins

The phylogenetic trees of LSD1-like proteins generated by the NJ and ME methods showed the same topology (data not show). In this analysis, the partial sequence of the *Triticum aestivum* LSD1-like protein (designated TaLSD) was included in the dataset, whereas the *Cynodon dactylon* LSD1-like protein was discarded because it is too short. BLASTP searches against the GenBank database detected five zf-LSD1 domain-containing proteins in parasites, which extended its presence into another kingdom.

The phylogenetic tree divided, with strong bootstrap support, the plant LSD1-like proteins into two distinct subfamilies, termed the S (single-zf-LSD1 domain-containing proteins) and M (multiple-zf-LSD1 domain-containing proteins) subfamilies, when parasite LSD1-like proteins were used to root the tree (Fig. 3). The M subfamily may be further divided into at least two groups, the D (two-zf-LSD1 domain-containing proteins) and T (three-zf-LSD1 domain-containing proteins) groups; the bootstrap value supporting this topology is high (99%; Fig. 3). Notably, within the plant clade, LSD1-like proteins from monocot and dicot species were not found in distinct groups but, instead, were interspersed, suggesting that the significant expansion of plant LSD1-like genes should have occurred before the divergence of monocots and dicots. In addition, several nucleotide sequences in pine (TC66760, TC73558, TC73563, and TC73556), *Physcomitrella patens* (contig14061, contig3048, contig3049, contig5971, and rpphb22b02), and green algae (BQ80970, AV622717, and BU647788) were found to be LSD1-like genes, an observation that can trace the presence of LSD1-like genes back to the early evolutionary stages of plant development. As for the parasites, it was not difficult to determine that the diversification of their LSD1-like proteins might precede the divergence of *T. cruzi*



```
(C)
                   *        20         *        40         *        60         *        80
OsMC1   : -----------MMMLIDCSGCRTPLQLPHGAPCIRCAICGAVTYVAA-----------------------------AAPPPAHGDP :   46
AtMC1   : MYPPPPSSIYAPPMLVNCSGCRTPLQLPSGARSIRCALCQAVTHIAD-----------------------------PRTAP----P :   53
AtMC2   : -----------MLLLVDCSSCRTPLHLPPGATRIRCAICHAFTLIAPEPRLQSHASASPFPFPNSSPAPSTFIYPPPTPSPYTHAP :   75
OsMC2   : -----MASARPPRGTAWCGGCGAYLAVPPGARSVRCALCRAVTRFKFRG--------------------------HHGGHGGALGF :   55
TcLSD1  : ---------MMFLGELVCAGCRKIISYPLGAISCRCRNCNTVNAAQN--------------------------------------- :   38
TcLSD3  : ---------MMFLGELVCAGCRKIISYPLGAISCRCRNCNTVNAAQN--------------------------------------- :   38
TbLSD2  : ---------MSQFGQLVCFKCRKILSYPLGAVSCRCRNCNTINPAQN--------------------------------------- :   38
TcLSD2  : -----MPHQPQSTGQIVCQGCQVTLAYPIGAPSVRCPMCTTITPVQQ--------------------------------------- :   42
TbLSD1  : -----MPSITRCTGQIICQGCQVTLAYPIGAPSVRCPLCASVTHVRQ--------------------------------------- :   42
             *       100       *       120       *       140       *       160       *
OsMC1   : ARGA----AGPGAVAPQHQAPGWG-------PPPPPAHGRKRAVICGISYKFSRHELKGCINDAKCMRHLLTTRFHFPDDSIIMLT :  121
AtMC1   : PQPS----SAPSPPPQIHAPPGQL-------PHP---HGRKRAVICGISYRFSRHELKGCINDAKCMRHLLINKFKFSPDSILMLT :  125
AtMC2   : HAPSPFNHAPPDSYPFTHAPPASSPFNHAPPGPPPPVHGQKRAVIVGVSYKNTKDELKGCINDANCMKFMLMKRFQFPESCILMLT :  161
OsMC2   : IKGLISAFARPPPLTPSAGAAAAA-------SYYPRVSGKKRALLVGISYAATGYELKGTVNDVNCMSFLLRERFAFPADCILVLT :  134
TcLSD1  : ---------------------------------------------MHLECGGCGQSILVPVNTLTFLCPCCATVTDIPQSLLPRVE :   79
TcLSD3  : ---------------------------------------------MHLECGCCGQSILVPVNTLTFLCPCCATVTDIPQPLLPRVD :   79
TbLSD2  : ---------------------------------------------LHITCGCCFRHILVPINTLTFLCPCCATITDIPQSLLPLVE :   79
TcLSD2  : ---------------------------------------------FSVTCVCCRCILILPQNTSLAMCPRCRTVMSIPAGIREGLT :   83
TbLSD1  : ---------------------------------------------FSVTCVQCRVVLILAQNTSLAMCPQCRVVMSIPACMRDGWS :   83
               180         *       200         *       220         *       240         *       2
OsMC1   : GNYPILPPYWLNSGSEEQTDPYKIPTKHNIRMAMYWLVQGCQPGDSLVFHYSGHGAQQRNYSGDEVDGMDETLCPLDFETQGMIVD :  207
AtMC1   : ---------------EEETDPYRIPTKQNMRMALYWLVQGCTAGDSLVFHYSGHGSRQRNYNGDEVDGYDETLCPLDFETQGMIVD :  196
AtMC2   : ---------------EEEADPMRWPTKNNITMAMHWLVLSCKPGDSLVFHFSGHGNNQMDDNGDEVDGFDETLLPVDHRTSGVIVD :  232
OsMC2   : ---------------QENGDPYRVPTRANLLAAMRWLVEGCSAGDSLVLHFSGHGVQKLDVDGDEADGYDEALCPVDFERAGVILD :  205
TcLSD1  : -------------------NPFIMGLNGELSSKTIYVTYPHGSSKKS----GLMQEAEGPERHKSREEE----------------- :  125
TcLSD3  : -------------------NPFIMGLNGELSSKTIYVTYPHGSSKKG----GLMQEAEGPERHKSREEE----------------- :  125
TbLSD2  : -------------------GPVSVGTETDRVVKTIYVTYPGKAKPKDKNRKGMKHDDDDDDDDRESEEEGCSSG------------ :  134
TcLSD2  : -------------------SQKPP-------KQCVYIDRPAVDASGL--------------------------------------- :  104
TbLSD1  : -------------------SLFPP-------EQCVYIERPVRAGASR--------------------------------------- :  104
          60         *       280         *       300         *       320         *       340
OsMC1   : DEINTALVRPLTPGVKLHALIDACHSGTALDLPFLCRMNRSGQYVWEDHRPRSGVWKGTSGGECISFSGCDDDQTSADTSALSKIT :  293
AtMC1   : DEINATIVRPLPHGVKLHSIIDACHSGTVLDLPFLCRMNRAGQYVWEDHRPRSGLWKGTAGGEAISISGCDDDQTSADTSALSKIT :  282
AtMC2   : DEINATIVRPLPYGVKLHAIVDACHSGTVMDLPYLCRMDRLGNYEWEDHRPKTGMWKGTSGGEVFSFTGCDDDQTSADTPQLSGSA :  318
OsMC2   : DEINETIVRPLVAGVKLHAIVDTCHSGTILDLPFLCRLSRTGYWQWENHCRRPELAKGTSGGLAISISGCGDSQTSSDTTAFSGGA :  291
TcLSD1  : -----------------EHEMVTVNKNTAAN---------------------------------------DGEESPRQKQQR--RR :  153
TcLSD3  : -----------------EHEMVTVNKNTAAN---------------------------------------DEEESPRQKQQR--RR :  153
TbLSD2  : ----------------GGGNIVMGGKATATGR-----------------------EGGGCKTGPLDGDDGDDDEQLVGDGQRSVSN :  181
TcLSD2  : -------------------SVAHLSVGTKLD---------------------------------------DDPNAPEVTRRR---- :  128
TbLSD1  : -------------------TLPRIAVGTKLD---------------------------------------DDEDP----------- :  121
               *       360         *       380         *       400         *       420         *
OsMC1   : STGAMTFCFIQAIER-GQGTTYGSILTSMRSTIRSTGD-SMG------SGGGAVTSLITMLLTGGSVS--------SGGLKQDPQL :  363
AtMC1   : STGAMTFCFIQAIERSAQGTTYGSLLNSMRTTIRNTGN-DGG------GSGGVVTTVLSMLLTGGSA---------IGGLRQEPQL :  352
AtMC2   : WTGAMTYAFIQAIER-GHGMTYGSLLNAMRSTVHEIFDKNKGRELVEVGGADFLSTLLGLLILGASPPDEEEEVNQAPQKTQEPQL :  403
OsMC2   : ATGAMTYSFIKAVET-EPGTTYGRLLSAMRATIRGGGG-EVG-------IPGPLGAFFRRVITFSCA--------------QEPQL :  354
TcLSD1  : QREQQSMVMVVGTRIL---------------------------------------------------------------------- :  169
TcLSD3  : QREQQSMVMVVGTRIL---------------------------------------------------------------------- :  169
TbLSD2  : RHEGSNVIMVVGTRIL---------------------------------------------------------------------- :  197
TcLSD2  : -------------------------------------------------------------------------------------- :    -
TbLSD1  : -------------------------------------------------------------------------------------- :    -
                 440
OsMC1   : TANEPFDVYAKPFSL :  378
AtMC1   : TACQTFDVYAKPFTL :  367
AtMC2   : SANEAFAVYEKPFSL :  418
OsMC2   : CASEPFDIYRKPFLL :  369
TcLSD1  : --------------- :    -
TcLSD3  : --------------- :    -
TbLSD2  : --------------- :    -
TcLSD2  : --------------- :    -
TbLSD1  : --------------- :    -
```

Fig. 2. Continued.

and *T. brucei*; in other words, at least one duplication event had occurred in their common ancestor.

Phylogenetic Analysis of zf-LSD1 Domains

The zf-LSD1 domain sequences were used to construct phylogenetic trees using the NJ and ME methods. The two trees exhibited slightly different topologies, with minor modifications at deep nodes. In the two trees, the parasite zf-LSD1 domains were tightly clustered together, a topology that was supported by a significantly higher bootstrap value in the NJ tree than in the ME tree (91% versus 67%). Accordingly, the NJ tree was mainly used for further analysis (Fig. 4). Our phylogenetic tree showed two distinct clades: the parasite and the plant clades. In the

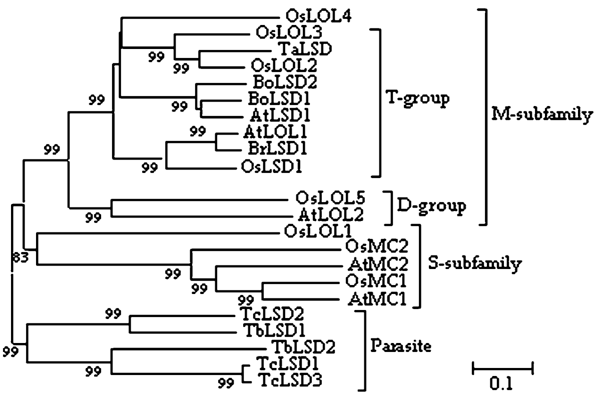


Fig. 3. Phylogenetic analysis of LSD1-like proteins using the neighbor-joining method wrapped in MEGA version 3.1. Numbers at the nodes indicate bootstrap support values (1000 replicates) > 50. The scale is in amino acid substitutions per site. T, three-zf-LSD1 domain-containing proteins; D, two-zf-LSD1 domain-containing proteins; S, one-zf-LSD1 domain-containing proteins. To identify the species of origin for each LSD1-like protein, a species acronym is included before the protein name: At, *Arabidopsis thaliana*; Os, *Oryza sativa*; Ta, *Triticum aestivum*; Bo, *Brassica oleracea*; Br, *Brassica rapa*; Cd, *Cynodon dactylon*; Tc, *Trypanosoma cruzi*; Tb, *Trypanosoma brucei*.

parasite clade, the topology of zf-LSD1 domains (Fig. 4) is as the same as in the tree constructed using full-length amino acid sequences (Fig. 3), suggesting that the phylogenetic analysis of parasite zf-LSD1 domains may reflect the evolution of their proteins. The plant clade can be further divided into four groups, termed the S (single zf-LSD1 domain), N (N-terminal zf-LSD1 domain), M (middle zf-LSD1 domain), and C (C-terminal zf-LSD1 domain) groups, although the bootstrap test of the N-group is not statistically significant (< 50% versus 54%). Nonetheless, the N, M, and C domains of monocot and dicot species were clearly classified into different clades, which implied that their diversification might precede the monocot-dicot split. The M and C groups were evolutionarily close to each other (76% in the ME tree), suggesting that they might share a common ancestor. In these groups, we found that the N- and C-terminal domains of two-zf-LSD1 domain-containing proteins was tightly clustered with the M- and C-terminal domains of three-zf-LSD1 domain-containing proteins, respectively (Fig. 4), a result supporting the common origin of the two classes of proteins. Compared with the M and C groups, the N-group domains should be more divergent, as the bootstrap value, unlike for the former groups, for clustering them together as one clade is low. Further, the N group might be evolutionary close to the S group, although the exact relationship between them would be uncertain because of the low bootstrap value. Notably, from Fig. 4, we can deduce that the diversification of S domains may be no younger than the approximately 200 million-year divergence time between monocots and dicots, unless these relationships are the result of horizontal gene transfer between *Arabidopsis* and rice.

Coevolution Among Domains of Three-zf-LSD1 Domain-Containing Proteins

Three copies of zf-LSD1 domain-containing proteins were used as material to investigate their possible domain-domain coevolution. Their full-length protein sequences were dissected into seven single-peptide domains, representing the N- and C-termini, the three zf-LSD1 domains, and the two interdomain regions. We paired each set of peptide sequences with each of the others and calculated correlation coefficients for all 21 possible pairings. The correlation coefficients obtained ranged from 0.163 to 0.850, with an average of 0.567 and a standard deviation of 0.214. Of all 21 possible pairings, 16 correlation coefficients are significantly greater than zero ($p <$ 0.05). However, Goh et al. (2000) suggested an upper cutoff of 0.8 for correlation coefficients for coevolution of domains in a single protein. Thus, the top five pairings with correlation coefficients higher than or close to 0.8 were further considered (Table 1). The combination of the N-terminal zf-LSD1 domain and the interregion between the middle and the C-terminal zf-LSD1 domains (INTER2) had the highest score (0.850), providing evidence for coevolution of these domains. Interestingly enough, a possible coevolution of the middle and C-terminal zf-LSD1 domains (0.841) was also found, a result supporting the deduction that they might have evolvde from a common ancestor. In addition, the following three pairs also indicated the possible coevolution of the C-terminus with the N-terminus, the middle, and the C-terminal zf-LSD1 domains, respectively. Overall, the coevolution analysis showed particular emphasis on the N-terminal, middle, and C-terminal zf-LSD1 domains and INTER2 (Table 1).

Reduced Functional Constraint in INTER2

To investigate the functional importance of INTER2 and the three zf-LSD1 domains for three-zf-LSD1 domain-containing proteins, their average d_N/d_S ratios were calculated using the Yang and Nielsen (2000) method. When strong functional constraints work on a region, d_S is greater than d_N, resulting in d_N/d_S being close to zero. These d_N/d_S values (0.454, $\sigma = 0.056$, for INTER2; 0.158, $\sigma = 0.021$, for the N-terminal domain; 0.111, $\sigma = 0.014$, for the middle domain; and 0.200, $\sigma = 0.055$, for the C-terminal domain) are < 1, indicating functional constraints working on these regions. For zf-LSD1 domains, it was easily determined that the middle zf-LSD1 domain should have endured more stronger purifying selection for functionality of the three-zf-LSD1 domain-containing proteins. However, it appears that functional constraint is less for INTER2, as the average d_N/d_S ratio for this region is at least two

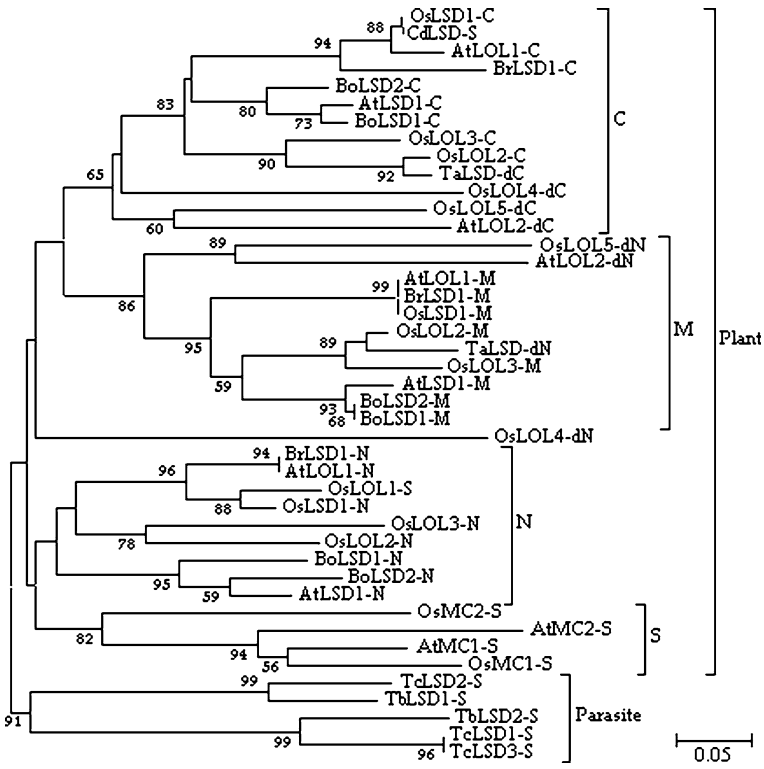


Fig. 4. Phylogenetic analysis of zf-LSD1 domains using the neighbor-joining method wrapped in MEGA version 3.1. Numbers at the nodes indicate bootstrap support values (1000 replicates) >50. The scale is in amino acid substitutions per site. N, M, and C indicate the N-terminal, middle, and C-terminal zf-LSD1 domains of three-zf-LSD1 domain-containing proteins, respectively. dN and dC indicate the N- and C-terminal domains of proteins that contain two copies of the zf-LSD1 domain, respectively. S indicates the single domain of proteins that contain only one copy of the zf-LSD1 domain. To identify the species of origin for each LSD1-like protein, a species acronym is included before the protein name: At, *Arabidopsis thaliana*; Os, *Oryza sativa*; Ta, *Triticum aestivum*; Bo, *Brassica oleracea*; Br, *Brassica rapa*; Cd, *Cynodon dactylon*; Tc, *Trypanosoma cruzi*; Tb, *Trypanosoma brucei.*

times higher than that for zf-LSD1 domains, and these differences are statistically significant (p < 0.001). This result implied that relatively relaxed constraints might work on the nucleotide sequences of this region, leading to the sequence variation found among compiled three-zf-LSD1 domain-containing proteins.

Functional Divergence Between LSD1-Like Proteins

Type I functional divergence between gene clusters of the LSD1-like gene family was estimated by posterior analysis using the program DIVERGE, which evaluates shifted evolutionary rates after gene duplication or speciation (Gu 2003). The estimation was based on a phylogenetic tree where the proteins that contain two zf-LSD1 domains were excluded, as groups with fewer than four sequences cannot be analyzed by this method. We found that the coefficient of type I functional divergence, θ, between any two clusters was statistically significant, which indicated a highly different site-specific altered selective constraint between them (Table 2). Particularly, it appears that the parasite protein cluster was clearly functionally divergent from the two plant gene clusters, suggesting that parasite LSD1-like proteins should have changed their ancestral function to a great extent over a long period of evolutionary time. On the other hand, the functional branch length (b_F) for each cluster was estimated by employing the least-squares method. When the level of altered selective constraint of each cluster was measured by this index, it was found to follow b_F (parasite; 0.797) > b_F (S; 0.636) > b_F (T; 0.245), implying that they may function in a different pattern. However, compared with single-zf-LSD1 domain-containing proteins, the evolutionary conservation of three-zf-LSD1 domain-containing proteins may be shifted at only a few important amino acid sites (0.636 versus 0.245), and their functional divergence should have occurred after the first round of plant LSD1-like gene duplication.

Table 1. Correlation coefficients of the coevolution analysis of domains in three-zf-LSD1 domain-containing proteins

Rank	Pair	Correlation coefficient
1	N domain/INTER2	0.850
2	M domain/C domain	0.841
3	N-terminus/C-terminus	0.807
4	M domain/C-terminus	0.801
5	C domain/C-terminus	0.794

Note. INTER2, the interregion between the middle and the C-terminal zf-LSD1 domains.

Discussion

The LSD1-like family members can be classified into several groups based on the copy numbers of the zf-LSD1 domain within their polypeptides. Proteins containing two or three zf-LSD1 domains are referred to as plant-specific zinc finger proteins, which may function as either transcriptional regulator or

Table 2. Functional divergence analysis of the LSD1-like gene family

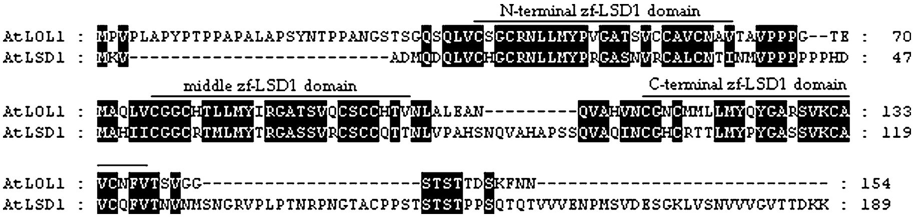
Group 1	Group 2	$\theta \pm$ SE	LRT	p
S	T	0.585 ± 0.159	13.55	< 0.005
S	P	0.762 ± 0.222	11.73	< 0.005
T	P	0.647 ± 0.189	11.68	< 0.005

Note. θ, coefficient of type I functional divergence between two gene clusters; LRT, likelihood ratio statistic; S, single zf-LSD1 domain-containing proteins; T, three zf-LSD1 domain-containing proteins; P, parasite proteins.

scaffold proteins (Dietrich et al. 1997). In contrast, in database searches, one-zf-LSD1 domain-containing proteins were found not only in plants, but also in two pathogenic trypanosomatids that are responsible for severe human diseases. In addition, although several ESTs with a high sequence similarity to LSD1-like genes have been identified in cotton, maize, etc., TBLASTN searches against the dbEST database and BLASTP searches against the nr protein database in NCBI did not find any zf-LSD1 domain-containing proteins in any other species except plant and trypanosomes to date. Interestingly, OsLOL1, whose exact function is unknown, contains only one conserved domain that distinguishes it from other plant one-zf-LSD1 domain-containing proteins. Phylogenetic analysis showed that the zf-LSD1 domain of the putative rice unique LSD1-like protein (OsLOL1) was evolutionarily close to the N-terminal zf-LSD1 domain of OsLSD1 instead of the S-group domains (Fig. 4), whereas the full-length OsLOL1 protein was clustered together with those one-zf-LSD1 domain-containing proteins, which suggested that OsLOL1 should possibly emerge after the divergence of one-zf-LSD1 domain-containing proteins and before the emergence of two-zf-LSD1 domain-containing proteins. Accordingly, it is reasonable that the unique protein structure of OsLOL1 is most likely due to a secondary loss of the peptidas_C14 domain.

LSD1-like protein phylogeny suggested that the diversification of LSD1-like family members should be no younger than the divergence of monocot and dicot species (Fig. 3). As reported, the rice genome has endured two rounds of ancient polyploidy events (Guyot and Keller 2004; Zhang et al. 2005): one occurred before the divergence of cereals, and the other before the monocot-dicot separation. These polyploidy events may act as a major force driving the expansion of gene families. It was found that rice LSD1-like genes were located on chromosomes 1, 3, 7, 8, 10, and 12 without any obvious clustering. With two exceptions (OsLOL5 and OsMC2), all members of the rice LSD1-like gene family were found to be located outside of the regions that were supposed to have undergone previous large-scale duplication events. Further, OsLOL5 and OsMC2 were found as single copies in their duplicated areas. This is not surprising, because massive gene losses and chromosome rearrangements, following large-scale genome duplications, have occurred in rice, leading to a loss of about 30%–65% of duplicated genes (X. Wang et al. 2005). Thus, the expansion of rice LSD1-like genes cannot be easily attributed to either individual gene duplication or chromosomal segmental duplication. Interestingly, several ESTs were identified in pine, moss, and green algae. If these findings cannot be interpreted as the result of horizontal gene transfer, the origin of the LSD1-like gene family would be dated back to the early evolutionary stage of plant development. According to the fact that LSD1-like genes should have been missing from all other eukaryotes and prokaryotes, trypanosome LSD1-like genes might be reasonably explained as the result of horizontal gene transfer from one-zf-LSD1 domain-containing proteins in plants. Another fact supporting this deduction is that they are evolutionarily close to each other. In addition, as reported, *T. cruzi* and *T. brucei* are two human- instead of plant-specific pathogenic trypanosomatids, and they share high genomic sequence similarity. When each species faced the same selective pressures, convergent evolution would have occurred. Thus, it can be speculated that trypanosome LSD1-like genes might evolve in a similar way so as to enhance virulence adaptation to their common host.

When parasite proteins were used to root the tree, plant LSD1-like proteins formed a tightly ordered clade (Fig. 3), where two independent duplications should have occurred: the first duplication gave rise to the S subfamily and a second lineage; the second lineage underwent a second duplication to produce the D- and T-group LSD1-like genes. These two gene duplication events must have occurred in the common ancestor of monocot and dicot species. The putative evolutionary track of plant LSD1-like genes indicated that the gain-of-the-zf-LSD1 domain model may be clearly appropriate for interpreting the evolution of this family in plants. Moreover, LSD1-like amino acid phylogeny identified several orthologous pairs, with significant bootstrap support, e.g., AtMC1/OsMC1, AtLOL2/OsLOL5, and AtLOL1/BrLSD1. It is beneficial for using orthologues with known functions to annotate uncharacterized LSD1-like genes in other species, because orthologues usually share similar functions. Interestingly, it was found that several orthologous pairs, such as OsLSD1/AtLOL1, OsLOL5/AtLOL2, and OsMC1/AtMC1 (Fig. 2), appear to share a similar gene structure (Fig. 1), although the exon/intron structure of rice and *Arabidopsis* LSD-like genes is quite variable in intron length, which implies that there might be a reasonable correlation between gene structure



```
AtLOL1 : MPVPLAPYPTPPAPALAPSYNTPPANGSTSCQSQLVCSGCRNLLMYPVGATSVCCAVCNAVTAVPPPG--TE :  70
AtLSD1 : MKV-------------------------ADMQDQLVCHGCRNLLMYPRGASNVRCALCNTINMVPPPPPPHD :  47

AtLOL1 : MAQLVCGGCHTLLMYIRGATSVQCSCCHTVNLALEAN---------QVAHVNCGNCMMLLMYQYGARSVKCA : 133
AtLSD1 : MAHIICGGCRTMLMYTRGASSVRCSCCQTTNLVPAHSNQVAHAPSSQVAQINCGHCRTTLMYPYGASSVKCA : 119

AtLOL1 : VCNFVTSVGG---------------------STSTTDSKFNN---------------------------- : 154
AtLSD1 : VCQFVTNVNMSNGRVPLPTNRPNGTACPPSTSTSTPPSQTQTVVVENPMSVDESGKLVSNVVVGVTTDKK : 189
```

Fig. 5. Sequence alignment of AtLOL1 and AtLSD1 using the CLUSTALW online server. Identical amino acid residues are shaded in black. The three zf-LSD1 domains are marked.

and phylogenetic relationship of matching rice/*Arabidopsis* gene pairs.

The method developed by Goh et al. (2000) for identifying protein-protein interaction pairs can be adapted to assess the intramolecular coevolution of peptide domains within a single protein family. Pazos and Valencia (2001) found that about 70% true interactions could be detected when the empiric cutoff value (0.8) suggested by Goh et al. (2000) was used for investigating coevolution. This method was also employed by Devoto et al. (2003) to investigate possible intramolecular domain-domain coevolution of the plant-specific MLO gene family. The authors provided evidence for coevolution of all cytoplasmic loops and C-terminal sequences. Interestingly, a recent experiment demonstrated that the interplay of cytoplasmic loops was quite necessary for functionality of the MLO protein (Elliott et al. 2005). The present study suggested that the middle- and C-terminal zf-LSD1 domains may have coevolved. The zf-LSD1 domain phylogeny showed that the evolution of this domain was probably shaped by two duplication events: the first duplication produced the N-terminal domain and the common ancestor of middle and C-terminal domains; the last two domains resulted from the second duplication (Fig. 4), and would evolve in a coordinated manner. In addition, we also found a possible coevolution pair between the N- and the C-termini. As reported, in *Arabidopsis*, AtLSD1 and AtLOL1 function in an antagonistic way for regulating PCD, although they all contain three zf-LSD1 domains (Epple et al. 2003). Sequence alignment showed that the major differences between them were located in the N- and C-termini and the INTER2 region (Fig. 5). Thus, it can be speculated that the regions determined probably to have coevolved should function for specificity of three-zf-LSD1 domain-containing proteins.

A maximum likelihood test of functional divergence was performed on the basis of Gu's (1999) method. The advantage of this method is that it is not sensitive to saturation of synonymous sites. Our results showed that three-zf-LSD1 domain-containing proteins were functionally divergent from those whose polypeptide contains only one zf-LSD1 domain, owing to the shifted evolutionary rate at some important amino acid residues. As demonstrated, three-zf-LSD1 domain-containing proteins may function as a negative (Dietrich et al. 1997; L. Wang et al. 2005) or positive (Epple et al. 2003) regulator of plant cell death. Further, rice OsLSD1 also plays a positive role in callus differentiation (L. Wang et al. 2005). In contrast, the S-subfamily proteins, except for OsLOL1, are putative metacaspases that function in the activation and/or execution of PCD, and are essential for plant embryogenesis (Suarez et al. 2004). Therefore, a reasonable explanation for this difference is that the two types of proteins should have evolved some new subfamily-specific functions over a long evolutionary time, although they have inherited the basic roles related to PCD. In addition, the evolution of D-group proteins was not studied with DIVERGE, as groups with fewer than four sequences cannot be analyzed by this method. To date, no detailed functional analysis has been performed on two-zf-LSD1 domain-containing proteins. Consequently, extensive biological studies are quite necessary to examine their exact biochemical roles, which will be of benefit for further investigation of the precise mechanism of LSD1-like family members in regulating PCD.

Acknowledgments. We thank Dr. Joshua Plotkin and anonymous reviewers for their valuable comments on and suggestions for the manuscript. This work was supported by grant from the China Postdoctoral Science Foundation (20060390348).

References

Coupe SA, Watson LM, Ryan DJ, Pinkney TT, Eason JR (2004) Molecular analysis of programmed cell death during senescence in *Arabidopsis thaliana* and *Brassica oleracea*: cloning broccoli LSD1, Bax inhibitor and serine palmitoyltransferase homologues. J Exp Bot 55:59–68

Devoto A, Hartmann HA, Piffanelli P, Elliott C, Simmons C, Taramino G, Goh CS, Cohen FE, Emerson BC, Schulze-Lefert P, Panstruga R (2003) Molucular phylogeny and evolution of the plant-specific seven-transmembrance MLO family. J Mol Evol 56:77–88

Dietrich RA, Delaney TP, Uknes SJ, Ward EJ, Ryals JA, Dangl JL (1994) *Arabidopsis* mutants simulating disease resistance response. Cell 77:565–578

Dietrich RA, Richberg MH, Schmidt R, Dean C, Dang JL (1997) A novel zinc finger protein is encoded by the *Arabidopsis LSD1* gene and functions as a negative regulator of plant cell death. Cell 88:685–694

Elliott C, Müller J, Miklis M, Bhat RA, Schulze-Lefert P, Panstruga R (2005) Conserved extracellular cysteine residues and cytoplasmic loop-loop interplay are required for functionality of the heptahelical MLO protein. Biochem J 385:243–254

Epple P, Mack AA, Morris VR, Dangl JL (2003) Antagonistic control of oxidative stress-induced cell death in *Arabidopsis* by two related, plant-specific zinc finger proteins. Proc Natl Acad Sci USA 100:6831–6836

Goh CS, Bogan AA, Joachimiak M, Walther D, Cohen FE (2000) Co-evolution of proteins with their interaction partners. J Mol Biol 299:283–293

Gu J, Wang Y, Gu X (2002) Evolutionary analysis for functional divergence of Jak protein kinase domains and tissue-specific genes. J Mol Evol 54:725–733

Gu X (1999) Statistical methods for testing functional divergence after gene duplication. Mol Biol Evol 16:1664–1674

Gu X (2003) Functional divergence in protein (family) sequence evolution. Genetica 118:133–141

Gu X, Velden KV (2002) DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. Bioinformatics 18:500–501

Guyot R, Keller B (2004) Ancestral genome duplication in rice. Genome 47:610–614

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Kliebenstein DJ, Dietrich RA, Martin AC, Last RL, Dangl JL (1999) LSD1 regulates salicylic acid induction of copper zinc superoxide dismutase in *Arabidopsis thaliana*. Mol Plant-Microbe Interact 12:1022–1026

Meehan L, Harkins K, Chory J, Rodermel S (1996) *Lhcb* transcription is coordinated with cell size and chlorophyll accumulation. Plant Physiol 112:953–963

Pazos F, Valencia A (2001) Similarity of phylogenetic trees as indicator of protein-protein interaction. Prot Eng 14:609–614

Suarez MF, Filonova LH, Smertenko A, Savenkov EI, Clapham DH, Arnold S, Zhivotovsky B, Bozhkov PV (2004) Metacaspase-dependent programmed cell death is essential for plant embryogenesis. Curr Biol 14:R339–R340

Vailleau F, Daniel X, Tronchet M, Montillet JL, Triantaphylides C, Roby D (2002) A R2R-MYB gene, AtMYB30, acts as a positive regulator of the hypersensitive cell death program in plants in response to pathogen attack. Proc Natl Acad Sci USA 99:10179–10184

Wang L, Pei Z, Tian Y, He C (2005) OsLSD1, a rice zinc finger protein, regulates programmed cell death and callus differentiation. Mol Plant-Microbe Interact 18:375–384

Wang X, Shi X, Hao B, Ge S, Luo J (2005) Duplication and DNA segmental loss in the rice genome: implications for diploidization. New Phytol 165:937–946

Yang ZH (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS 13:555–556

Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17:32–43

Zhang Y, Xu G, Guo X, Fan L (2005) Two ancient rounds of polyploidy in rice genome. J Zhejiang Univ SCI 6B(2):87–90