IDENTIFICAÇÃO E ANÁLISE FILOGENÉTICA DA FAMÍLIA GÊNICA LSD (*LESION SIMULATING DISEASE*) EM VIRIDIPLANTAE E CARACTERIZAÇÃO DOS GENES IDENTIFICADOS EM SOJA [*GLYCINE MAX* (L.) MERRILL]

CAROLINE CABREIRA

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

IDENTIFICAÇÃO E ANÁLISE FILOGENÉTICA DA FAMÍLIA GÊNICA LSD (*LESION SIMULATING DISEASE*) EM VIRIDIPLANTAE E CARACTERIZAÇÃO DOS GENES IDENTIFICADOS EM SOJA [*GLYCINE MAX* (L.) MERRILL]

Caroline Cabreira

Porto Alegre-RS
2013

Caroline Cabreira

Identificação e análise filogenética da família gênica LSD (L*esion Simulating Disease*) em Viridiplantae e caracterização dos genes identificados em soja [*Glycine max* (l.) merrill]

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul como requisito parcial para obtenção do título de Mestre em Genética e Biologia Molecular

Orientadora: Dra. Maria Helena Bodanese Zanettini
Co-orientadora: Dra. Márcia Pinheiro Margis

PORTO ALEGRE-RS
2013

Os resultados apresentados neste trabalho foram obtidos no Laboratório de Genética Molecular Vegetal e no Laboratório de Evolução Molecular, no Departamento de Genética da UFRGS.

O estudo contou com apoio financeiro do GENOSOJA-CNPq e projeto BIOTECSUR II-MCT.

A meu amado pai, meu herói, meu eterno amor.

“Têm coisas que tem seu valor
Avaliado em quilates, em cifras e fins
E outras não têm o apreço
Nem pagam o preço que valem pra mim
Tenho uma velha saudade
Que levo comigo por ser companheira
E que aos olhos dos outros
Parecem desgostos por ser tão caseira...
Tenho amigos que o tempo
Por ser indelével, jamais separou
E ao mesmo tempo revejo
As marcas de ausência que ele me deixou...
Daz vozes dos outros eu levo a palavra
Dos sonhos dos outros eu tiro a razão
Dos olhos dos outros eu vejo os meus erros
Das tantas saudades eu guardo a paixão
Sempre que eu quero, revejo meus dias
E as coisas que eu posso, eu mudo ou arrumo
Mas deixo bem quietas as boas lembranças
Vidinha que é minha, só pra o meu consumo...”

Gujo Teixeira e Luis Marenco

AGRADECIMENTOS

Às minhas duas queridas orientadoras, professora Maria Helena e professora Márcia, agradeço imensamente toda dedicação, apoio, incentivo e carinho durante esses dois anos. Serei eternamente grata pela oportunidade de desfrutar de todo conhecimento e experiência profissional proporcionada por vocês, que certamente vão além da ciência. Agradeço a confiança depositada em mim ao longo desse período e o esforço constante para proporcionar um local de trabalho agradável com todo estrutura necessária. Tenho certeza que esses dois anos foram de extrema valia para meu crescimento como profissional e como pessoa. Muitíssimo obrigada!

À professora Loreta e seu grupo de pesquisa, agradeço toda a gentileza de me receber em seu laboratório, onde tive todo suporte físico e educacional para que eu pudesse desenvolver as análises filogenéticas. Muito obrigada pela paciência comigo e com meus "genes", que não facilitaram muito para que as análises pudessem ser concluídas. Seus ensinamentos agregaram muito a minha formação e sem o seu auxílio, a filogenia não teria sido possível. Muito obrigada!

Ao meu namorado e em breve "Doutor Alexandro". Não tenho palavras para agradecer todo esforço e apoio dedicado a mim, que muitas vezes fez a questão profissional vir antes do laço afetivo. Foram muitas as conversas, dicas, auxílios, críticas e "discussões" científicas. Obrigada por estar do meu lado e me entender nos momentos difíceis e de insegurança. Tu me fazes ver que a construção de bons resultados na vida é consequência de dedicação e muita entrega.

Agradeço ao amigo e colega de laboratório Lauro, por todo auxílio científico e apoio incansável durante esse período, em que estiveste sempre disponível a ajudar. Nossas conversas profissionais e pessoais sempre me fizeram pensar e muito somaram para eu chegar até aqui. Obrigada por tudo!

As "sojinhas" do Laboratório de Genética Vegetal, em especial as queridas Bea, Graci e Ci, que se tornaram mais do que colegas. Agradeço de coração toda ajuda, conversa e conselhos profissionais e pessoais.

Aos colegas dos grupos “MM”, “bactérias” e “bromélias” pela agradável companhia e convivência diária no laboratório e nos deliciosos almoços do RU. As “queris” Graci, Fer L, Leila e Camille, agradeço a companhia no chimarrão, as inúmeras conversas e muitas risadas. Nossa convivência tornou minha vida em Porto Alegre muito mais divertida!

Agradeço as fontes financiadoras desse trabalho, GENOSOJA/CNPq e projeto BIOTECSUR III/MCT.

Ao Programa de Pós Graduação em Genética e Biologia Molecular, que nos proporciona uma formação de excelente qualidade e reconhecimento acadêmico. Ao Elmo, pelo auxílio e imensa disponibilidade sempre. Aos professores do PPGBM, PPG em Biologia Celular e Molecular e PPG em Fitotecnia da UFRGS que contribuíram para a melhoria da minha formação. A banca avaliadora da minha dissertação, Dra. Andréia Turchetto, Dr. Giancarlo Pasquali e Dr. Rogério Margis, agradeço a disponibilidade e as sugestões dadas a este trabalho.

As amigas que a UNISINOS me trouxe Tizi, Aline e Cíntia. As amizades que ficaram na querência, aos amigos do PTG Rancho da Amizade, por não deixar a distância ultrapassar nosso laço afetivo.

A grande família Cabreira, a qual eu não trocaria por nada. Agradeço ao meu primo Maninho, que é a minha família na capital dos pampas. Em especial agradeço aos meus irmãos Jeca e Ito: vocês são exemplos a serem seguidos. Minhas queridas cunhadas Mi e Sil, por serem tão amigas e fazerem parte da minha vida. Meus adoráveis sobrinhos Gui, Leo, Dudu e Matheus: o sorriso e o amor de vocês me fazem ver que todo esse esforço vale SIM à pena.

Por fim, meu maior e mais puro agradecimento, a minha amada Mãe Judite, meu grande exemplo de cordialidade, coragem, determinação, dedicação e caráter. Sem o seu amor e apoio nada disso seria possível. Muito obrigada por me ensinar valores essenciais de amor, respeito, lealdade e gratidão. Essa vitória é nossa! A senhora dedico meu amor infinito.

SUMÁRIO

LISTA DE ABREVIATURAS

aa: aminoácidos

ASR: A*sian Soybean Rust*

At: *Arabidopsis thaliana*

Ca: Cálcio

CONAB: Companhia Nacional de Abastecimento

CuZnSOD: enzima detoxificadora de espécies reativas de oxigênio (*copper zinc superoxide dismutase*)

ETI: *effector-triggered immunity*

EST: *expressed sequence tag*

Gm: *Glycine max*

HR: *hypersensitive response*

H_2O_2: peróxido de hidrogênio

hpi: *hours pos inoculation*

Kb: quilo pares de bases

LSD: L*esion Simulating Disease*

MAMP: *microbe-associated molecular pattern*

MAPK: *mithogen-activated Protein Kinase*

NB-LRR: *nucleotide-binding domain-leucine-rich repeat*

Mb: mega bases

Osa: *Oryza sativa*

PCD: *programmed cell death*

Psa: *Pisum sativa*

PRR: *pattern recognition receptors*

ROS: *reactive oxygen species*

RPK: *receptor-like kinases*

Rpp: genes de resistência

RT-qPCR: trascrição reversa associada a PCR quantitativa (do inglês *reverse transcription-quantitative polymerase chain reaction*)

SA: *salicylic acid*

SAR: *systemic acquired resistance*

RESUMO

LSD (*Lesion Simulating Disease*) é uma família de proteínas dedo de zinco descritas como importantes na resposta hipersensitiva (HR – *hipersensitive response*), limitando a área da lesão e a morte celular programada (PCD – *programmed cell death*), regulando negativamente as espécies reativas de oxigênio (ROS – *reative oxygen species*) geradas em condições de estresse biótico e abiótico. Até o presente momento, a maioria dos estudos incluindo genes *LSD* foi realizada em *Arabidopsis thaliana*, havendo poucos relatos em outros organismos. Além disto, a identificação completa dos genes *LSD* não foi ainda descrita, o que dificulta a reconstrução da história evolutiva desta família. Assim, os objetivos deste estudo foram identificar sequências de genes *LSD* em diferentes genomas, realizar a análise filogenética dessa família e avaliar o perfil de expressão de genes *GmLSD* em diferentes órgãos de soja e sob condições de estresse. Nossos resultados permitiram a identificação de 117 possíveis genes *LSD* exclusivamente em Viridiplantae. Os organismos basais *Volvox carteri* e *Chamydomonas reinhardtii* apresentaram uma cópia de genes *LSD*. Além disso, a expansão do número de cópias gênicas parece ter ocorrido no clado Embriófita. As sequências identificadas foram analisadas quanto à presença do domínio dedo de zinco característico, sendo encontradas proteínas com uma, duas e três cópias de domínios LSD. Foi observada uma ampla conservação das três sequências de domínios LSD, indicando a manutenção destas ao longo da evolução da família. A análise filogenética mostrou que a arquitetura de proteínas com três domínios LSD representa a condição mais ancestral, sendo presente desde os organismos basais. As proteínas com dois domínios LSD foram identificadas somente no clado Embriófita e proteínas que possuem um domínio LSD foram altamente representadas no clado gramíneas. Os genes de monocotiledôneas e dicotiledôneas não formaram grupos separados, indicando que a expansão da família LSD foi anterior à diversificação entre esses clados. A análise de expressão dos genes *GmLSD* em diferentes órgãos de soja mostrou que estes tem expressão variável dependente do órgão. Foi avaliada a expressão dos genes *GmLSD* em resposta a *Phakopsora pachyrhizi*, o agente causador da ferrugem asiática da soja (ASR – A*sian Soybean Rust*). A abordagem de RNA-seq (no qual foi analisada somente a região da lesão) permitiu a identificação de *GmLSD1*,

GmLSD3 e *GmLSD4* nos genótipos suscetível (EMBRAPA48) e resistente (PI561356), enquanto que *GmLSD6* e *GmLSD8* foram detectados somente no genótipo resistente PI561356. Por outro lado, o experimento de RT-qPCR (em que foram analisadas folhas de soja infectadas por *P. pachyrhizi*) revelou que todos os genes *Gm*LSD foram responsivos a inoculação do fungo, embora *GmLSD1*, *GmLSD4* e *GmLSD8* tenham sido modulados exclusivamente no genótipo resistente PI561356. A análise do perfil de expressão dos genes *GmLSD* em raízes e folhas de plantas de soja em condições de déficit hídrico mostrou que a maioria dos genes *GmLSD* foi modulada em resposta ao estresse. No entanto os genes *GmLSD*5 (folhas), *GmLSD1* e *GmLSD2* (raízes) foram diferencialmente expressos somente na cultivar tolerante EMBRAPA48. Esses resultados sugerem um possível envolvimento dos genes *GmLSD* na PCD causada por estresses bióticos e abióticos, assim como observado em *Arabidopsis thaliana*. Análises futuras com os genes *GmLSD* são particularmente interessantes para avaliar o potencial biotecnológico destes genes, sobretudo visando o desenvolvimento de plantas de soja mais tolerantes/resistentes à diferentes condições de estresse.

ABSTRACT

LSD (Lesion Simulating Disease) is a family of zinc finger proteins described as important in the hypersensitive response (HR), limiting the area of injury and the programmed cell death (PCD), negatively regulating the Reactive oxygen species (ROS) generated under biotic and abiotic stress conditions. To date, most studies including *LSD* genes were performed in *Arabidopsis thaliana*, with few reports in other organisms. Moreover, the complete identification of *LSD* genes has not yet been described, which difficult the reconstruction of the evolutionary history of this family. The goals of this study were to identify *LSD* gene sequences in different genomes, perform a phylogenetic analysis of this family and evaluate the expression profile of *GmLSD* genes in different soybean organs and under stress conditions. Our results allowed the identification of 117 putative *LSD* genes exclusively in Viridiplantae. *Volvox carteri* and *Chamydomonas reinhardtii* basal organisms showed one copy of *LSD* gene. Furthermore, the expansion of the number of genes copies seems to be occurred in Embryophyte clad. The identified sequences were analyzed for the presence of the characteristic zinc finger domain, being found proteins with one, two and three copies of LSD domains. We observed a wide conservation of the three sequences of LSD domains, indicating their maintenance throughout the evolution of the family. Phylogenetic analysis showed that the architecture of proteins with three LSD domains is the most ancestral condition, being present since the basal organisms. Proteins with two LSD domains have been identified only in Embryophyte clad and proteins having one LSD domain were highly represented in the Grass clad. The monocots and dicots genes not formed separate groups, indicating that the expansion of the LSD family was prior to diversification between these clades. The expression analysis of *GmLSD* genes in different soybean organs showed that these having an organ dependent variable expression. We evaluated the expression of *GmLSD* genes in response to *Phakopsora pachyrhizi*, the Asian Soybean Rust (ASR) causal agent. The RNA-seq approach (wherein only the lesion site was analyzed) allowed the identification of *GmLSD1*, and *GmLSD3 GmLSD4* in both susceptible (EMBRAPA48) and resistant (PI561356) genotypes, while *GmLSD6* and *GmLSD8* were detected only in PI561356 resistant genotype. Moreover, the RT-qPCR experiment (in which soybean leaves infected by *P. pachyrhizi* were analyzed) showed that all *GmLSD* were

responsive to fungus inoculation, although *GmLSD1*, *GmLSD4* and *GmLSD8* were differentially expressed only in PI561356 resistant genotype. The analysis of *GmLSD* genes expression profile in roots and leaves of soybean plants under water deficit conditions showed that the majority *GmLSD* genes were modulated in response to stress. However, *GmLSD5* (leaves), and *GmLSD1 GmLSD2* (roots) genes were differentially expressed only in EMBRAPA48 tolerant cultivar. These results suggest a putative involvement of *GmLSD* genes in the PCD caused by biotic and abiotic stresses, as observed in *Arabidopsis thaliana*. Further analyzes with *GmLSD* genes are particularly interesting to evaluate the biotechnological potential of these genes, especially for the development of soybean plants more tolerant / resistant to various stress conditions.

CAPÍTULO I

INTRODUÇÃO GERAL

1. INTRODUÇÃO GERAL

1.1 SOJA

Na safra 2011/12, o Brasil plantou sua maior área de grãos da história, alcançando 51 milhões de hectares, com produção recorde de 166 milhões de toneladas (Reetz 2012). Previsões indicam que a produção agrícola brasileira vai crescer mais durante a próxima década do que qualquer outra no mundo, aumentando em mais de 40% até o ano de 2019 (Tollefson, 2010). O alto investimento em pesquisa, tanto por parte de setores públicos quanto privados vem impulsionando a ampliação da área agrícola, o incremento na produtividade, o aumento da qualidade do grão e a redução dos custos de produção (Reetz 2012).

A soja [*Glycine max* (L.) Merrill] é uma espécie anual, de autofecundação, pertencente à família Fabaceae e representa uma cultura de grande importância no cenário mundial. A espécie originou-se no noroeste da China durante o século XI a.C. (Goellner et al., 2010). O cultivo em nosso país iniciou-se em 1960, sendo que em duas décadas, a cultura tornou-se a principal lavoura brasileira. Embora atualmente o Brasil seja o segundo maior produtor deste grão, atrás apenas dos EUA, estima-se que na safra de 2012/2013 o país irá assumir a liderança em produção (Reetz, 2012). A cultura é utilizada para prover alimentos para humanos e para animais, e mais recentemente, como matéria-prima na produção de biodiesel (www.biodiesel.gov.br).

Segundo dados da CONAB (www.conab.gov.br), a produção brasileira de soja na safra 2010/2011 foi estimada em 75,04 milhões de toneladas, mantendo o ritmo de crescimento das últimas safras. Este volume é 9,2% superior à produção obtida na safra 2009/10, quando foram colhidas 68,69 milhões de toneladas. O fator climático foi o principal responsável por este resultado. No entanto, a safra brasileira de 2011/2012 foi marcada pela estiagem, resultando numa produção 11,9% menor do que na safra anterior, sendo que a região Sul foi a mais afetada pelo déficit hídrico (Reetz, 2012).

Devido a sua grande importância, a comunidade científica internacional recomendou a soja, além de *Lotus japonicus* e *Medicago truncatula* como plantas modelo para estudos genéticos e moleculares em leguminosas (Gepts et al., 2005). Recentemente, foi realizado o sequenciamento do genoma da cultivar Willians 82 de soja, compreendendo 950 Mb, o que representa uma cobertura de 85% do genoma total da espécie (Schmutz et al., 2010). Contendo 20 cromossomos, foram preditos 46.430 genes que codificam proteínas, ou seja, 70% a mais do que é predito para *A. thaliana*. Dois eventos de duplicação do genoma parecem ter ocorrido e estes datam de 59 e 13 milhões de anos, resultando em um genoma altamente duplicado, com 75% dos genes apresentando-se em múltiplas cópias. Segundo Schmutz et al. (2010), os dois eventos de duplicação do genoma foram seguidos por inúmeros rearranjos cromossômicos, perdas e diversificação gênica. Estes resultados vêm possibilitando uma série de estudos visando uma maior compreensão das respostas da planta de soja às inúmeras alterações ambientais, bem como um melhor entendimento da biologia do organismo como um todo.

1.2 AMEAÇAS ÀS CULTURAS AGRÍCOLAS: ESTRESSES BIÓTICOS E ABIÓTICOS

Embora o cenário agrícola seja bastante positivo, inúmeros estresses ambientais afetam a agricultura, ocasionando perdas severas em produtividade e rendimento. Em se tratando da cultura da soja, muitos fatores podem ocasionar perdas na lavoura, podendo-se citar os estresses bióticos ocasionados por vírus, nematóides, insetos, fungos, bactérias, ervas daninhas e os estresses abióticos como seca ou alagamento, salinidade, alta/baixa temperatura e toxicidade por minerais no solo (Dita et al., 2006). Esses estresses podem não ocorrer isoladamente na planta, visto que plantas sob condições de estresse abiótico tornam-se mais suscetíveis aos danos causados por ervas daninhas, insetos e doenças, aumentando consideravelmente as perdas. A baixa variabilidade genética do germoplasma da soja é responsável por grandes limitações na busca por genótipos mais tolerantes/resistentes à estresses ambientais (Barros e Borges, 2007; Barros, 2009).

No que se refere aos estresses bióticos, estima-se que em torno de 40 doenças causadas por fungos, bactérias, nematóides e vírus já foram observadas na lavoura de

soja no Brasil (Vidor et al., 2004). Entre diversos patógenos, pode-se destacar *Fusarium semitectum, Colletotrichum dematium* var. truncata, *Cercospora kikuchii*, que ocasionam danos à semente, *Rhizoctonia solani*, que afeta tanto plântulas quanto raízes e folhas, *Sclerotium rolfsii*, que ataca tanto plântulas quanto raízes, *Sclerotinia sclerotiorum*, que pode afetar hastes, vagens e pecíolos. Embora a utilização de genótipos com resistência a estresses seja a melhor estratégia para a diminuição de danos, para a grande maioria das doenças ainda não há cultivares resistentes disponíveis.

Atualmente, uma das maiores ameaças à cultura da soja é a ferrugem asiática da soja (ASR), causada pelo basidiomiceto biotrófico *Phakopsora pachyrhizi*. Acredita-se que o fungo originou-se na Ásia, espalhando-se para África, posteriormente para América do Sul e finalmente colonizando a América do Norte (Panthee et al., 2007). No Brasil, a doença foi identificada na safra de 2002/2003 e perdas em torno de dois milhões de reais foram relatadas nesse primeiro período de infecção (Goellner et al., 2010). As perdas em rendimento de grãos podem variar em 10% a 80%, dependendo do estágio da planta em que a doença se estabelece e das condições climáticas da safra (Vidor et al., 2004; Goellner et al., 2010). Quanto mais cedo o patógeno incidir na cultura, maiores são os danos em produtividade (Reetz, 2012). O estabelecimento da doença causa rápida diminuição da área foliar e consequente desfolhação prematura, o que se correlaciona negativamente com os níveis de produtividade desejados (Hartman et al., 1991). A facilidade de colonização do patógeno encontra-se no fato de que este é disseminado pelo vento (Twizeyimana et al., 2011) e períodos longos de umidade favorecem o estabelecimento da infecção (Goellner et al., 2010).

P. pachyrhizi é capaz de infectar mais de 150 espécies de plantas de mais de 53 gêneros, incluindo a soja e outras espécies da família Fabaceae (Hartman et al., 2011), o que confere ao patógeno uma ampla distribuição e variabilidade. O melhoramento para resistência à *P. pachyrhizi* vem sendo realizado baseado nos três fenótipos de resposta possíveis observados em hospedeiros: a) reação de susceptibilidade, caracterizada por lesões de coloração marrom (denominadas lesões TAN), decorrentes da esporulação das urédias; b) reação de hipersensibilidade (HR – do inglês *hipersensitive response*), caracterizada por lesões avermelhadas (RB – do

inglês *red brown*), decorrentes da morte celular programada (do inglês, *programmed cell death* – PCD) e c) resposta imune (Figura 1) (Miles et al., 2006; Goellner et al., 2010). Genes *Rpp*, capazes de conferir resistência a alguns isolados de *P. pachyrhizi* já foram identificados (Cheng e Chan, 1968; Hidayat e Somaatmadia, 1977; Singh e Thapliyal, 1977; Bromfield e Melching, 1982; Hartwig, 1986; Garcia et al., 2008; Li et al., 2012). No entanto, Hartmann et al. (2011) sugerem que a interação de genótipos de soja com o fungo parece seguir a teoria gene a gene, em que genes de resistência conferem resistência a isolados específicos do patógeno. Assim, a obtenção de cultivares resistentes à ASR tem sido bastante dificultada, uma vez que a efetividade da maioria dos *Rpp* identificados já foi quebrada por diferentes isolados do patógeno ao redor do mundo (Twizeyimana et al., 2011).

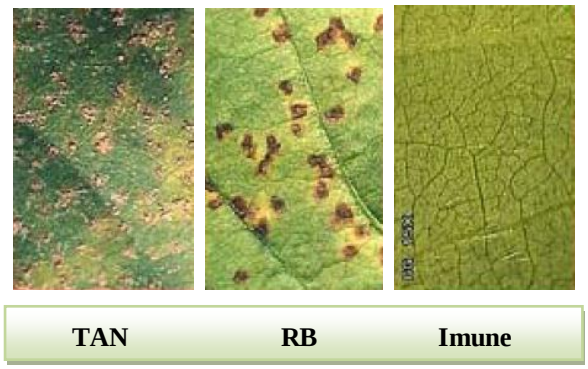


Figura 1: Fenótipo de resposta observada em folhas de soja inoculadas com *P. pachyrhizi.* TAN: reação de susceptibilidade; RB: resposta de hipersensibilidade, ocasionada por PCD; Imune: sem lesões visíveis. Modificado de Miles, Frederick e Hartman (2006).

Quanto aos estresses abióticos, a seca é um dos principais fatores que afetam a produtividade (Le et al., 2011a). Esse estresse afeta a cultura da soja em todos os estágios de desenvolvimento e crescimento, apresentando impacto direto na qualidade da semente, podendo acarretar perdas significativas de até 40% (Tran e Mochida, 2010). No que se refere à fixação de nitrogênio em nódulos, a planta também é prejudicada pelo déficit hídrico (Ladrera et al. 2007).

Em condições de déficit hídrico, folhas e raízes são importantes sinalizadores de dano, visto que o desbalanço entre a transpiração nas folhas e a captação de água pelas raízes são sinais do início do estresse (Aroca et al., 2011). A percepção de déficit

hídrico nas raízes faz com que este órgão seja o primeiro sinalizador do estresse, que envia o alerta de dano para toda a planta, através de hormônios sinalizadores como ABA e etileno, que induzem o fechamento estomatal (Dubos e Plomion, 2003). A plasticidade das raízes em condições de seca é muito importante, visto que estas têm papel fundamental na recuperação da planta quando o estresse cessa (Porcel e Ruiz-Lozano, 2004).

Após o início do estresse hídrico, uma série de mecanismos de defesa é desencadeada, levando à ativação de cascatas de transdução de sinal que ativam vários fatores de transcrição e *cis*-elementos presentes em promotores de genes responsivos a estresse (Le et al., 2011a). Em situações de seca, frio ou salinidade, uma via de transdução de sinal genérica é desencadeada e pode ocorrer da seguinte maneira: o primeiro sinal ao dano é ainda extracelular e é percebido por receptores de membrana, que ativam cascatas de sinalização por meio de mensageiros secundários, como íons Ca^{2+}, espécies reativas de oxigênio (*reactive oxygen species* – ROS) e inositol fosfato; os mensageiros secundários modulam o nível intracelular de Ca^{2+}, o qual é percebido por proteínas Ca^{2+}-sensores, as quais mudam sua conformação e interagem com seus respectivos parceiros, iniciando uma cascata de fosforilação, que ativa fatores de transcrição e genes responsivos ao estresse; o produto desses genes leva, em ultima análise, à adaptação da planta ao estresse por meio de sinais sinérgicos à planta como um todo (Figura 2) (Huang et al., 2012).

As pesquisas atuais buscam a compreensão das diversas vias de sinalização que atuam nas células afetadas pelo déficit hídrico bem como o entendimento das interações entre as diferentes vias. Os resultados dessas complexas vias de transdução de sinal relacionadas ao estresse e o impacto potencial que exercem sobre a agricultura moderna têm feito com que essa abordagem ganhe atenção especial (Le et al., 2011b).

1.3 EXPLORANDO A FUNÇÃO GÊNICA POR MEIO DA GENÔMICA

A genômica é um campo da ciência que pode ser dividida em três grandes áreas: a) estrutural, a qual se concentra na estrutura de genomas; b) comparativa, a qual consiste na análise e comparação de genomas de diferentes espécies, a fim de entender a evolução de regiões gênicas e intergênicas e c) funcional, a qual inclui o

perfil do transcritoma global, genética reversa e clonagem baseada em mapeamento (Gutterson e Zhang, 2004). Essas abordagens permitem uma melhor compreensão das respostas das plantas às condições ambientais, bem como o entendimento da intrincada rede de relações entre genes envolvidos na resposta a estresses abióticos e bióticos.

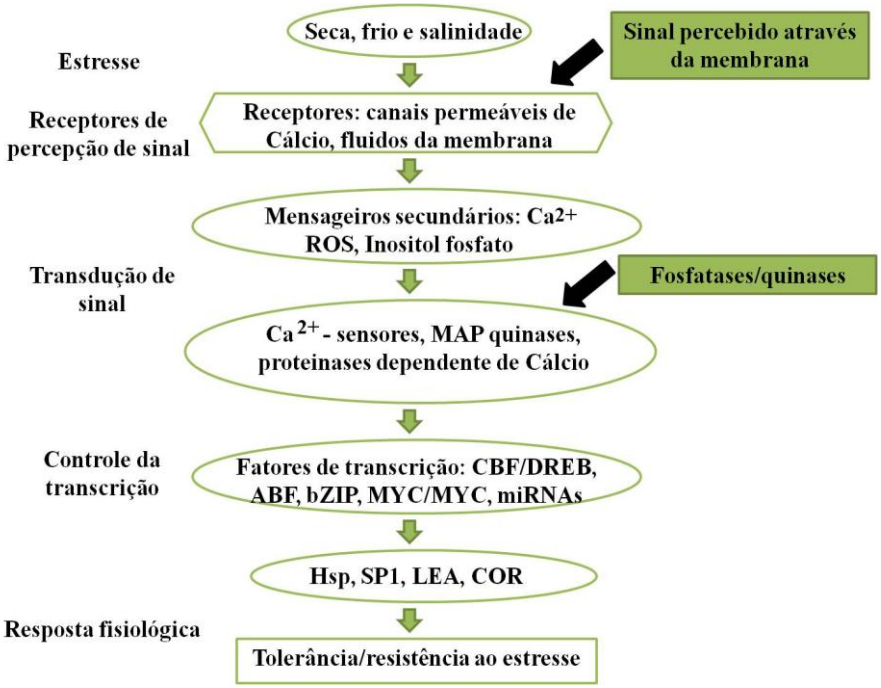


Figura 2: Via genérica de resposta das plantas a estresses abióticos. O sinal de estresse extracelular é percebido por receptores de membrana, os quais ativam inúmeras cascatas de sinalização, como a geração de mensageiros secundários. A cascata de sinal leva à ativação de fatores de transcrição, os quais ativam genes alvos, de forma direta ou indireta, que atuam na resposta de adaptação da planta. Modificado de Huang et al. (2012).

Sabe-se que as mudanças bioquímicas e fisiológicas que ocorrem em plantas sob condições de estresse estão relacionadas com alterações na expressão gênica (Saibo et al., 2009). Entretanto, a questão chave em estudos da expressão gênica é o alto grau de relação e sobreposição de diferentes rotas metabólicas (Gutterson e Zhang, 2004). Assim, a função de um único gene pode interferir em inúmeras respostas da planta aos sinais ambientais. Neste contexto, muitos fatores de transcrição estão sabidamente envolvidos na adaptação da planta às condições de dano e na

modulação da expressão de genes responsivos a diferentes estresses (Saibo et al., 2009).

A família de fatores transcricionais DREB/CBFs, por exemplo, regula a expressão gênica em resposta à seca, salinidade e baixas temperaturas (Li et al., 2005). As famílias MYB e MYC são relacionadas à resposta a estresses bióticos e suas lesões, sendo responsivos ao ácido jasmônico (Saibo et al., 2009). Outra família gênica bastante estudada, a família que codifica os fatores de transcrição WRKY, está envolvida na regulação negativa e positiva da defesa a estresses, crescimento e desenvolvimento em plantas (Agarwal et al., 2010).

A família conhecida como "*lesion simulating disease*" ou fenótipo LSD vem sendo bastante estudada em *A. thaliana*. A família recebe esse nome devido ao fato de incluir uma das seis classes de mutantes que mimetizam lesões (*lesions mimic mutants*) já isolados (Lorrain et al., 2003; Moeder e Yoshioka, 2008). Alguns mutantes formam lesões que se assemelham àquelas resultantes em doenças, embora o fenótipo ocorra mesmo na ausência de patógeno (Dietrich et al., 1994).

Dietrich et al. (1994) foram os primeiros a descrever os genes *LSD* por meio da análise de mutantes em *A. thaliana*. Os mutantes analisados apresentavam marcadores bioquímicos e moleculares de resposta de resistência à doença e os autores relacionaram o fenótipo da lesão à resposta de resistência sistêmica adquirida (SAR – do inglês *systemic acquired resistance*) e à PCD. Uma descrição mais detalhada do gene *LSD1* foi feita posteriormente por Dietrich et al. (1997). Os autores identificaram o domínio característico da família CxxCxRxxLMYxxGASxVxC, o qual foi designado domínio dedo de zinco LSD. Além disso, nesse trabalho foi demonstrado o envolvimento do gene na regulação negativa da HR, via um sinal dependente de superóxido. Desde então, muitos trabalhos vêm demonstrando o papel importante dos genes *LSD* no controle da HR ocasionada por estresses bióticos e abióticos.

A HR é um dos mecanismos envolvidos na percepção da planta ao ataque do patógeno, desencadeando o fenômeno conhecido como PCD no local da infecção e, consequentemente, ajudando a combater o crescimento de agentes patogênicos (Rusterucci et al., 2001; Mur et al., 2008). Segundo Coll et al. (2011), a HR se dá pela

seguinte maneira: quando em contato com um patógeno, a primeira linha de defesa em plantas ocorre por meio de receptores PRR (do inglês *pattern recognition receptors*), como receptores RPK (do inglês *receptor-like kinases*), os quais reconhecem MAMPs (do inglês *microbe-associated molecular pattern*), que sinalizam para a resposta imune. Patógenos capazes de suprimir PRR liberam inúmeros efetores, que podem ser reconhecidos, ativando ETI (do inglês *effector-triggered immunity*). ETI é mediada por moléculas como NB-LRR (do inglês *nucleotide-binding domain-leucine-rich repeat*), proteínas de resistência à doença, as quais impedem o desenvolvimento do patógeno por meio da HR e pela ativação de genes relacionados a esta e à defesa (Figura 3). A HR se dá a partir da geração de ROS no apoplasto, cloroplastos e mitocôndrias. Além disso, muitos sinais de dano são disparados, como o aumento do nível de cálcio celular, encolhimento citoplasmático, condensação da cromatina, vacuolização e rompimento do cloroplasto (Figura 4).

Quando ocorre o desenvolvimento de lesões, as necroses num local específico, delimitadas por tecido saudável, dão inicio à SAR, que faz com que a resistência ao ataque do patógeno seja aumentada nos tecidos próximos e nos mais distantes do local da infecção (Rusterucci et al. 2001). Porém, quando não há regulação da extensão da morte celular, essa pode progredir além do limite necessário da HR e então o órgão da planta pode ser totalmente destruído (Epple et al. 2003; Moeder e Yoshioka 2008). Esse fenômeno é observado quando genes LSD são mutados, levando a uma acelerada morte celular (Dietrich et al. 1994; Dietrich et al. 1997; Rusterucci et al. 2001; Kaminaka et al. 2006; He et al. 2011). Além disso, alguns mutantes *lsd* formam lesões espontâneas mesmo na ausência de patógeno (Dietrich et al., 1994; Weymann et al., 1995; Hunt et al., 1997; Morel e Dangl, 1999; Chaerle et al., 2001).

Alguns trabalhos vêm demonstrando a interação de *At*LSD1 com outras proteínas relacionadas a PCD. Coll et al. (2010) demonstraram a interação de *At*LSD1 com a metacaspase do tipo I (*At*MC1), que atua regulando positivamente PCD. *At*LSD1 retém *At*MC1 inativa no citoplasma, impedindo a propagação de PCD nas células vizinhas à infecção (Figura 5). A interação de *At*LSD1 com *At*bZIP10 (do inglês *basic region leucine zipper 10*), um regulador positivo de PCD, ocorre também via retenção deste no citoplasma, impedindo que o mesmo se desloque para o núcleo da célula,

ativando PCD (Kaminaka et al. 2006). Por outro lado, a interação de *At*LSD1 com outro regulador negativo, *At*GILP também já foi descrita (He et al. 2011b). Coll et al. (2011), via *screening* de duplo híbrido em levedura, propuseram que *At*LSD1 atua como um "*deathosome*", interagindo com outros reguladores de PCD (Figura 6). Assim, as evidências vêm apontando *At*LSD1 como um gene chave no controle da PCD induzida por HR.

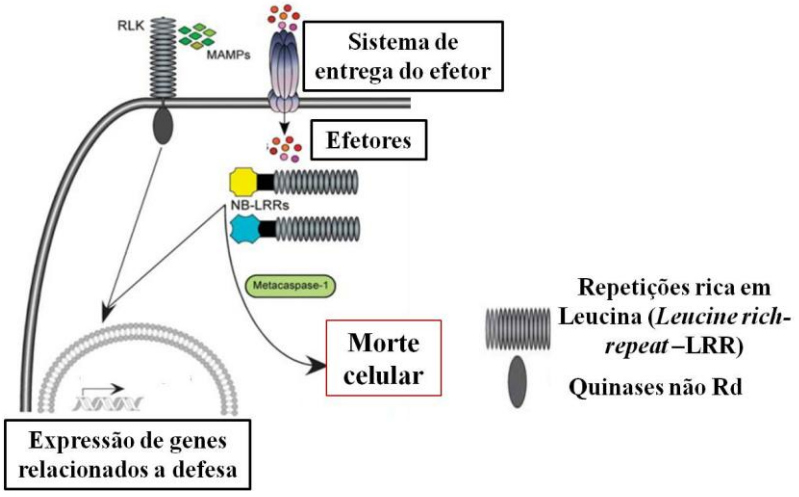


Figura 3: Via de resposta imune em células vegetais: do reconhecimento do patógeno a sinalização. As proteínas RLKs são receptores transmembrana com dominios LRR, reconhecem MAMPs, que sinalizam para a resposta imune. RLK associam-se com quinases não-RD, as quais tipicamente carregam uma cisteina ou glicina no lugar da arginina, antes do residuo aspartato catalítico. Liberação de efetores pelos patógenos, os quais são reconhecidos pelas NB-LRR. Essa cascata de sinalização impede o desenvolvimento do patógeno por meio da HR e pela ativação de genes relacionados com a defesa. Modificado de Coll et. al. (2011).

Outro sinal relacionado à PCD é o aumento de ROS, que são moléculas chaves na transdução de sinal em resposta à infecção, fazendo com que sejam induzidos diversos mecanismos de defesa (Mittler et al., 2004). No entanto, o excesso de produção e a falta de manutenção do balanço de ROS existente na célula fazem com que esta sofra dano oxidativo, o que leva a defeitos no crescimento e morte celular (Torres et al., 2005). Alguns pesquisadores têm demonstrado que a família LSD tem relação com o controle do estresse oxidativo, por meio de indução de enzimas responsáveis pela detoxificação de ROS (como superóxido dismutase e

catalase, por exemplo), fazendo assim com que o dano oxidativo não ocorra (Jabs et al., 1996; Kliebenstein et al., 1999; Epple et al., 2003; Mateo et al., 2004).

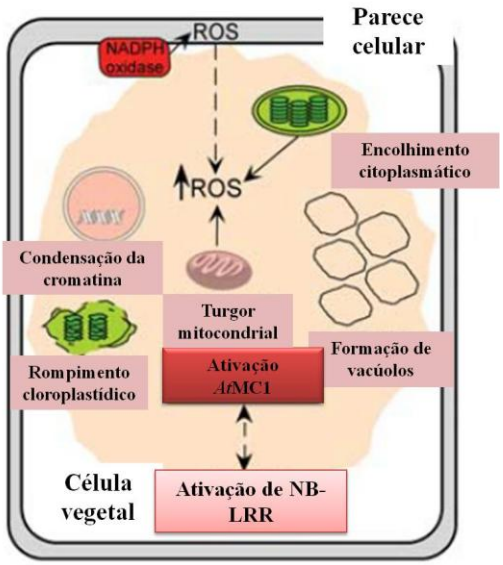


Figura 4: HR em plantas. Geração de ROS no apoplasto por meio de NADPH oxidase. Este é essencial para o desenvolvimento da HR e ativação da resposta imune. ROS também é gerado nos cloroplastos e nas mitocôndrias. O encolhimento citoplásmico, condensação da cromatina, turgor mitocondrial, formação de vacúolos e rompimento cloroplastídico são outros sinais que ocorrem na HR. Adaptado de Coll et. al. (2011).

Além do papel em PCD, alguns estudos vêm demonstrando a importância da família LSD em diversos processos relacionados a estresses abióticos, como aclimatação às condições que promovem excesso de excitação energética (Mateo et al. 2004; Muhlenbock et al. 2008), formação de aerênquima sob condições de hipóxia (Muhlenbock et al. 2007) e regulação da morte celular condicionada por frio (Huang et al. 2010).

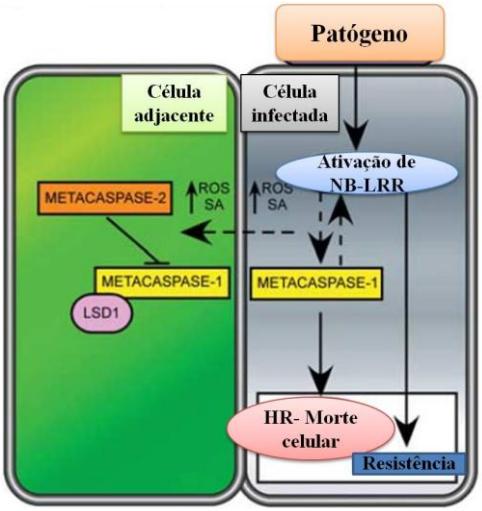


Figura 5: Modelo de atuação de *At*LSD1 e *At*MC1. Após a percepção do ataque de patógenos via NB-LRR na célula inicial da infecção, moléculas sinalizadoras de dano são liberadas como ROS e SA. *At*MC1 atua ativando HR e morte celular, que em ultima análise, leva à resistência. Esses sinais são propagados às células vizinhas à célula inicial da infecção. A fim de evitar a extensão da HR e da PCD (que pode levar a destruição do órgão da planta), *At*LSD1 interage via domínio dedo de zinco LSD (o qual também é presente em *At*MC1) com *At*MC1, retendo esta no citoplasma, impedindo a propagação da PCD. A mesma interação não ocorre com *At*MC2, a qual regula negativamente *At*MC1, por sua vez regulando negativamente PCD, tal qual *At*LSD1. Modificado de Coll et al. (2011).

Até o momento, a maioria das pesquisas com a família LSD foi realizada em *A. thaliana*. Existem alguns poucos trabalhos em outros organismos, como *Brassica oleracea* (Coupe et al., 2004), *Oryza sativa* (Wang et al., 2005), *Hordeum vulgare* (Keiša et al., 2008), *Pisum sativa* (He et al., 2011a) e *Bambusa oldhamii* (Yeh et al., 2011). Porém, em soja ainda não foram publicados estudos referentes a essa família gênica. Por outro lado, a identificação da família LSD ainda não foi realizada, o que dificulta a reconstrução filogenética dessa família. Além disso, não existem trabalhos relatando o envolvimento de genes *LSD* na resposta da soja a estresses desencadeados por ASR e déficit hídrico.

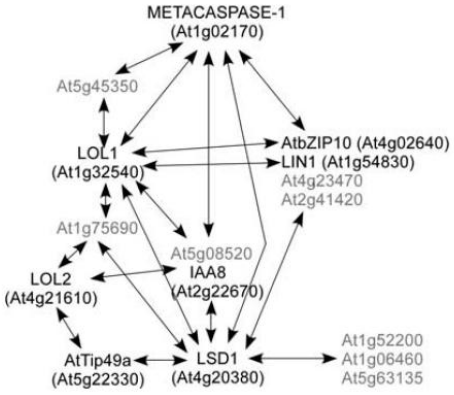


Figura 6: "*deathosome*" de *At*LSD1. Diagrama representado a interação de reguladores conhecidos de PCD com *At*LSD1 via duplo hibrido de levedura. Os genes escritos em cinza ainda não tem função conhecida (Coll *et al.*, 2011).

Portanto, é de especial interesse a anotação gênica da família LSD, o que poderá contribuir para o entendimento da evolução dessa família e revelar potenciais genes-alvos para abordagens biotecnológicas. A caracterização do perfil de expressão dos genes *LSD* de soja, (*Glycine max* LSD - *GmLSD*), possibilitará uma melhor compreensão acerca da família LSD, bem como a elucidação do envolvimento destes na resposta da planta a estresses bióticos e abióticos.

OBJETIVOS

1.4) OBJETIVO GERAL

O objetivo geral deste estudo foi identificar genes pertencentes à família LSD, com ênfase na caracterização dos genes identificados em soja [*Glycine max* (L.) Merrill].

1.5) OBJETIVOS ESPECÍFICOS

Os objetivos específicos deste trabalho foram:

a) Identificar os genes da família LSD nos genomas disponíveis em bancos de dados;

b) Analisar o nível de conservação dos domínios dedo de zinco LSD dos genes identificados;

c) Realizar análise filogenética dos genes identificados, a fim de reconstruir a história evolutiva da família;

d) Analisar a estrutura gênica dos genes *LSD* de soja (*GmLSD*);

e) Realizar análise *in silico* da região promotora dos genes *GmLSD*;

f) Determinar o padrão de expressão dos genes *GmLSD* em diferentes órgãos da planta de soja;

g) Determinar o padrão de expressão dos genes *GmLSD* em plantas de soja submetidas à infecção por ASR e ao déficit hídrico.

CAPÍTULO II

Evolution of the Lesion Simulating Disease (LSD) gene family in Viridiplantae and evidence of the involvement of *LSD* genes in the responses to biotic and abiotic stresses in soybean

---

Artigo submetido ao periódico *Functional and Integrative Genomics*

Evolution of the Lesion Simulating Disease (LSD) gene family in Viridiplantae and evidence of the involvement of *LSD* genes in the responses to biotic and abiotic stresses in soybean

Caroline Cabreira[1], Alexandro Cagliari[1], Lauro Bücker-Neto[1], Beatriz Wiebke-Strohm[1], Loreta B. de Freitas[1], Francismar C. Marcelino-Guimarães[2], Mayra C. C. G. de Carvalho[2], Alexandre L. Nepomuceno[2], Ricardo V. Abdelnoor[2], Márcia M. A. N. Margis-Pinheiro[1] and Maria H. Bodanese-Zanettini[1]

[1] Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil [2] Empresa Brasileira de Pesquisa Agropecuária - Embrapa Soja, Londrina, PR, Brazil.

Corresponding author: Maria Helena Bodanese-Zanettini, e-mail: mhbzanettini@yahoo.com.br, phone: 0055-51-33086725, fax: 0055-51-33089823, address: Departamento of Genetics, Federal University of Rio Grande do Sul - UFRGS, P. O. Box 15053, CEP 91501-970, Porto Alegre, RS, Brazil.

Abstract

The *Lesion Simulating Disease* (*LSD*) genes encode a family of zinc finger proteins that are reported to play an important role in the hypersensitive response and programmed cell death (PCD) that are caused by biotic and abiotic stresses. In the present study, 117 putative LSD family members were identified in Viridiplantae. We identified genes with one, two or three conserved LSD domains. Phylogenetic analysis showed that genes with three LSD domains, which are found in basal organisms, represent the ancestral condition. Genes with two LSD domains were identified only in the Embryophyte clade, and genes possessing one LSD domain were highly represented in grass species. Monocot and eudicot genes were not found to be separate groups, indicating that the expansion of the LSD family occurred prior to the diversification of these clades. Expression analyses by real-time quantitative

polymerase chain reaction (RT-qPCR) of *Glycine max LSD* (*GmLSD*) genes indicated that they are not ubiquitously expressed in soybean and that their expression patterns are instead organ-dependent. RNA sequencing method (RNA-seq) and RT-qPCR analyses indicated that the expression of the majority of *GmLSD* genes is modulated in soybean during *Phakopsora pachyrhizi* infection. In addition, RT-qPCR data showed that the expression of some *GmLSD* genes is modulated in plants under drought stress. These results suggest the involvement of *GmLSD* genes in the response of soybean to both biotic and abiotic stresses.

Key words: *Lesion Simulating Disease* (*LSD*), zinc finger LSD domain, LSD family phylogeny, soybean expression analysis, *Phakopsora pachyrhizi*, drought

Introduction

Zinc finger proteins play key roles in a variety of cellular functions, such as transcriptional regulation, apoptosis control, RNA binding and protein-protein interactions. The *Lesion Simulating Disease* (*LSD*) protein family comprises well-characterized zinc finger proteins that contain the conserved zinc finger LSD domain CxxCRxxLMYxxGASxVxCxxC (Dietrich et al. 1997). The *LSD* genes were first characterized in *Arabidopsis thaliana* mutants that spontaneously form necrotic lesions in the absence of pathogenic infection. These genes were shown to negatively regulate programmed cell death (PCD) (Dietrich et al. 1994; Dietrich et al. 1997). Lesion formation in *lsd* mutants is associated with the expression of histochemical and molecular markers of the plant disease response, and in some cases, the mutants showed a significant increase in their resistance to pathogen attack (Dietrich et al. 1994). The well-characterized *A. thaliana LSD1* gene (*AtLSD1*) has been described as a key regulator of abiotic and biotic stress responses in plants. *AtLSD1* participates in the signaling pathway that induces the expression of copper zinc superoxide dismutase (CuZnSOD) in response to salicylic acid (SA), and it also negatively regulates reactive oxygen species (ROS) generated under stress conditions (Kliebenstein et al. 1999), thereby limiting the accumulation of superoxide in the cell (Jabs et al. 1996). The Arabidopsis *lsd1* mutant exhibited impaired spreading of the hypersensitive response (HR) throughout the course of a pathogen attack, which consequently expanded

beyond the infection site engulfing the entire leaf (Dietrich et al. 1997). During photooxidative stress, Arabidopsis *lsd1* mutants showed high levels of ROS, reduced stomatal conductance and low peroxisomal catalase activity, which suggests that *LSD1* is required for acclimation to conditions that promote excess excitation energy (Mateo et al. 2004). Together with the *Enhanced Disease Susceptibility 1* (*EDS1*) and *Phytoalexin Deficient 4* (*PAD4*) genes, *AtLSD1* acts as an ROS/ethylene homeostatic switch during light acclimation and pathogen defense (Muhlenbock et al. 2008). Under hypoxia conditions, *AtLSD1* regulates lysigenous aerenchyma formation (Muhlenbock et al. 2007), while under low temperature conditions, it regulates the response to cell death (Huang et al. 2010). Thus, *AtLSD1* has been proposed to act as a cellular hub, making a central contribution to the oxidative stress response in plants (Kaminaka et al. 2006).

Soybean (*Glycine max*) is one of the most economically important crops in the world. Biotic and abiotic stresses severely restrict soybean yield. With regard to biotic stress, Asian Soybean Rust (ASR), caused by the *Phakopsora pachyrhizi* fungus, is one of the main diseases affecting soybean production. The pathogen develops in leaves, stems and pods and is able to defoliate soybean plants in a few days, leading to drastic crop losses (Goellner et al. 2010). Soybean genes determining resistance to *P. pachyrhizi* (*Rpp* genes) have been identified (Cheng and Chan 1968; Hidayat and Somaatmadia 1977; Singh and Thapliyal 1977; Bromfield and Hartwig 1980; Hartwig 1986; Garcia et al. 2008; Kim et al. 2012; Li et al. 2012). Because no available commercial soybean cultivar is resistant to all pathotypes of *P. pachyrhizi*, the application of fungicides is the only method of controlling the spread of infection. Regarding abiotic stress, drought is the most important condition that affects soybean production, especially during flower establishment (Meckel et al. 1984). Moreover, drought stress affects biological nitrogen (N) fixation, which leads to a reduced supply of nitrogen for protein production (Ledrera et al. 2007; Manavalan et al. 2009).

Although some *LSD* genes have been reported to play important roles in the defense against fungal isolates (Dietrich et al. 1994; Weymann et al. 1995; Hunt et al. 1997; Aviv et al. 2002; Wang et al. 2005; Yeh et al. 2011), bacterial isolates (Epple et al. 2003; Bhatti et al. 2008) and abiotic stress (Mateo et al. 2004; Muhlenbock et al. 2008; Huang et al. 2010), the potential of *Glycine max LSD* genes (*GmLSD*) to protect plants against environmental stresses remains to be determined. Furthermore, there is

no complete identification of the LSD gene family, which hampers the reconstitution of its evolutionary history. Thus, the present study reports the identification and phylogenetic reconstruction of the LSD gene family in Viridiplantae and presents evidence of the involvement of *GmLSD* genes in the responses to *P. pachyrhizi* infection and drought conditions.

Material and Methods

Viridiplantae *LSD* gene annotation

To identify the *LSD* genes present in Viridiplantae, including representatives of the monocot, eudicot, moss, lycophyte and algae species, the complete nucleotide and protein sequences of the well-characterized *AtLSD1* gene (Dietrich et al. 1994) were used as the query in tBLASTx and BLASTn searches conducted against the Phytozome (http://www.phytozome.org/) and PLAZA (http://bioinformatics.psb.ugent.be/plaza/) databases. Homologous sequences exhibiting greater than 50% homology with the query sequence in this first-round BLAST search were then used as query sequences in a second-round BLAST search against the National Center of Biotechnology Information database (NCBI) (http://www.ncbi.nlm.nih.gov/). The putative LSD protein sequences identified were examined for the presence of the previously reported zinc finger LSD domain (Dietrich et al. 1997). These analyses were performed using the SMART (http://smart.embl-heidelberg.de/) and InterProScan Signature (http://smart.embl-heidelberg.de/) databases. Sequences containing one, two or three LSD domains were annotated as LSD proteins. The degree of conservation of the zinc finger LSD domain was analyzed using the MEME suite (http://meme.sdsc.edu/meme/).

To identify putative pseudogenes, the coding sequences (CDS) of the *LSD* genes identified were used as query sequences in a BLAST search against the NCBI expressed sequence tag (EST) database (http://www.ncbi.nlm.nih.gov/nucest/). The identities of the resulting ESTs were confirmed using the Phytozome (http://www.phytozome.org/) and PLAZA (http://bioinformatics.psb.ugent.be/plaza/) databases.

Phylogenetic analysis

The complete consensus sequences of the LSD proteins identified were aligned manually using the Molecular Evolutionary Genetics Analysis (MEGA) 5.05 program (Tamura et al. 2011) and were back-translated to nucleotide sequences for analysis at the DNA level. Contiguous insertion-deletion events (indels) were treated as single mutations (Simmons and Ochoterena 2000). To construct the phylogenetic tree, two different methods were used. The method based on distance Neighbor-Joining (NJ) was performed in MEGA 5.05 using p-distance methods, pairwise deletion, as the basis for testing each node with 1000 bootstrap replicates. A Bayesian approach was also taken using the software program BEAST v1.4.7 (Drummond and Rambaut 2007), in which a run of 30^7 chains was performed, along with sampling every 1000 generations. The Yule tree prior, the HKY substitution model and the uncorrelated *log*-normal relaxed clock were used in the BEAST analysis. The software program TRACER v1.4 (http://beast.bio.ed.ac.uk/Tracer) was used to check the convergence of Monte Carlo Markov Chains (MCMC) and adequate effective sample sizes (>200) after the first 10% of generations had been deleted as burn-in. The final joint sample was used to estimate the maximum clade credibility tree using the TreeAnnotator program, which is part of the BEAST package. The statistical support for the clades was determined by assessing the Bayesian posterior probability. The trees were visualized using FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Soybean exon-intron structure and promoter sequence analysis

To assess the structural conservation of *GmLSD* genes, their exon/intron structure was analyzed. The putative promoter region from the 2000 base pairs (bp) upstream of the transcription start site (TSS) of each *GmLSD* gene was used to search for putative *cis*-elements. The analysis was performed using the Plant Care database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/), and the *cis*-elements identified were classified based on their putative biological functions.

Expression analysis of *GmLSD* genes

To confirm *GmLSD* gene expression, different plant tissues were analyzed using real time quantitative polymerase chain reaction (RT-qPCR). Leaves, roots and stems of plants in the vegetative (V) phase and seeds, pods and flowers before and after fertilization were harvested from the MGBR-46 Conquista Brazilian soybean cultivar grown under greenhouse conditions. The tissues were harvested and immediately frozen in liquid nitrogen before being stored at -80º C until RNA extraction. Three biological replicates, with three plants per replicate, were sampled for each plant organ. The relative expression levels of the transcripts were statistically analyzed using variance analysis, with data transformation when necessary. Means were compared using the Tukey multiple comparison test. Statistical Analysis System (SAS) 9.2 and the Statistical Package for the Social Sciences (SPSS/PASWSTAT) 18 were also used.

Two experiments were conducted to analyze *GmLSD* gene expression profiles in response to *P. pachyrhizi* infection. These experiments were carried out in Embrapa Soja, Londrina, Parana, Brazil. In the experiment I, *GmLSD* gene expression profiles were analyzed by RT-qPCR. In this experiment, the susceptible cultivar EMBRAPA48 (van de Mortel et al. 2007) and the resistant genotype PI561356 (Almeida et al. 2012) were used. The PI561356 genotype presents the *Rpp1* gene (responsible for immune response against some isolates), mapped at the chromosome 18 (Kim et al. 2012). Plants in the V2 stage, which consist of a fully developed trifoliolate leaf at the node above the unifoliolate nodes (Fehr and Caviness 1977), were grown in a greenhouse and sprayed with a fungal spore suspension according to the method of Wiebke-Strohm et al. (2012). Leaves sprayed with a solution lacking fungal spores were used as mock-treated control. One trifoliolate leaf from each plant was collected at 12, 24, 48, 96 and 192 hours post-inoculation (hpi) and was frozen in liquid nitrogen before being stored at -80ºC. Three biological replicates were analyzed from each genotype per treatment. The relative expression levels of *Gm**LSD* genes were statistically compared by variance analysis with factorial treatments based on three factors: genotype, time and pathogen presence. When necessary, data were transformed using the weighted least squares method. Means were compared using the Bonferroni multiple comparison test. The SAS 9.2 and the SPSS/PASWSTAT 18 software packages were used to perform this analysis.

In the experiment II, RNA sequencing method (RNA-seq) of microdissected lesions was performed. The *P. pachyrhizi*-susceptible variety BRS231 (Yamanaka et al. 2011) and the resistance PI561356 genotype at V2 growth stage were used. Ten days after infection with *P. pachyrhizi*, foliar segments (1 cm^2) containing lesions were collected and fixed on ice in Farmer's solution (Kerk et al. 2003) before being dehydrated and embedded in paraffin (Cai and Lashbrook 2006). Serial sections of 12 µm were made using a rotary microtome and transferred to microscope membrane slides. Twenty sections containing a variable number of rust lesions were prepared for each biological replicate under each treatment condition. The PixCell II LCM system (Arcturus, Sunnyvale, CA, USA) and the CapSure Macro LCM system (Arcturus) were used to collect foliar cells within the lesion. RNA extraction using the PicoPure RNA isolation kit (Arcturus) was performed on cells collected from a variable number of infection sites for each biological replicate. Synthesis of cDNA was carried out, and high performance paired-end (108 bp) sequencing was performed on the Illumina Genome Analyzer GAIIx (San Diego, CA, USA). Low quality RNA-Seq reads were discarded. A total of 86,301,242 reads were aligned against the soybean genome, and the corresponding genes were predicted using the TopHat (Trapnell et al. 2009) and SOAP2 (Li et al. 2009) aligners. Gene expression was calculated using the fragments per kilobase of exon per million fragments mapped (FPKM) values (Mortazavi et al. 2008). To identify differentially expressed genes, pairwise comparisons between the FPKM values from each genotype were performed using a *t*-test. A 99% confidence interval was used as the threshold value to identify differentially expressed genes. This experiment was carried out as part of the work of the Biotecsur Consortium (Apoyo al desarrolo de las biotecnologias en el Mercosur), and the results are available to members of the consortium (http://bioinfo.cnpso.embrapa.br/genosoja/).

An additional experiment was carried out to analyze changes in the expression profile of *GmLSD* genes in response to water deficit. A highly sensitive BR16 cultivar and a slightly sensitive EMBRAPA48 cultivar (Oya et al. 2004) were grown in a greenhouse according to the method described by Kulcheski et al. (2011) and were submitted to drought stress as described by Martins et al. (2008). Briefly, seedlings in the V2 stage were removed from a hydroponic solution and kept in a tray in the dark without nutrients. Leaves and roots were collected at 0 (control), 25, 50, 75, 100, 125 and 150 minutes after the initiation of drought stress and were kept in liquid nitrogen

at -80°C until RNA extraction. Three biological replicates (three plants/replicate) were sampled for each organ/genotype/treatment point. The relative expression level of *GmLSD* genes in each organ was statistically compared by variance analysis considering drought stress time and cultivar. Data were transformed using the weighted least squares method. Means were compared using the Bonferroni multiple comparison test. The SAS 9.2 and the SPSS/PASWSTAT 18 programs were used.

RNA extraction, reverse transcription and quantitative PCR

Total RNA was extracted from each sample using TRIzol reagent (Invitrogen, Carlsbad, USA), and 2 µg of DNA-free RNA of each sample was subsequently treated with DNAse I (Promega, Madison, USA). The first-strand cDNA synthesis reaction was performed using the M-MLV Reverse Transcriptase System (Invitrogen) and a 24-mer oligo dT anchored primer.

RT-qPCR was performed in a StepOne Applied Biosystem Real-Time Cycler. For all experiments, the RT-qPCR thermocycling began with a 5 min initial denaturation step at 94°C, after which 40 cycles consisting of a 10 s denaturation step at 94°C, a 15 s annealing step at 60°C and a 15 s extension step at 72°C were performed, and this was followed by a final extension step of 2 min at 40°C. To identify different products, a melting curve analysis over a range of 55-99°C and with a stepwise temperature increase of 0.1°C per second was performed at the end of each PCR run. The final volume of each reaction was 25 µl and comprised 12.5 µl of diluted DNA template, 1 X PCR buffer (Invitrogen), 50 mM $MgCl_2$, 10 mM of each dNTP, 10 µM of each primer, 2.5 µl SYBR-Green solution (1:100,000, Molecular Probes Inc., Eugene, USA) and 0.06 U Platinum Taq DNA Polymerase (Invitrogen). The first-strand cDNA reaction product was diluted 1:100 for evaluation the relative expression analyses. Technical quadruplicates were performed for each sample and negative control was used in all reactions.

Primer 3 software (http://frodo.wi.mit.edu/) was used to design specific oligonucleotides corresponding to each *GmLSD* gene (Electronic Supplementary Material (ESM) Table S1). The annealing temperature used for all primers was 60°C. The specificity of the amplicons obtained in the RT-qPCR was confirmed by sequencing on an ABI PRISM 3100 Genetic Analyzer automatic sequencer (Applied

Biosystems) in the ACTGene Laboratory (Centro de Biotecnologia-UFRGS, Porto Alegre, Rio Grande do Sul, Brazil). Several housekeeping genes were selected for the normalization of mRNA levels in RT-qPCR. The *ACTII*, *CYP2* (Jian et al. 2008) and *metalloprotease* (Libault et al. 2008) genes were used for normalization in the plant organ experiment, while the *metalloprotease* and *f-box* genes (Libault et al. 2008) were used in the *P. pachyrhizi* assay, and the *f-box*, *ACTII* and *ELF* genes (Jian et al. 2008) were used in the drought assay. The stability of the housekeeping genes was examined using Genorm software (http://medgen.ugent.be/~jvdesomp/genorm/). The relative gene expression was analyzed using the 2-ΔΔCt method (Livak and Schmittgen 2001).

Results

LSD genes are widely distributed in Viridiplantae

We identified a total of 117 putative *LSD* genes that are widely distributed in Viridiplantae species (Fig. 1). The genes were assigned acronyms according to the BLAST output order, and these are shown in ESM Table S2. The number of genes in each species ranged from one to eight, and the Embryophyte node was the earliest point in the phylogeny at which increases in the copy numbers of the genes were detected. In the most ancestral genomes, *Volvox carteri* and *Chlamydomonas reinhardtii*, only one *LSD* gene was identified. In the moss *Physcomitrella patens* and in the lycophyte *Selaginella moellinorffii*, two and one *LSD* genes were found, respectively. Among the monocot clade, *Setaria italica* was the species with the greatest number of LSD family genes, with six members identified, while soybean and *Linum usitatissimum* had the most *LSD* genes among the eudicot species, with eight genes identified in each. In contrast, genes belonging to the LSD family were not identified in fungi or animals, including humans.

The identified *LSD* sequences genes were further analyzed to determine the number of LSD domains present in their encoded protein products. Our analysis showed that the LSD proteins possess one, two or three zinc finger LSD domains, although the proteins with three such domains constitute the majority of the LSD sequences identified (Fig. 1). A decrease in the number of encoded LSD domains per

protein was identified only in the Embryophyte clade. In addition, except for the *Mdo2* (*Malus domestica 2*) gene, the proteins with only one LSD domain were restricted to the grass clade. To investigate the conservation between LSD domains, regions corresponding to the consensus sequence CxxCRxxLMYxxGASxVxCxxC (Dietrich et al. 1997) were compared. The results of this analysis showed that the LSD domains exhibit broad conservation across all *LSD* genes and that the cysteine residues at positions 1, 4, 19 and 22 are particularly well conserved (Fig. 2). The neutral amino acids cysteine, leucine, tyrosine, glycine, alanine and valine were more often conserved than were the other amino acids within the LSD domains.

The presence of at least one of the following features indicates that a gene is potentially a pseudogene: a lack of transcript support, a short coding region, a long untranslated region, and either a lack of a paralog or the presence of a significantly shorter paralog for genes residing within a segmental duplication (Thibaud-Nissen et al. 2009). With this in mind, we investigated the expression of the putative *LSD* genes by searching for corresponding ESTs in the database held by the NCBI. We identified ESTs for 87 of the 117 genes analyzed, and these included ESTs corresponding to genes with only one or two LSD domains (ESM Table S2).

Genes encoding three LSD domains represent the ancestral condition of the LSD gene family

In our analysis of the evolution of the LSD gene family, unrooted phylogenetic trees generated by either the NJ method or the Bayesian method possessed the same topology, but the tree generated by the Bayesian method was better supported than the NJ tree, as expected (data not shown). We observed the formation of three clusters, one of which comprised the more basal organisms (algae, pteridophytes and bryophytes) and some members of monocot species, while the other two clusters included members of both the monocot and the eudicot clades (Fig. 3). The topology of the trees and the presence of three LSD domains in genes found in the most basal organisms suggest that to have three such domains is the basal condition of the LSD gene family.

The distribution of the *LSD* genes within the phylogeny is structured in a manner that corresponds to the number of LSD domains encoded by the genes. While genes

with three domains formed two separate clusters, the genes with one or two LSD domains were grouped in two different subclusters. The *Rco2* (*Ricinus communis 2*) gene was a unique exception to this pattern because it encodes two LSD domains but was located in a group of genes that each encode three LSD domains.

The *GmLSD* genes encoding three LSD domains were separated into two subclusters. *GmLSD1*, *GmLSD2*, *GmLSD6* and *GmLSD7* were located in one cluster, while *GmLSD3*, *GmLSD4* and *GmLSD5* were present in the other. *GmLSD8*, which encodes a protein with two LSD domains, was included in the cluster that contains genes that encode two LSD domains (Fig. 3).

GmLSD genes have different exon-intron organization and their promoter regions are enriched for _cis_-elements related to stress responses

Investigation of the properties of exons and introns can provide a better understanding of the mechanisms that determine exon-intron architecture (Zhu *et al.*, 2009). Structural analysis of the *GmLSD* genes revealed three distinct groups: the *GmLSD1*, *GmLSD2* and *GmLSD4* genes all contain six exons, while the *GmLSD3*, *GmLSD5* and *GmLSD7* genes have five exons, and the *Gm*LSD6 and *GmLSD8* genes have four exons (Fig. 4). The sizes of the exons are very similar, while the introns are quite variable, even among genes that have the same number of introns (data not shown).

Analysis of the putative promoter regions (defined as the 2000 bp region upstream of the TSS) of the *GmLSD* genes identified an enrichment for putative *cis*-elements that are related to the response of the plant to abiotic and biotic stresses. Several regulatory *cis*-elements that are known to be responsive to hormones and to a variety of stress factors and signaling pathways were found in *GmLSD* gene promoters (Table 1). *Cis*-elements related to the response to light, to circadian control, and to the regulation of expression in the meristem, the seed and the endosperm were also identified (data not shown). Two putative *cis*-elements involved in the abscisic acid (ABA) response were also identified. The abscisic acid-responsive element (ABRE) and the salicylic acid-responsive element (TCA) were both found in the putative promoters of the *GmLSD2*, *GmLSD3* and *GmLSD7* genes, while the corresponding regions of *GmLSD5* and *GmLSD6* contained two and four copies of the TCA element,

respectively. The defense and stress responsive element (TC-rich repeat) was found in the promoters of all *GmLSD* genes except *GmLSD8*, and between one and five copies of this element were present in these promoters. Moreover, two sequences involved in the response to signaling by the methyl jasmonate hormone (MeJA), namely, the CGTCA and TGAC elements, were identified in the putative promoters of the *GmLSD2*, *GmLSD3*, *GmLSD4*, *GmLSD7* and *GmLSD8* genes. The gibberellin hormone response element (GARE) was found in *GmLSD1*, *GmLSD2*, *GmLSD3* and *GmLSD7*, while the ethylene-responsive element (ERE) was found in the upstream regions of the *GmLSD3*, *GmLSD4*, *GmLSD5*, *GmLSD7* and *GmLSD8* genes. In addition, *cis*-elements related to the specific response to pathogen attack were also identified. The Box-W1 *cis*-element was found in the putative promoters of both *GmLSD3* and *GmLSD5*, whereas the Elicitor-box 3 (ELI-box3) element was found exclusively in the *GmLSD8* gene. The wound-responsive element (WUN) was observed only upstream of the *GmLSD4* gene. The MBSII element, an MYB transcription factor binding site found in flavonoid biosynthesis genes, was found only in the putative promoters of the *GmLSD3* and *GmLSD6* genes. With regard to abiotic stress response *cis*-elements, the MBS element, an MYB transcription factor binding site associated with drought response, was identified in the *GmLSD1*, *GmLSD3*, *GmLSD4*, *GmLSD5* and *GmLSD6* genes, but no further elements of this class were found. Overall, the promoter region of the *GmLSD3* gene exhibited the greatest diversity in its putative *cis*-elements related to stress responses because only the ELI-box 3 and WUN elements were not found in its sequence, while the *GmLSD7* promoter contained the greatest number of *cis*-elements.

Modulation of the expression of *GmLSD* genes varies in different organs and under different stress conditions

We examined the expression profile of *GmLSD* genes in different organs (Fig. 5). For each gene, the lowest transcript level detected was used to normalize the transcript levels in other organs and thereby to quantify transcript accumulation. Our statistical analysis indicated that there was a significant association ($p<0.001$) between the organ factors, except *GmLSD6*.

The *GmLSD1* gene had higher transcript levels in roots, stems, leaves and flowers after fertilization, while *GmLSD2* showed higher expression levels in roots, stems and flowers after fertilization. The transcript levels of the *GmLSD3* and *GmLSD5* genes were increased in leaves, while those of *GmLSD4* were increased in roots, stems and pods. Transcripts of *GmLSD6* were detected only in roots and flowers after fertilization. The expression level of *GmLSD7* was higher in stems, pods and leaves. Lastly, the level of *GmLSD8* transcription was elevated in stems. Interestingly, under our experimental conditions, the transcripts of most genes were not detected in flowers before fertilization. Exceptions to this were the *GmLSD7* and *GmLSD8* genes, in which the relative expression levels were statistically similar.

We also performed a comparative analysis of all *GmLSD* genes in each organ (ESM Fig. S1). For each organ, the gene exhibiting the lowest expression level was used to normalize the transcript levels of the other genes. The statistical analysis indicated significant differences among the transcript levels of the different *GmLSD* genes in each organ ($p<0.001$), with the exception of flowers, after fertilization. The *GmLSD8* gene exhibited peak transcript levels in all organs analyzed, but other *GmLSD* genes had the same high level of expression in leaves, stems and flowers before fertilization.

To detect changes in *GmLSD* gene expression patterns occurring in response to *P. pachyrhizi* infection, two independent experiments were carried out. In experiment I, the expression levels of *GmLSD* genes during the course of a fungal infection were analyzed by RT-qPCR, and the responses of the EMBRAPA48 susceptible cultivar and the PI561356 resistant genotype were compared (Fig. 6). The transcript levels in mock-inoculated plants were used to normalize the transcript levels detected in inoculated plants. Plant tissues were sampled at a series of time-points that comprised the plant basal response (in which the first peak of plant gene expression occurs), fungal appressorium formation and epidermal cell penetration (until 12 hpi), the quiescent period between 24 and 48 hpi (when the fungus proceeds the early infection process but does not cause strong differential gene expression in the plant) and the time-point (usually after 72 hpi) at which a second round of gene expression can occur in the plant due to fungal colonization or the formation of lesions or uredinia (van de Mortel et al. 2007; Schneider et al. 2011). Our statistical analysis indicated a significant interaction among genotype, time-point and treatment for the most *GmLSD*

genes ($p<0.05$). *GmLSD3*, *GmLSD5* and *GmLSD7* were notable exceptions to this: a significant interaction ($p<0.05$) was found only between the treatment condition and the time-point for these genes. *GmLSD1* expression was induced only in PI561356 at 12, 24 and 48 hpi. Increased expression of the *GmLSD2* gene was detected only at 12 hpi in the EMBRAPA48 cultivar and at 24 hpi in PI561356 plants. *GmLSD3* was induced at 48 hpi in EMBRAPA48, while in PI561356 it was induced at 48 hpi, repressed at 96 hpi and induced again at 192 hpi. Expression of the *GmLSD4* gene was only induced in PI561356 plants and only at 48 hpi. *GmLSD5* expression was induced at 48 hpi in EMBRAPA48 and at 24 and 48 hpi in PI561356. The *GmLSD6* gene exhibited an induction response at 192 hpi in EMBRAPA48 plants, while it was found to be induced at 12, 24, 48 and 96 hpi in the PI561356 genotype. No differential expression was observed for *GmLSD7* in either genotype. Finally, the *GmLSD8* gene was induced at 24 and 48 hpi, but only in PI561356 plants.

Regarding experiment II, the expression profile of the *GmLSD* gene family was examined by RNA-seq analysis of microdissected lesions taken from soybean plants of the disease-susceptible BRS231 and disease-resistant PI561356 genotypes at ten days after infection (ESM Fig. S2). This time-point corresponds to the period during which it is possible to observe symptoms of the disease and at which the first uredospores are released, which signifies the end of the life cycle of the pathogen (Koch et al. 1983; Tremblay et al. 2009). Statistical analysis showed no difference in gene expression levels between the cultivars. *GmLSD1*, *GmLSD3* and *GmLSD4* were expressed in plants of both genotypes, while *GmLSD2*, *GmLSD5* and *GmLSD7* expression was not detected for either genotype in this experiment. *GmLSD6* and *GmLSD8* were expressed only in plants of the resistant PI561356 genotype.

The expression profile of the *GmLSD* genes under drought conditions was examined in the roots and leaves of the BR16 (drought sensitive) and EMBRAPA48 (drought tolerant) cultivars at a series of time-points during the 150 minutes after the beginning of drought stress. Fig 7 and Fig 8 provide details of the changes in *GmLSD* gene expression observed in the leaves and roots, respectively. Expression data for *GmLSD6* in roots were not submitted to statistical analysis because the transcripts were not detected at some of the time-points analyzed. A significant interaction between the genotype and the time-point was observed for all other *GmLSD* genes in the leaves ($p<0.01$) and the roots ($p<0.05$). The transcript levels in plants at T0 were

used to normalize the transcript levels for the subsequent time-points. Expression of *GmLSD1* was induced at T125 and T150 in the leaves of both cultivars. In the roots, *GmLSD1* it was not induced in BR16, but was induced at T100 and T150 in EMBRAPA48. *GmLSD2* expression was induced in the leaves at T50, T100 and T125 in BR16 and at T150 in EMBRAPA48. In contrast, *GmLSD2* was not induced in the roots of BR16, while in EMBRAPA48 it was induced at T75, T100 and T150 in this organ. *GmLSD3* transcription was induced at T25, T50, T100, T125 and T150 in the leaves of BR16 but only at T100 in those of the EMBRAPA48 cultivar, while in the roots *GmLSD3* was induced at T100 and T150 in BR16 and at T75, T100 and T150 in EMBRAPA48. *GmLSD4* expression was induced in leaves at T50, T100, T125 and T150 in BR16 plants and at T75, T125 and T150 in plants of the EMBRAPA48 cultivar. In the roots, expression of the *GmLSD4* gene was detected only in the BR16 cultivar, in which it was induced at T150. *GmLSD5* was not differentially expressed in the leaves and roots of BR16. However, we found that *GmLSD5* transcription was repressed in the leaves of EMBRAPA48 plants, except at T150. On the other hand, *GmLSD5* expression was induced only at T150 in the EMBRAPA48 roots. Transcripts from the *GmLSD6* gene were not detected in the leaves of the EMBRAPA48 cultivar, and the expression of this gene was found to be repressed at T25, T50 and T100 in BR16 leaves. In the roots, *GmLSD6* was expressed at T0, T25 T50 and T125 in BR16 and at T0, T75 and T150 in EMBRAPA48. The *GmLSD*7 gene was repressed at T25 and T150 in the leaves of BR16 plants and at T100, T125 and T150 in those of EMBRAPA48 plants, while in the roots *GmLSD7* it was induced only at T150 in BR16 plants and was not differentially expressed in EMBRAPA48 plants. *GmLSD8* was not differentially expressed in BR16 leaves from T0 to T125 but was found to be repressed at T150, while in EMBRAPA48 leaves it was repressed at T25, T100 and T125. In the roots, no differential expression of *GmLSD8* was observed for the EMBRAPA48 cultivar, but an induction of *GmLSD8* expression in this organ was observed at T150 in BR16 plants.

Discussion

Here, we report the identification of the LSD gene family in the available eukaryotic genomes. Our data show that the *LSD* genes are exclusive to the

Viridiplantae, suggesting that LSD proteins might have emerged and evolved as a strategy to subvert specific cell death modalities in plants. The diversification of the LSD gene family seems to have occurred in the Embryophyte clade, whose species possess greater numbers of *LSD* genes than are found in basal organisms. The fact that soybean and *Linum usitatissimum* have the greatest numbers of *LSD* genes is not surprising because soybean underwent two genome duplication events approximately 59 million and 13 million years ago (Schmutz et al. 2010) and *Linum usitatissimum* underwent a whole genome duplication approximately 5-9 million years ago (Wang et al. 2012).

The loss and gain of LSD domains support the theory that plant genomes have passed by dynamic and progressive evolution (Zhang et al. 2012). In our analysis of *LSD* genes in algae and plants, we observed that a gene structure consisting of three LSD domains represents the ancestral condition because it is present in algae and other basal organisms (Fig. 1). Although the majority of proteins have three LSD domains, the identification of genes with one or two LSD domains indicates that these structures have also been equally conserved during evolution. However, the consequence of the number of LSD domains on the function of the *LSD* genes remains unknown. Xu and He (2007) reported an investigation of LSD1-*like 2* of *Oryza sativa* (*OsLOL2*), a sequence with two LSD domains. Transgenic rice expressing the antisense sequence of *OsLOL2* showed a dwarf phenotype and lower resistance to pathogen attack. However, the dwarfism was restored by exogenous application of gibberellin (GA), suggesting that *OsLOL2* supports the regulation of GA biosynthesis, playing a role in the growth of rice plants. Moreover, the overexpression of *OsLOL2* conferred resistance against bacteria and induced a hypersensitive response in tobacco (Bhatti et al. 2008). Thus, the presence of three LSD domains is not a requirement for a role in pathogen and stress responses in rice. The *At*LSD1 protein, which contains three LSD domains, interacts with *Arabidopsis* Metacaspase 1 (AtMC1) (Coll et al. 2010) and with *Arabidopsis* basic region leucine zipper 10 (bZIP10) (Kaminaka et al. 2006) through its second and third LSD domains only. However, although *Os*LOL2 and *At*LSD1 are able to function with just two LSD domains, the presence of three LSD domains appears to be essential for certain functions of other LSD proteins. An example is the LSD1 protein of *Pisum sativa* (*Psa*LSD1), in which the three domains are necessary for nuclear export of the protein (He et al. 2011a). Interestingly,

although these three aforementioned LSD proteins can function with two or three LSD domains, the third domain (the C terminal) is maintained in all three cases. We observed that the third domain was present in all species and proteins analyzed (data not shown), which provides further evidence of the maintenance of this domain during *LSD* family evolution and suggests that it could be essential for LSD protein function.

Protein domains are structural, functional and evolutionary units and tend to be more stable than the surrounding regions of a protein during evolution (Zhang et al. 2012). Our analysis of consensus sequences showed that the structure and composition of the three LSD domains are highly conserved in all species analyzed (Fig. 2). The cysteine residues are conserved in all domains and species analyzed, forming the typical structure C2C2, which confirms the classification of the LSD domain as a C2C2-type zinc finger. This amino acid is a zinc-chelating residue involved in the binding of zinc ions, which is extremely important for the tertiary structure of the LSD domain (He et al. 2011b). Cysteine residues play important roles in protecting the cell against oxidative damage through the thiol functional group in the active sites of proteins (Requejo et al. 2010). Thus, the function of *At*LSD1 in oxidative stress (Jabs et al. 1996; Kliebenstein et al. 1999; Mateo et al. 2004) might be mediated by the conserved cysteine residues in the protein, which could explain the high degree of maintenance of the LSD domains encoded by the LSD gene family. Overall, the high degree of sequence conservation in LSD domains throughout plant evolution suggests that the maintenance of these sequences is important to ensure the functionality of *LSD* genes and their involvement in biological processes.

Our phylogenetic analyses demonstrated that the distribution of the *LSD* genes among the Viridiplantae species reflects the number of LSD domains they encode, and this distribution was well supported by the posterior probability values (Fig. 3). Additionally, the incongruence observed in the location of Rco2 (with two LSD domains) could be explained by the fact that while other proteins with two LSD domains comprise the second and third domains (i.e., the core and C-terminal domains), Rco2 is unique in comprising the first and third LSD domains (i.e., the N- and C-terminal domains) (data not shown). Thus, despite the absence of the second domain, Rco2 most closely resembles proteins with three LSD domains.

Monocot and eudicot *LSD* genes were interspersed along the phylogenetic tree, suggesting that the diversification of the *LSD* gene family occurred in the ancestral

lineage of the monocots and eudicots. With regard to *GmLSD* genes, we observed that even the evolutionarily closest members showed distinct behavior in our expression analysis. The upregulation of *GmLSD6* expression and the absence of modulation of *GmLSD7* expression in response to *P. pachyrhizi* infection this in experiment I are an example of this. The presence of different putative *cis*-elements in the promoters of these genes (Table 1) suggests a possible explanation of this behavior, as it might lead to transcriptional neofunctionalization or subfunctionalization of gene family members (Haberer et al. 2004).

There is a strong tendency towards gene duplication and retention in stress responsive genes (Zou et al. 2009). After duplication, pseudogenization is the most common fate of duplicated genes, although they can alternatively undergo neofunctionalization (when one copy acquires a novel function) or subfunctionalization (in which both copies are mutated and adopt complementary functions) (Cagliari et al. 2011). Both neofunctionalization and subfunctionalization are required to explain the retention of some duplicated genes (Zou et al. 2009). In the course of our EST database searches, we found ESTs for the majority of the putative *LSD* genes, which suggests that all the LSD structures that we found (consisting of one, two or three LSD domains) are likely to be important in plant development. Based on the apparent process of evolution from an ancestral protein structure consisting of three LSD domains to structures consisting of one or two such domains, we suggest that even though the loss of LSD domains has occurred during LSD gene family evolution, the basic function of these proteins has been maintained. The three structural classes observed might represent an evolutionary innovation that was required for the further development of diverse plant species, as well as an evolutionary adaptation to diverse environmental conditions and to biotic and abiotic stresses in particular.

In recent years, extensive promoter analyses have identified a large number of *cis*-elements that are components in the transcriptional regulatory networks that regulate biological processes such as development and the responses to stresses and hormones (Mochida et al. 2009). In soybean, *in silico* analysis of promoter regions has previously been reported for transcription factors (Mochida et al. 2009), the *chalcone synthases 7* (*CHS7*) and *CHS8* (Yi et al. 2010) and the *Bowman-Birk protease inhibitor* (de Almeida Barros et al. 2011). The analysis of *cis*-elements can be useful in

predicting the stress responsive genes, which may then be used in the genetic engineering of plants better suited to conditions of stress (Tran and Mochida 2010). Genes that are differentially expressed in response to various environmental stimuli will possess a greater number of distinct regulatory elements than are found in genes that respond to fewer environmental factors (Walther et al. 2007). In our *in silico* analysis, several *cis*-elements involved in different stress responses were identified (Table 1), which indicates that the *LSD* genes might respond to a variety of signaling pathways when plants are under stress conditions. The ABRE is a major *cis*-element involved in the response to osmotic stress (Yamaguchi-Shinozaki and Shinozaki 2006). This element interacts with the DRE *cis*-element in the *Rd29A* promoter region and induces a response to drought, high salt and cold stresses in addition to ABA treatment (Narusaka et al. 2003). The presence of the ABRE *cis*-element in the putative promoters of *GmLSD2, GmLSD3* and *GmLSD7,* together with the modulation seen in their expression in response to drought, suggests a possible stress response function for these genes that may be regulated by ABA.

At least one copy of the TC-rich repeat element, which is related to pathogen defense and stress response, was identified in the majority of *GmLSD* genes. Moreover, the number of copies of the TC-rich repeat ranged from one to five, which could explain the differences in modulation observed for each *GmLSD* gene in response to drought and *P. pachyrhizi* infection because the number of copies of a *cis*-element in the promoter region affects the function of these regulatory sequences (Mehrotra et al. 2005). According to Mehrotra et al. (2005), the ACGT *cis*-element functioned as a positive activator of transcription when tobacco leaves were transiently transformed with one or two copies, while in the case of the GT *cis*-element, a second copy interfered negatively with the positive effect of the first copy.

The presence of Box-W1, ERE, TCA and MBS elements in the promoter regions of the *GmLSD3* and *GmLSD5* genes indicates that these genes might respond to fungal attack, ethylene, ABA, and drought stress. In our experiments, the expression levels of *GmLSD3* and *GmLSD5* were modulated in response to drought and fungal infection, which was consistent with the *cis*-elements identified. Similar results have been reported for the *cotton mitogen-activated protein kinase 16* (*MPK16)* gene, wherein these *cis*-elements were identified and the expression of the gene was modulated in

response to pathogens, drought stress and additional molecules such as SA, MeJA, and ABA (Shi et al. 2011).

The highly significant positive correlation between the increase in the density of *cis*-elements upstream of the TSS and the number of conditions in which a gene was differentially regulated has already been reported elsewhere (Walther et al. 2007). Our *in silico* analysis indicated that the promoter region of *GmLSD3* exhibited greater diversity in its putative *cis*-elements than was found in other *GmLSD* genes, which suggests that this gene could respond to a wide range of environmental conditions. In fact, we observed modulation of *GmLSD3* expression in all experiments performed. However, additional experiments are necessary to demonstrate the *in vivo* function of the *cis*-elements we have identified in *GmLSD* genes.

To our knowledge, ours is the first report describing a comprehensive expression analysis of all of the *LSD* gene family members in a species. The previously reported expression analysis of *AtLSD1*, *AtLOL1* and *OsLSD1* showed that these genes are constitutively expressed in plant organs (Epple et al. 2003; Wang et al. 2005), while our data showed that the expression of the *GmLSD* genes is organ-dependent (Fig. 5). A similarly variable expression pattern was also reported for the *LOL1* gene in the developed shoots of *Bambusa oldhamii* (Yeh et al. 2011), which suggests that *LSD* genes can present a distinct modulation pattern depending on the species and organs analyzed.

In terms of organ-specific expression, a notable observation is that the majority of the *GmLSD* genes were not expressed in the flowers before fertilization, whereas all *GmLSD* members were expressed in the flowers after fertilization. To our knowledge, *LSD* genes have not previously been shown to be involved in the flowering process, and thus, further analysis of this correlation is important because it has direct consequences on the yield of crop plants such as soybeans (Jung et al. 2012).

The *GmLSD8* gene was highly expressed relative to other *GmLSD* genes (except in leaves, stems and flowers before fertilization). Interestingly, *GmLSD8* is a unique *GmLSD* gene with two zinc finger LSD domains and was located in a separate group in our phylogenetic analysis (Fig. 3). Beyond its role in responding to *P. pachyrhizi* infection and drought stress, *GmLSD8* may have evolved novel biological functions. Its ubiquitous expression in the majority of soybean organs points to putative roles in

different regulatory pathways during plant development. Further experiments are necessary to better understand the role of *GmLSD8* in soybean.

Studies have suggested a distinct biphasic response of genes against *P. pachyrhizi* infection. The first peak in this response, which occurs within the first 12 hpi, is related to basal defense and is a non-specific modulation that is transient or is suppressed when the early steps in colonization are completed in both the susceptible and the resistant genotypes (van de Mortel et al. 2007). A second peak in gene expression occurs early in resistant genotypes, approximately 72 hpi, while in susceptible genotypes, the response is later and continuing unaltered until 96 hpi and allows rapid fungal growth (van de Mortel et al. 2007; Schneider et al. 2011). Our data suggest that *GmLSD1*, *GmLSD2*, *GmLSD4*, *GmLSD5* and *GmLSD8* may be involved in the basal response because their expression was modulated at early time points and returned to levels similar to those of mock-treated controls after 48 hpi (Fig. 6). In contrast, the modulation of *GmLSD3* and *GmLSD6* expression was maintained after 48 hpi, which suggests their involvement in the networks of basal and specific defense against fungal attack. Panthee et al. (2007) performed a microarray analysis of soybean plants in the V2 stage at 72 hpi and reported that the majority of the up-regulated genes were related to defense and stress signaling, which supports our hypothesis that *GmLSD3* and *GmLSD6* have specific functions against the development and expansion of ASR infection.

RNA-Seq analysis of microdissected lesions allowed the identification of *GmLSD* genes expressed inside rust infection sites. In their microarray analysis of microdissected palisade and mesophyll cells of susceptible soybean plants 10 days after infection, Tremblay et al. (2010) reported many up-regulated genes associated with basal defense and down-regulated genes associated with metabolic pathways, thereby showing that *P. pachyrhizi* strongly affects plant metabolism and suggesting that soybean futilely fights to establish a resistance response, even in the advanced stages of an infection. Thus, *GmLSD1*, *GmLSD3* and *GmLSD4* expression in both genotypes could be related to the basal defense response against pathogen attack. By contrast, the specific response involving the *GmLSD6* and *GmLSD8* genes in the resistant PI561356 genotype seems to be related to the maintenance of plant protection against fungal penetration.

In regard to the two different gene expression responses seen in plants infected by *P. pachyrhizi*, we have shown that most *GmLSD* genes are modulated in response to ASR. Our data indicate that *GmLSD1*, *GmLSD2*, *GmLSD4*, *GmLSD5* and *GmLSD8* might be involved in basal defense, while *GmLSD3* and *GmLSD6* might be involved in both the basal and the specific responses to fungal infection and thereby prevent the expansion of ASR. However, additional experiments are necessary to elucidate the function of these genes in response to ASR.

Recently, RNA-Seq analysis of drought-treated and well-watered fertilized ovary and basal leaf meristem tissues of maize identified significant changes in the expression patterns of several known PCD genes under these conditions (Kakumanu et al. 2012). Among these PCD genes, the maize homolog of *AtLSD1* was reported to exhibit a decrease in its transcript abundance under drought stress. To our knowledge, this was the first study associating *LSD* gene expression with drought stress.

In this study, we characterized the changes in the expression patterns of *GmLSD* genes in both the leaves and the roots of plants subjected to drought conditions. These organs play an important role in signaling dehydration. The imbalance between the rate of transpiration in the leaves and the uptake of water by the roots is an important signal that triggers the response to drought stress (Aroca et al. 2011). The plasticity of roots in dry conditions is critical for the response of the plant to this stress because they are the primary site of drought signaling and play a fundamental role in recovery after stress (Porcel and Ruiz-Lozano 2004). Moreover, dry perception by roots leads to signal transduction throughout the plant via ABA or ethylene hormones, which induces the closing of the stomata (Dubos and Plomion, 2003). Although all *GmLSD* genes showed a response to drought stress, some of these responses are especially interesting. The detection of transcripts of *GmLSD6* in leaves and of *GmLSD4* in roots only in the drought-sensitive BR16 plants suggests that these genes are involved in a damage response that is specific to this cultivar. Similarly, the differential expression of *GmLSD5* in the leaves and of *GmLSD1* and *GmLSD2* in the roots of the drought-tolerant EMBRAPA48 plants following exposure to drought stress indicates an important function for these genes in drought tolerance in this soybean plant. Future analysis will be useful to clarify the specific role of these genes in the drought stress response.

In summary, we identified 117 genes belonging to the *LSD* gene family in Viridiplantae. We showed that the LSD domain is widely conserved in Viridiplantae and that the protein structure comprising three LSD domains represents the ancestral condition. The phylogenetic reconstruction generated new insights into the evolution of the LSD gene family in suggesting that its expansion occurred prior to the diversification of monocots and eudicots and that both neofunctionalization and subfunctionalization may have occurred during the evolution of this gene family. Expression analyses of *GmLSD* genes indicated that they are modulated during the response to *P. pachyrhizi* infection and drought stress. These findings indicate that these genes could be useful in efforts to improve the stress tolerance and disease resistance of the soybean plant.

The authors declare that they have no conflict of interest.

This work was supported by grants from the GENOSOJA/CNPq and BIOTECSUR II/MCT.

References

Almeida BB, da Silva WG, Moreira MA. de Barros EG (2011) In silico characterization and expression analysis of the multigene family encoding the Bowman-Birk protease inhibitor in soybean. Mol Biol Rep 39:327-334

Almeida HO, Barbosa MO, Marques AE, Pereira THA, Júnior MJM, Tessarollo NG, Games PD, de Barros EG, tolf-Moreira R, Marcelino-Guimarães FC, Abdelnoor RV, Pereira PRG. Baracat-Pereira MC (2012) Differentially regulated induced resistance marker enzymes in soybean genotypes resistant and susceptible to Asian soybean rust. Pesquisa Agropecuária Brasileira 47:163-172

Aroca R, Porcel R, Ruiz-Lozano JM (2011) Regulation of root water uptake under abiotic stress conditions. J Exp Bot 63:43-57

Aviv DH, Rusterucci C, Holt BF III, Dietrich RA, Parker JE Dangl JL (2002) Runaway cell death, but not basal disease resistance, in lsd1 is SA- and NIM1/NPR1-dependent. Plant J 29:381-391

Bhatti KH, Xu C, Wu J, and He C (2008) Overexpression of rice OsLOL2 gene confers disease resistance in tobacco to *Pseudomonas syringae* pv. tabaci. Progress in Natural Science 18:807-812

Bromfield KR, Hartwig EE (1980) Resistance to soybean rust and mode of inheritance. Crop Sci 20:254-255.

Cagliari A, Turchetto-Zolet AC, Maraschin FS, Loss G, Margis R, Margis-Pinheiro M (2011) The Evolutionary History of CBF Transcription Factors: Gene Duplication of CCAAT – Binding Factors NF-Y in Plants In: Felix Friedberg (ed) Gene duplication, 1, pp 1-26

Cai S, Lashbrook CC (2006) Laser capture microdissection of plant cells from tape-transferred paraffin sections promotes recovery of structurally intact RNA for global gene profiling. Plant J 48:628-637

Cheng YW, Chan KL (1968) The breeding of 'Tainung 3' soybean. Journal of Taiwan Agricultural Research, Taiwan 17:30-35

Coll NS, Vercammen D, Smidler A, Clover C, Van Breusegem F, Dangl JL, Epple P (2010) *Arabidopsis* type I metacaspases control cell death. Science 330:1393-1397

Dietrich RA, Delaney TP, Uknes SJ, Ward ER, Ryals JA, Dangl JL (1994) *Arabidopsis* mutants simulating disease resistance response. Cell 77: 565-577

Dietrich RA, Richberg MH, Schmidt R, Dean C, Dangl 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

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214-221

Dubos C, Plomion C (2003) Identification of water-deficit responsive genes in maritime pine (Pinus pinaster Ait.) roots. Plant Mol Biol 51:249-262

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

Fehr WR, Caviness CE (1977) Stages on soybean development. Iowa State University/Cooperative Extention Service, Ames 11p (Special report, 80)

Garcia A, Calvo ES, de Souza Kiihl RA, Harada A, Hiromoto DM, Vieira LG (2008) Molecular mapping of soybean rust (*Phakopsora pachyrhizi*) resistance genes: discovery of a novel locus and alleles. Theor Appl Genet 117:545-553

Goellner K, Loehrer M, Langenbach C, Conrath U, Koch E, Schaffrath U (2010) *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust. Mol Plant Pathol 11:169-177

Haberer G, Hindemitt T, Meyers BC, Mayer KF (2004) Transcriptional similarities, dissimilarities, and conservation of cis-elements in duplicated genes of Arabidopsis. Plant Physiol, 136:3009-3022

Hartwig EE (1986) Identification of a fourth major gene conferring resistance to soybean rust. Crop Sci 26:1135-1136

He S, Huang K, Zhang X, Yu X, Huang P, An C (2011a) The LSD1-type zinc finger motifs of *Pisum sativa* LSD1 Are a novel nuclear localization signal and interact with Importin Alpha. PLoS One 6:e22131

He S, Tan G, Liu Q, Huang K, Ren J, Zhang X, Yu X, Huang P, An C (2011b) The LSD1-interacting protein GILP is a LITAF domain protein that negatively regulates hypersensitive cell death in *Arabidopsis*. PLoS One 6:e18750

Hidayat OO, Somaatmadia S (1977) Screening of soybean breeding lines for resistance to soybean rust (*Phakopsora pachyrhizi* Sydow). Soybean Rust Newsletter 1:9-22

Huang X, Li Y, Zhang X, Zuo J, Yang S (2010) The *Arabidopsis* LSD1 gene plays an important role in the regulation of low temperature-dependent cell death. New Phytol, 187, 301-312

Hunt MD, Delaney TP, Dietrich RA, Weymann KB, Dangl JL, Ryals JA (1997) Salicylate-independent lesion formation in *Arabidopsis lsd* mutants. Mol Plant Microbe Interact 10:531-536

Jabs T, Dietrich RA, Dangl JL (1996) Initiation of runaway cell death in an *Arabidopsis* mutant by extracellular superoxide. Science 273:1853-1856

Jian B. Liu B. Bi Y. Hou W, Wu C, Han T (2008) Validation of internal control for gene expression study in soybean by quantitative real-time PCR. BMC Mol Biol 9:1-14

Jung CH, Wong CE, Singh MB, Bhalla PL (2012) Comparative genomic analysis of soybean flowering genes. PLoS One 7:e38250

Kakumanu A, Ambavaram MM, Klumas C, Krishnan A, Batlang U, Myers E, Grene, R, Pereira A (2012) Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. Plant Physiol 160:846-867

Kaminaka H, Nake C, Epple P, Dittgen J, Schutze K, Chaban C, Holt BF III, Merkle T, Schafer E, Harter K, Dangl JL (2006) bZIP10-LSD1 antagonism modulates basal defense and cell death in *Arabidopsis* following infection. EMBO J 25:4400-4411

Kerk NM, Ceserani T, Tausta SL, Sussex IM, Nelson TM. (2003) Laser capture microdissection of cells from plant tissues. Plant Physiol 132:27-35

Kim KS, Unfried JR, Hyten DL, Frederick RD, Hartman GL, Nelson RL, Song Q, Diers BW (2012) Molecular mapping of soybean rust resistance in soybean accession PI 561356 and SNP haplotype analysis of the *Rpp*1 region in diverse germplasm. Theor Appl Genet 125:1339-1352

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

Koch E, Ebrahimnesbat F, Hoppe HH (1983) Light and electron-microscopic studies on the development of soybean rust (*Phakopsora pachyrhizi* Syd) in susceptible soybean leaves. Phytopathology 106:302-320

Kulcheski FR, de Oliveira LF, Molina LG, Almerao MP, Rodrigues FA, Marcolino J, Barbosa JF, Stolf-Moreira R, Nepomuceno AL, Marcelino-Guimaraes FC, Abdelnoor R, Nascimento LC, Carazzolle MF, Pereira GA, Margis R (2011) Identification of novel soybean microRNAs involved in abiotic and biotic stresses. BMC Genomics 12: 307:324

Ledrera R, Marino D, Larrainzar E, Gonzales EM, Aresse-Igor C (2007) Reduced carbon availability to bacteroids and elevated ureides in nodules, but not in shoots, are involved in the Nitrogen fixation response to early drought in soybean. Plant Physiology 145:539-546

Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J (2009) SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25:1966-1967

Li S, Smith JR, Ray JD, Frederick RD (2012) Identification of a new soybean rust resistance gene in PI 567102B. Theor Appl Genet 125:133-142

Libault M, Thibivilliers S, Bilgin DD, O R, Benitez M, Clough SJ, Stacey G (2008) Identification of four soybean reference genes for gene expression normalization. PLant Genome 1:44-54

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25:402-408

Manavalan LP, Guttikonda SK, Tran LS, Nguyen HT (2009) Physiological and molecular approaches to improve drought resistance in soybean. Plant Cell Physiol 50:1260-1276

Martins PK, Jordão BK, Yamanaka N, Farias JRB, Beneventi MA, Binneck E, Fuganti R, Stolf R, Nepomuceno AL (2008) Differential gene expression and mitotic cell analysis of the drought tolerant soybean (*Glycine max* L. Merrill Fabales, Fabaceae)

cultivar MG/BR46 (Conquista) under two water deficit induction systems. Genetics and Molecular Biology 31:512-521

Mateo A, Muhlenbock P, Rusterucci C, Chang CC, Miszalski Z, Karpinska B, Parker JE, Mullineaux PM, Karpinski S (2004) LESION SIMULATING DISEASE 1 is required for acclimation to conditions that promote excess excitation energy. Plant Physiol 136:2818-2830

Meckel L, Egli DB, Phillips RE, Radcliffe D, Leggett JE (1984) Effect of moisture stress on seed growth in soybeans. Agronomy Journal 75:1027-1031

Mehrotra R, Kiran K, Chaturvedi CP, Ansari SA, Lodhi N, Sawant , Tuli R (2005) Effect of copy number and spacing of the ACGT and GT *cis* elements on transient expression of minimal promoter in plants. J Genet 84:183-187

Mochida K, Yoshida T, Sakurai T, Yamaguchi-Shinozaki K, Shinozaki K, Tran LS (2009) *In silico* analysis of transcription factor repertoire and prediction of stress responsive transcription factors in soybean. DNA Res 16:353-369

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621-628

Muhlenbock P, Plaszczyca M, Mellerowicz E, Karpinski S (2007) Lysigenous aerenchyma formation in *Arabidopsis* is controlled by LESION SIMULATING DISEASE1. Plant Cell 19:3819-3830

Muhlenbock P, Szechynska-Hebda M, Plaszczyca M, Baudo M, Mateo A, Mullineaux PM, Parker JE, Karpinska B, Karpinski S (2008) Chloroplast signaling and LESION SIMULATING DISEASE1 regulate crosstalk between light acclimation and immunity in *Arabidopsis*. Plant Cell 20:2339-2356

Narusaka Y, Nakashima K, Shinwari ZK, Sakuma Y, Furihata T, Abe H, Narusaka M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Interaction between two *cis*-acting

elements, ABRE and DRE, in ABA-dependent expression of *Arabidopsis* rd29A gene in response to dehydration and high-salinity stresses. Plant J 34:137-148

Oya T, Nepomuceno AL, Neumaier N, Farias JRB, Tobita S, Ito O (2004) Drought tolerance characteristics of Brazilian soybean cultivars – evaluation and characterization of drought tolerance of various Brazilian soybean cultivars in the field. Plant Production Science 7:129-237

Panthee DR, Yuan JS, Wright DL, Marois JJ, Mailhot D, Stewart CN Jr (2007) Gene expression analysis in soybean in response to the causal agent of Asian soybean rust (*Phakopsora pachyrhizi* Sydow) in an early growth stage. Funct Integr Genomics 7:291-301

Porcel R, Ruiz-Lozano JM (2004) Arbuscular mycorrhizal influence on leaf water potential, solute accumulation, and oxidative stress in soybean plants subjected to drought stress. J Exp Bot 55:1743-1750

Requejo R. Hurd TR, Costa NJ, Murphy MP (2010) Cysteine residues exposed on protein surfaces are the dominant intramitochondrial thiol and may protect against oxidative damage. FEBS J 277:1465-1480

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178-183

Schneider KT, van de Mortel M, Bancroft TJ, Braun E, Nettleton D, Nelson RT, Frederick RD, Baum TJ, Graham MA, Whitham SA (2011) Biphasic gene expression changes elicited by *Phakopsora pachyrhizi* in soybean correlate with fungal penetration and haustoria formation. Plant Physiol 157:355-371

Shi J, Zhang L, An H, Wu C, Guo X (2011) GhMPK16, a novel stress-responsive group D MAPK gene from cotton, is involved in disease resistance and drought sensitivity. BMC Mol Biol 12:22-36

Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic analyses. Syst Biol 49:369-381

Singh BB, Thapliyal PN (1977) Breeding for resistance to soybean rust in India. FORD, R.E.; SINCLAIR, J.B. (Ed.). Rust of soybean: the problem and research needs. Urbana: University of Illinois, (INTSOY. Series, 12), 62-65

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 28(10):2731-2739

Thibaud-Nissen F, Ouyang S, Buell CR (2009) Identification and characterization of pseudogenes in the rice gene complement. BMC Genomics 10:317-329

Tran LS, Mochida K (2010) Functional genomics of soybean for improvement of productivity in adverse conditions. Funct Integr Genomics 10:447-462

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105-1111

Tremblay A, Li S, Scheffler BE, Matthews BF (2009) Laser capture microdissection and expressed sequence tag analysis of uredinia formed by *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust. Physiological and Molecular Plant Pathology 73:163-174

Tremblay A, Hosseini P, Alkharouf NW, Li S, Matthews BF (2010) Transcriptome analysis of a compatible response by *Glycine max* to *Phakopsora pachyrhizi* infection. Plant Science 179:183-193

van de Mortel M, Recknor JC, Graham MA, Nettleton D, Dittman JD, Nelson RT, Godoy CV, Abdelnoor RV, Almeida AM, Baum TJ, Whitham SA (2007) Distinct biphasic mRNA changes in response to Asian soybean rust infection. Mol Plant Microbe Interact 20:887-899

Walther D, Brunnemann R, Selbig J (2007) The regulatory code for transcriptional response diversity and its relation to genome structural properties in *A. thaliana*. PLoS Genet 3:e11

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 Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins S, Neutelings G, Datla R, Lambert G, Galbraith DW, Grassa CJ, Geraldes A, Cronk QC, Cullis C, Dash PK, Kumar PA, Cloutier S, Sharpe AG, Wong GK, Wang J, Deyholos MK (2012) The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. Plant J 72:461-473

Weymann K, Hunt M, Uknes S, Neuenschwander U, Lawton K, Steiner HY, Ryals J (1995) Suppression and Restoration of Lesion Formation in *Arabidopsis* lsd Mutants. Plant Cell 7:2013-2022

Wiebke-Strohm B, Pasquali G, Margis-Pinheiro M, Bencke M, Bucker-Neto L, Becker-Ritt AB, Martinelli AH, Rechenmacher C, Polacco JC, Stolf R, Marcelino FC, Abdelnoor RV, Homrich MS, Del Ponte EM, Carlini CR, De Carvalho MC, Bodanese-Zanettini MH (2012) Ubiquitous urease affects soybean susceptibility to fungi. Plant Mol Biol 79:75-87

Xu C, He C (2007) The rice OsLOL2 gene encodes a zinc finger protein involved in rice growth and disease resistance. Mol Genet Genomics 278:85-94

Yamaguchi-Shinozaki K, Shinozaki K (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu Rev Plant Biol 57:781-803

Yamanaka N, Lemos NG, Akamatsu H, Yamaoka Y, Silva DCG, Passianotto ALL, Abdelnoor RV, Soares RM, Suenaga K (2011) Soybean breeding materials useful for resistance to soybean rust in Brazil. JARQ 45:385-395

Yeh SH, Lin CS, Wu FH, Wang AY (2011) Analysis of the expression of BohLOL1, which encodes an LSD1-like zinc finger protein in *Bambusa oldhamii*. Planta 234:1179-1189

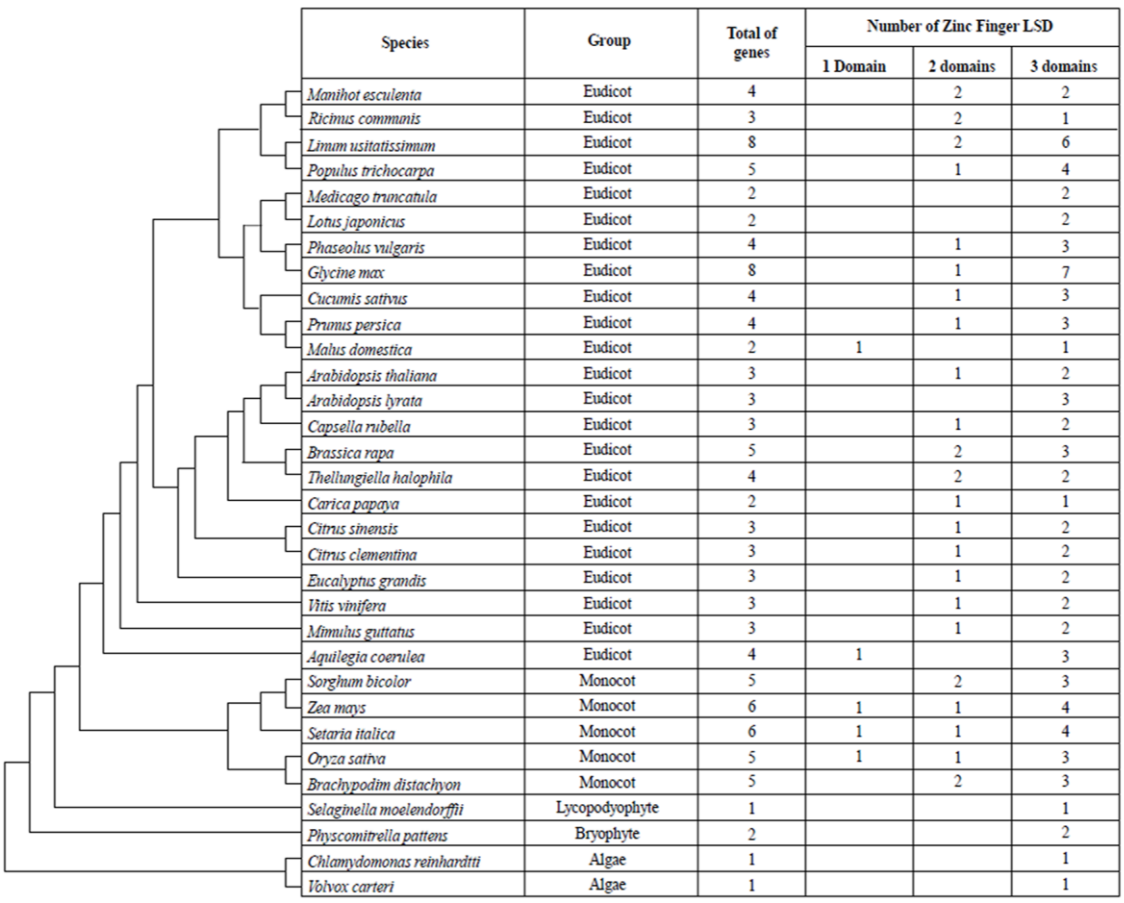
Yi J, Derynck MR, Chen L, Dhaubhadel S (2010) Differential expression of CHS7 and CHS8 genes in soybean. Planta 231:741-753

Zhang XC, Wang Z, Zhang X, Le MH, Sun J, Xu D, Cheng J, Stacey G (2012) Evolutionary dynamics of protein domain architecture in plants. BMC Evol Biol 12:6-17

Zhu L, Zhang Y, Zhang W, Yang S, Chen JQ, Tian D (2009) Patterns of exon-intron architecture variation of genes in eukaryotic genomes. BMC Genomics 10:47-58

Zou C, Lehti-Shiu MD, Thomashow M, Shiu SH (2009) Evolution of stress-regulated gene expression in duplicate genes of *Arabidopsis thaliana*. PLoS Genet 5:e1000581

Captions:



Species	Group	Total of genes	Number of Zinc Finger LSD		
			1 Domain	2 domains	3 domains
Manihot esculenta	Eudicot	4		2	2
Ricinus communis	Eudicot	3		2	1
Linum usitatissimum	Eudicot	8		2	6
Populus trichocarpa	Eudicot	5		1	4
Medicago truncatula	Eudicot	2			2
Lotus japonicus	Eudicot	2			2
Phaseolus vulgaris	Eudicot	4		1	3
Glycine max	Eudicot	8		1	7
Cucumis sativus	Eudicot	4		1	3
Prunus persica	Eudicot	4		1	3
Malus domestica	Eudicot	2	1		1
Arabidopsis thaliana	Eudicot	3		1	2
Arabidopsis lyrata	Eudicot	3			3
Capsella rubella	Eudicot	3		1	2
Brassica rapa	Eudicot	5		2	3
Thellungiella halophila	Eudicot	4		2	2
Carica papaya	Eudicot	2		1	1
Citrus sinensis	Eudicot	3		1	2
Citrus clementina	Eudicot	3		1	2
Eucalyptus grandis	Eudicot	3		1	2
Vitis vinifera	Eudicot	3		1	2
Mimulus guttatus	Eudicot	3		1	2
Aquilegia coerulea	Eudicot	4	1		3
Sorghum bicolor	Monocot	5		2	3
Zea mays	Monocot	6	1	1	4
Setaria italica	Monocot	6	1	1	4
Oryza sativa	Monocot	5	1	1	3
Brachypodim distachyon	Monocot	5		2	3
Selaginella moelendorffii	Lycopodyophyte	1			1
Physcomitrella pattens	Bryophyte	2			2
Chlamydomonas reinharditi	Algae	1			1
Volvox carteri	Algae	1			1

Fig. 1: The total number of *LSD* genes annotated per species and the number of proteins with one, two or three zinc finger LSD domains. Adapted from Phytozome (http://www.phytozome.org/).

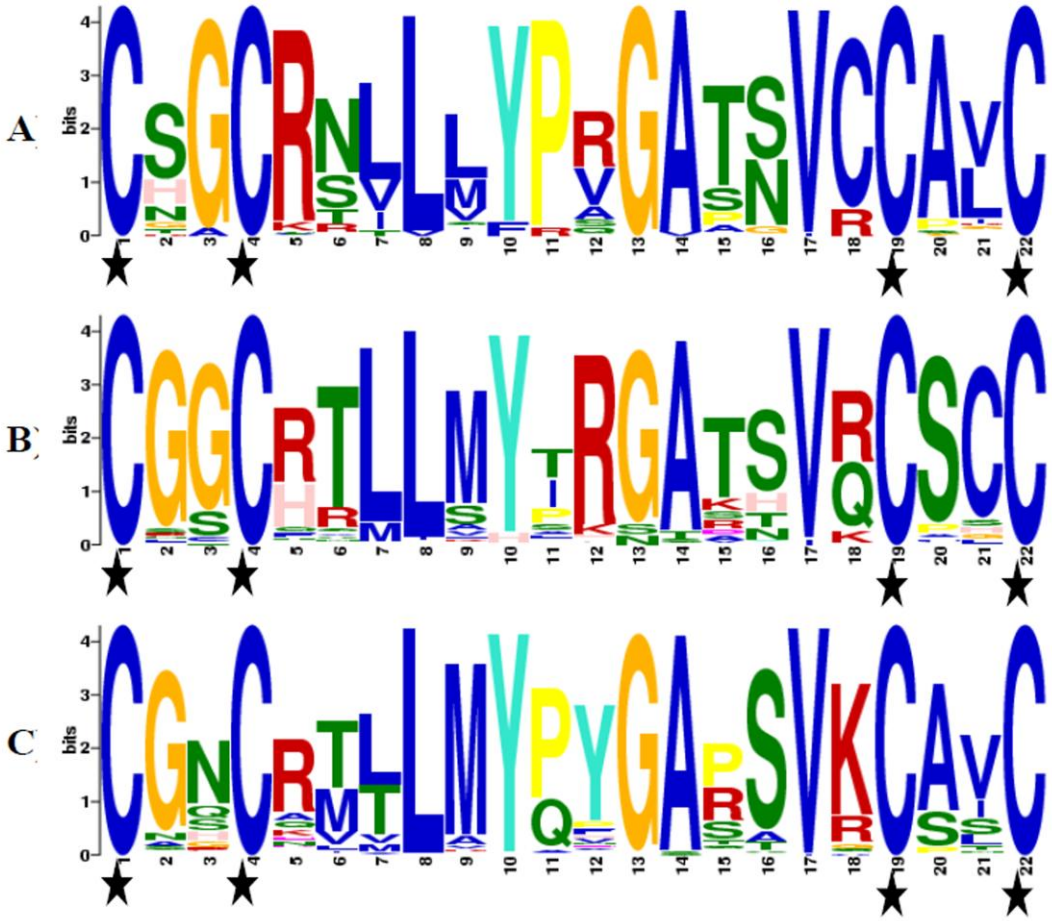


Fig. 2: Conservation analysis of the consensus sequence of the zinc finger LSD domains. Analysis of the 117 *LSD* genes identified was performed using the MEME suite. The overall height in each stack indicates the sequence conservation at each position. The height of each residue letter is proportional to the relative frequency of the corresponding residue. Amino acids are colored according to their chemical properties: green for polar, non-charged, non-aliphatic residues (N, Q, S and T), blue for the most hydrophobic residues (A, C, F, I, L, V and M), red for positively charged residues (K and R), pink for histidine (H), orange for glycine (G), yellow for proline (P) and turquoise for tyrosine (Y). Black stars indicate the conserved cysteine residues of the C2C2 zinc finger. (A) 84 sequences of the first domain (N-terminal). (B) 111 sequences of the second domain (core). (C) 117 sequences of the third domain (C-terminal).

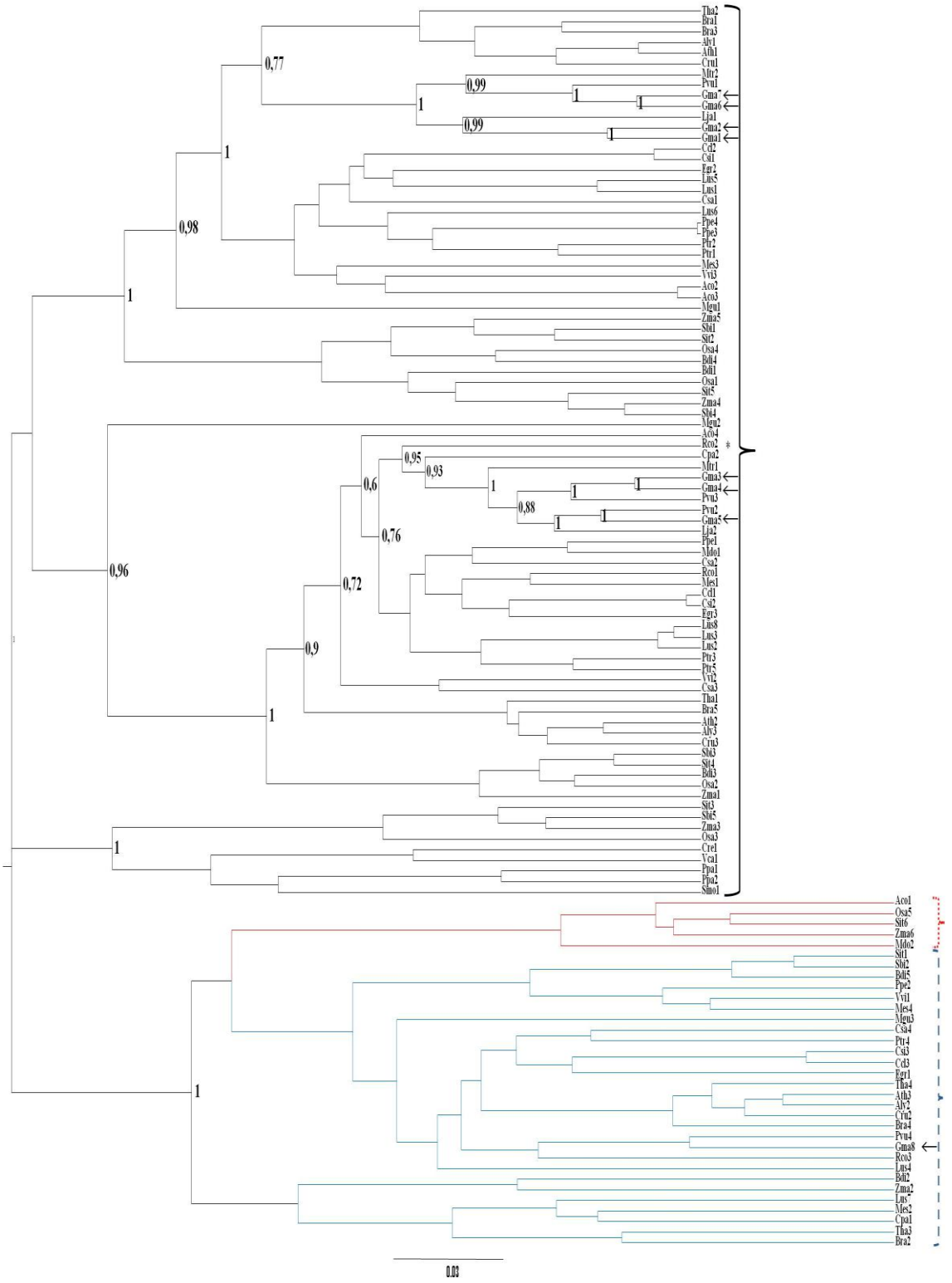

Tha2
Bra1
Bra3
Alr1
Ath1
Cru1
Mtr2
Pvu1
Gma7
Gma6
Lja1
Gma2
Gma1
Cdl2
Csi1
Egr2
Lus5
Lus1
Csa1
Lus6
Ppe4
Ppe3
Ptr2
Ptr1
Mes3
Vvi3
Aco2
Aco3
Mgu1
Zma5
Sbi1
Sit2
Osa4
Bdi4
Bdi1
Osa1
Sit5
Zma4
Sbi4
Mgu2
Aco4
Rco2
Cpa2
Mtr1
Gma3
Gma4
Pvu3
Pvu2
Gma5
Lja2
Ppe1
Mdo1
Csa2
Rco1
Mes1
Cdl1
Csi2
Egr3
Lus8
Lus3
Lus2
Ptr3
Ptr5
Vvi2
Csa3
Tha1
Bra5
Ath2
Alr3
Cru3
Sbi3
Sit4
Bdi3
Osa2
Zma1
Sit3
Sbi5
Zma3
Osa3
Cre1
Vca1
Ppa1
Ppa2
Sbi01
Aco1
Osa5
Sit6
Zma6
Mdo2
Sit1
Sbi2
Bdi5
Ppe2
Vvi1
Mes4
Mgu3
Csa4
Ptr4
Csi3
Cdl3
Egr1
Tha4
Ath3
Alr2
Cru2
Bra4
Pvu4
Gma8
Rco3
Lus4
Bdi2
Zma2
Lus*
Mes2
Cpa1
Tha3
Bra2
0,77
0,99
0,99
1
0,98
0,95
0,93
0,6
0,88
0,76
0,96
0,72
0,9
0.03

Fig. 3: Phylogenetic relationships among the 117 *LSD* genes identified. The unrooted tree was generated using BEAST v1.4.7 and visualized using Fig Tree v1.3.1. The Bayesian posterior probabilities are shown inside the branches. Taxa terminologies are abbreviated using the first letter of the genus and the first two letters of the species name as follows: *Manihot esculenta* (Mes), *Ricinus communis* (Rco), *Linum usitatissimum* (Lus), *Populus trichocarpa* (Ptr), *Medicago truncatula* (Mtr), *Lotus japonicus* (Lja), *Phaseolus vulgaris* (Pvu), *Glycine max* (Gma), *Cucumis sativus* (Csa), *Prunus persica* (Ppe), *Malus domestica* (Mdo), *Arabidopsis thaliana* (Ath), *Arabidopsis lyrata* (Aly), *Capsella rubella* (Cru), *Brassica rapa* (Bra), *Thellungiella halophila* (Tha), *Carica papaya* (Cpa), *Citrus sinensis* (Csi), *Citrus clementina* (Ccl), *Eucalyptus grandis* (Egr), *Vitis vinifera* (Vvi), *Mimulus guttatus* (Mgu), *Aquilegia coerulea* (Aco), *Sorghum bicolor* (Sbi), *Zea mays* (Zma), *Setaria italica* (Sit), *Oryza sativa* (Osa), *Brachypodium distachyon* (Bdi), *Selaginella mollendorffii* (Smo), *Physcomitella patens* (Ppa), *Chlamydomonas reinhardtii* (Cre) and *Volvox carteri* (Vca). The solid black shading indicates the sequences with three LSD domains. The red dotted shading indicates the sequences with one LSD domain. The blue dashed shading indicates the sequences with two LSD domains. The arrows indicate the *GmLSD* genes. * Incongruence was observed between the clustering pattern and the number of LSD domains encoded.

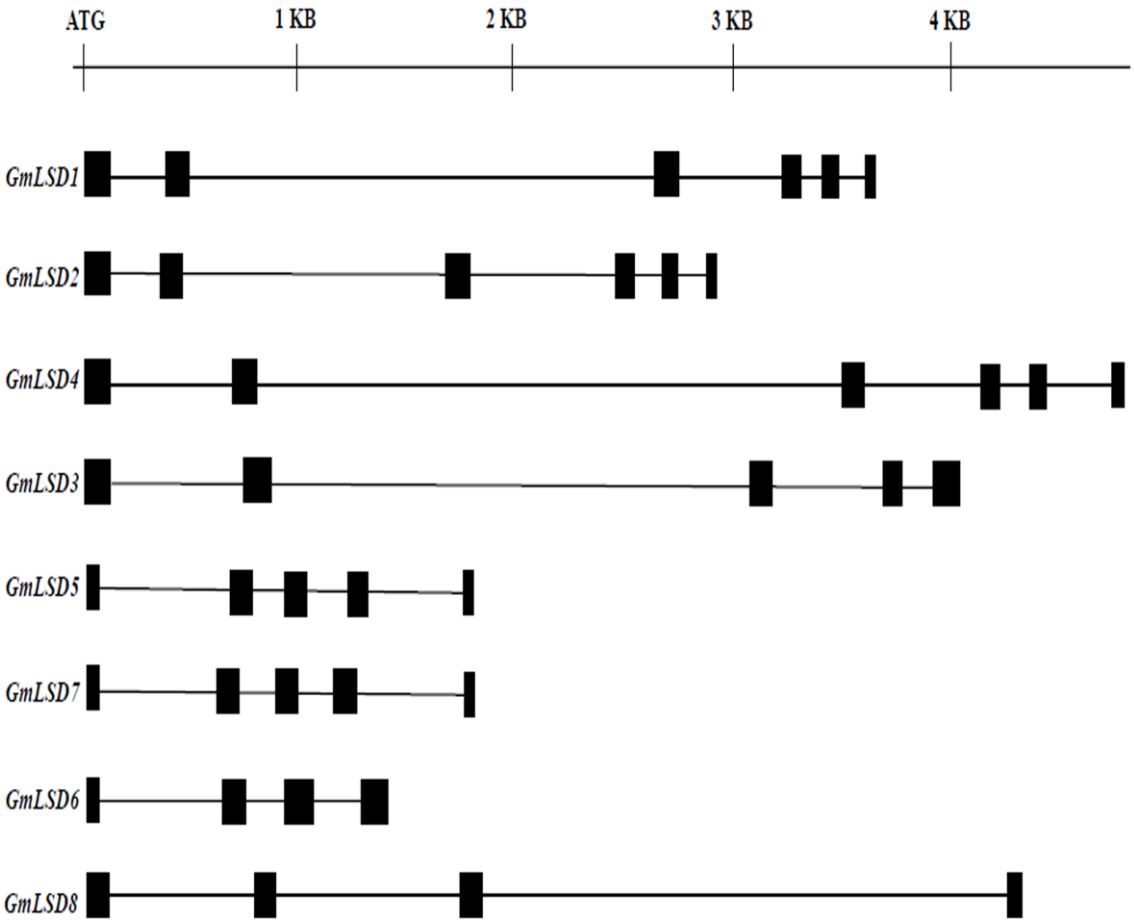


Fig. 4: Exon-intron structure of the *GmLSD* genes. The boxes represent the exons, and the lines connecting them represent the introns. The sequences are drawn to scale.

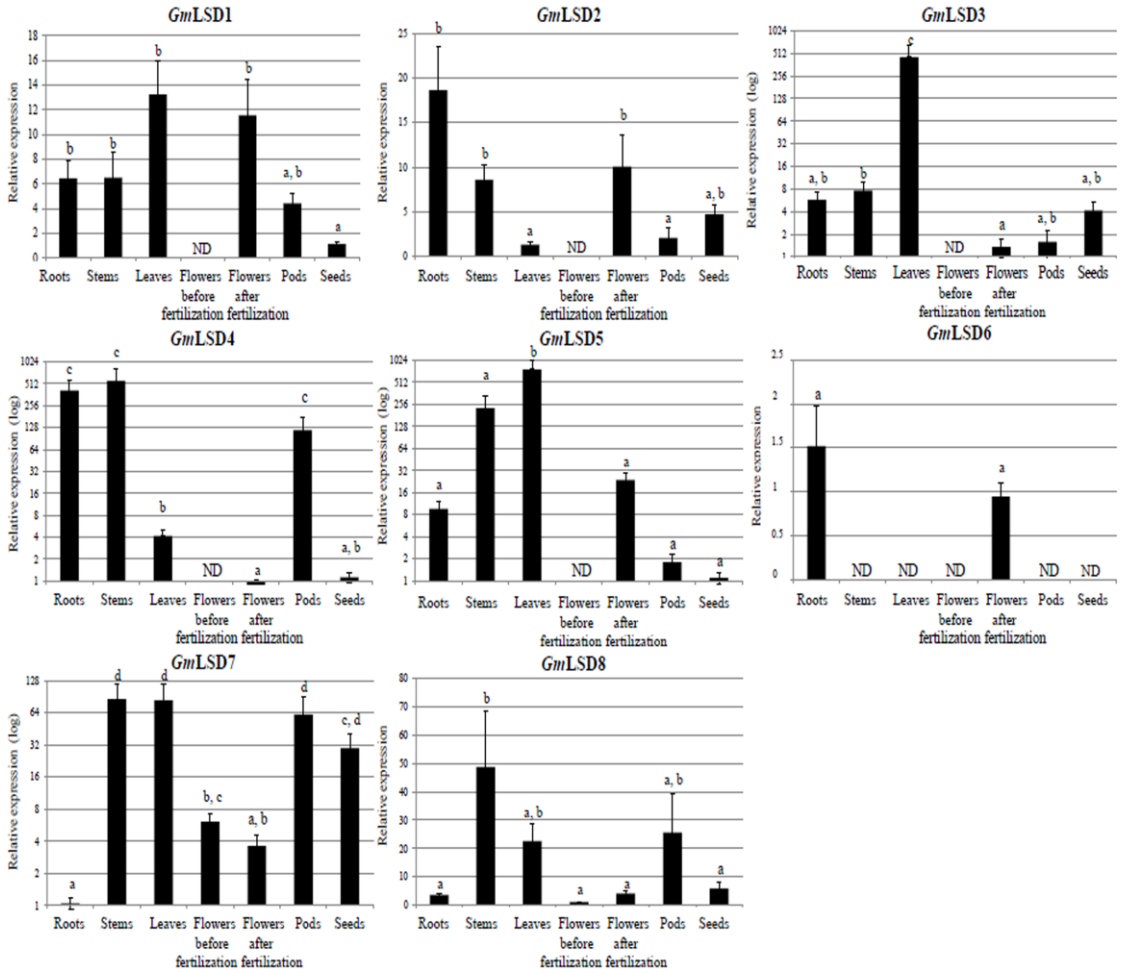


Fig. 5: Relative expression levels of the *GmLSD* genes in different organs of the Conquista Brazilian soybean cultivar. The values are the means of three biological replicates with four technical replicates each. The means that are labeled identically (with a letter) do not differ significantly (Tukey comparison test, $p<0.05$). The *ACTII*, *CYP2* and *metalloprotease* reference genes were used as internal controls to normalize for the amount of mRNA present in each sample. For each gene, the lowest transcript level was used to normalize the transcript levels in other organs. To allow a better comparison of groups of genes with large differences in their expression levels, some graphs of fold-change are shown in log2 scale, and this is indicated at the y axis. ND: not detected

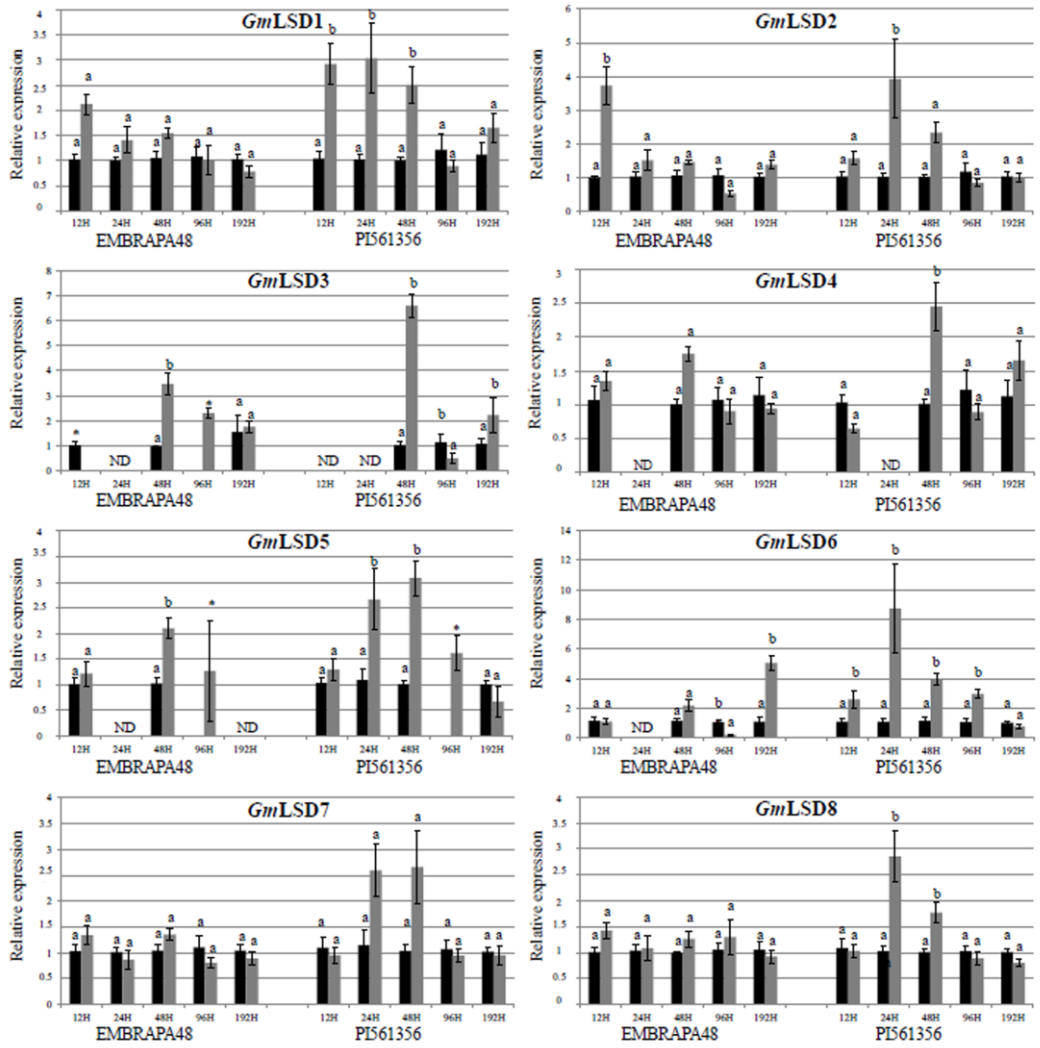


Fig. 6: *GmLSD* gene expression profiles in response to *Phakopsora pachyrhizi* infection in the leaves of EMBRAPA48 (susceptible) and PI561356 (resistant) soybean genotypes. The relative expression levels of *GmLSD* genes were measured by RT-qPCR at 12, 24, 48, 92 and 196 hpi (hours post-inoculation). The black bars represent the mock plants (non-infected), and the grey bars represent the infected plants. The values are the means of three biological replicates with four technical replicates each. The means that are labeled identically (with a letter) in the same cultivar do not differ significantly (Bonferroni multiple comparison test, $p<0.05$). The *f-box* and *metalloprotease* reference genes were used as internal controls to normalize for the amount of mRNA present in each sample. The transcript levels from the mock-inoculated plants were used to normalize the transcript levels from the inoculated plants. ND: not detected

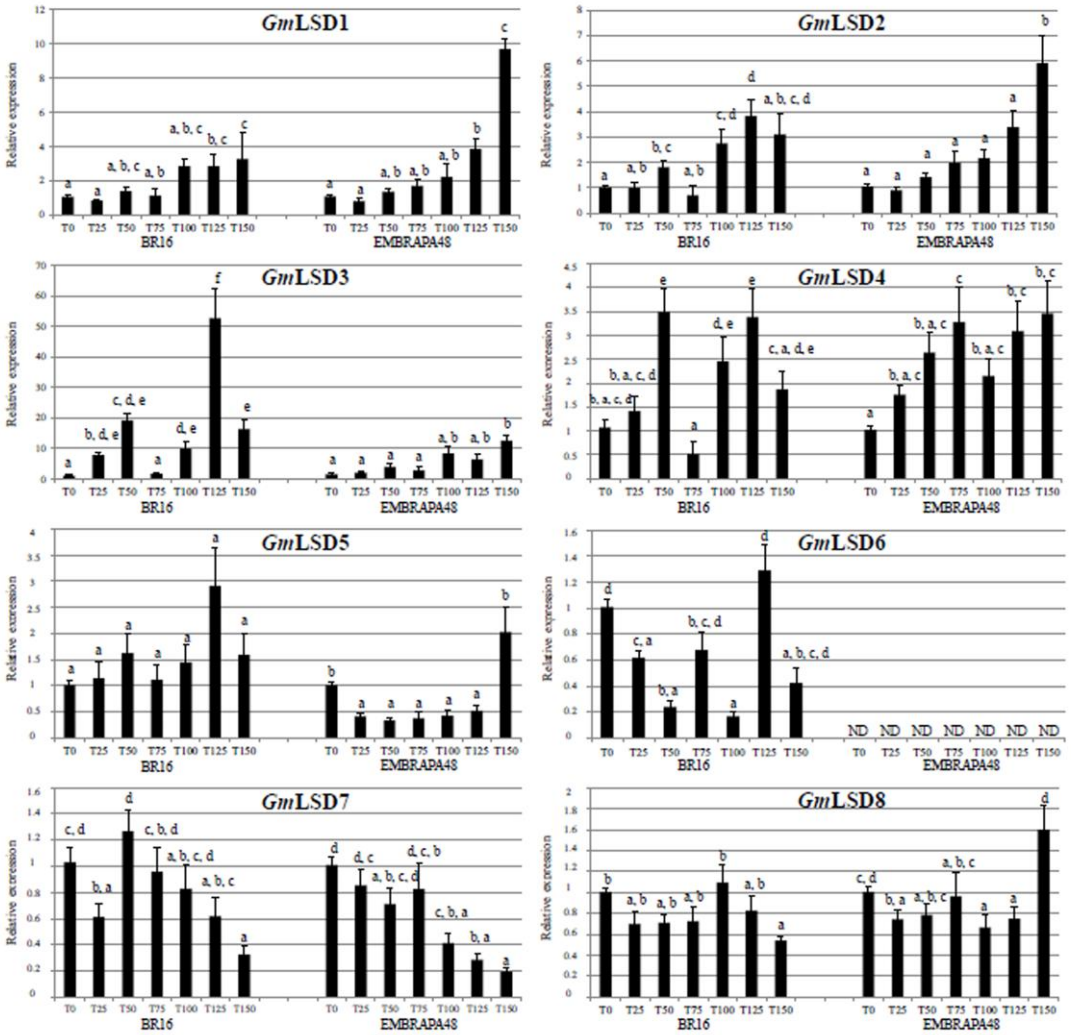


Fig. 7: *GmLSD* gene expression in the leaves of BR16 (sensitive) and EMBRAPA48 (tolerant) cultivars in response to drought stress. The relative expression levels of the *GmLSD* genes were measured by RT-qPCR at T0 (control), T25, T50, T75, T100, T125 and T150 minutes of drought stress. The values are the means of three biological replicates with four technical replicates each. The means that are labeled identically (with a letter) in the same cultivar do not differ significantly (Bonferroni multiple comparison test, $p<0.05$). The *f-box*, *ACTII* and *ELF* reference genes were used as internal controls to normalize for the amount of mRNA present in each sample. The transcript levels from the plants at T0 were used to normalize the transcript levels from the plants subjected to drought stress. ND: not detected

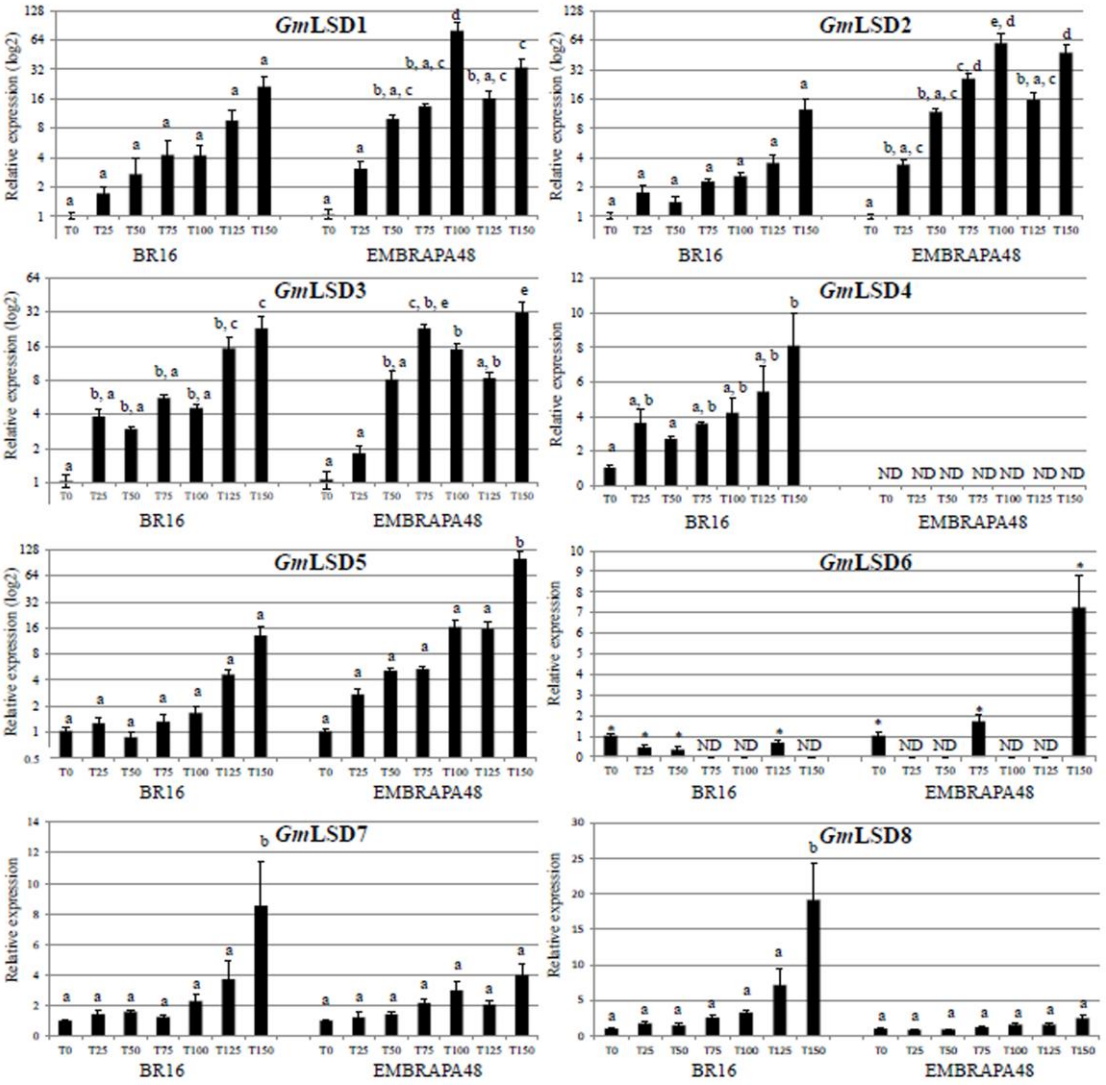
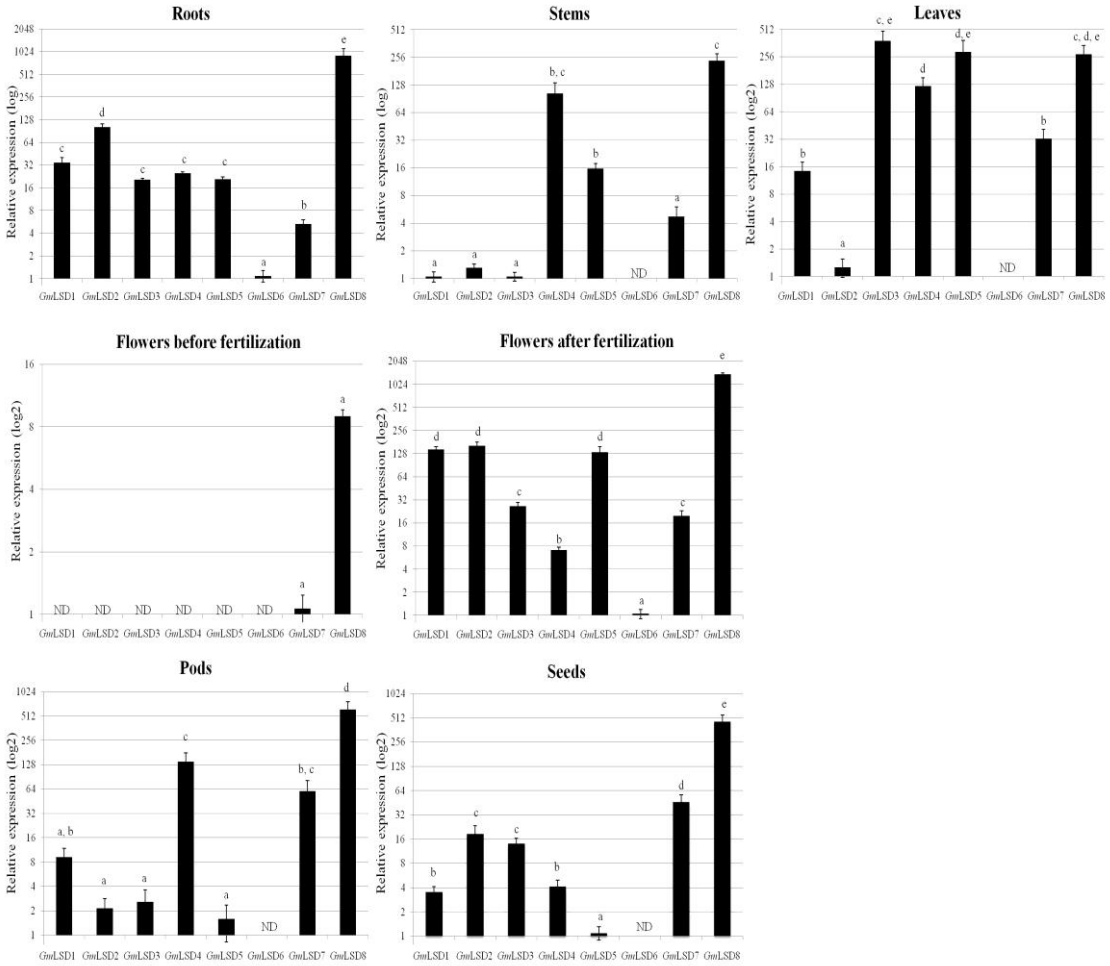


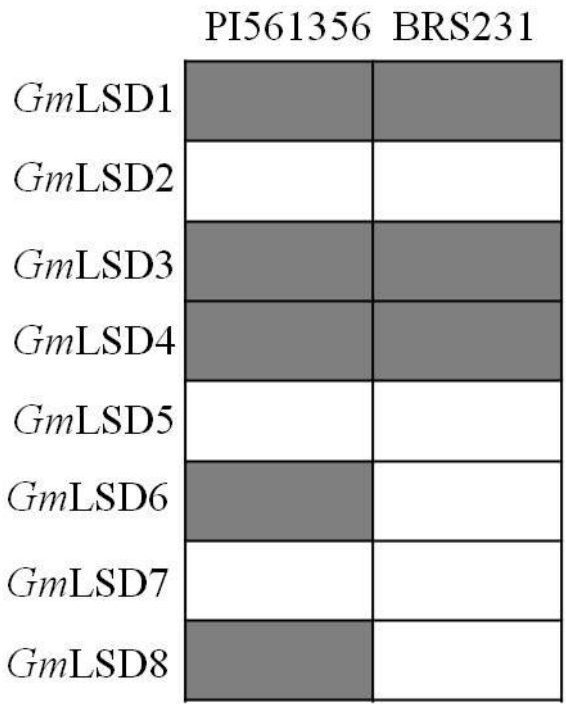
Fig. 8: *GmLSD* gene expression in the roots of BR16 (sensitive) and EMBRAPA48 (tolerant) cultivars in response to drought stress. The relative expression levels of the *GmLSD* genes were measured by RT-qPCR at T0 (control), T25, T50, T75, T100, T125 and T150 minutes of drought stress. The values are the means of three biological replicates with four technical replicates each. The means that are labeled identically (with a letter) in the same cultivar do not differ significantly (Bonferroni multiple comparison test, $p<0.05$). The *f-box*, *ACTII* and *ELF* reference genes were used as internal controls to normalize for the amount of mRNA present in each sample. The transcript levels from plants at T0 were used to normalize the transcript levels from plants subjected to drought stress. To allow a better comparison of groups of genes with large differences in their expression levels, some graphs of fold-change are shown in log2 scale, and this is indicated at the y axis. ND: not detected

Cis-element	Function	*GmLSD1*	*GmLSD2*	*GmLSD3*	*GmLSD4*	*GmLSD5*	*GmLSD6*	*GmLSD7*	*GmLSD8*
ABRE	Absicic acid response		2	1				1	
Box W1	Fungi elicitor response			2		1			
CGTCa	Methyl jasmonate response		1	2	2			6	1
ELI-box 3	Elicitor response								1
ERE	Ethylene response			3	1	1		1	1
GARE	Gibberellin response	2	5	2				2	
MBS	MYB binding site of drought response	1		1	1	1	2		
MBSII	MYB binding site of flavonoid biosynthetic			1			1		
TC-rich repeats	Defense and stress response	5	2	2	4	5	1	3	
TCA	Absicic acid response		2	2		2	4	1	
TGAC	Methyl jasmonate response		1	2	2			6	1
WUN	Wounding response				1				

Table 1: Number of copies and biological function of the putative *cis*-elements related to stress that were identified in the *GmLSD* genes. Analysis of a 2000 bp region upstream of the transcription start site (TSS) of each gene was performed using the Plant Care database.



ESM Figure S1: Relative expression levels of the *GmLSD* genes in each different organ of the Conquista Brazilian soybean cultivar. The values are the means of three biological replicates with four technical replicates each. The means that are labeled identically (with a letter) do not differ significantly (Tukey comparison test, $p<0.05$). *ACTII*, *CYP2* and *metalloprotease* reference genes were used as internal controls to normalize for the amount of mRNA present in each sample. In each organ, the gene with the lowest transcript level was used to normalize the transcript levels of the other genes. To allow a better comparison of groups of genes that exhibited large differences in their expression levels, some graphs of fold-change are shown in log2 scale, and this is indicated at the y axis. ND: no detected



ESM Figure S2: Expression levels of *GmLSD* genes measured by RNA-seq analysis of lesions microdissected from soybean leaves at ten days after infection by *P. pachyrhizi*. The susceptible variety BRS231 and the resistant PI561356 genotype were examined. Gray coloring indicates expression, and white coloring indicates an absence of expression.

Gene	5′- 3′ Primer sequence
*Gm*LSD1_F	CTTCCAAGTCCTAGACTAGTAAATGCTAC
*Gm*LSD1_R	TGTCTTATCCAAAATTTTCAAATGAC
*Gm*LSD2_F	TGATTACCAAAGCCTACTTTGTTG
*Gm*LSD2_R	AATGCTTGCACAAGATCATCA
*Gm*LSD3_F	CTCTTCTTTGGTGACTTGATTCG
*Gm*LSD3_R	TCTCATTTTTCTTTCTAATTTTCTTCTCC
*Gm*LSD4_F	CCAACTCTGGAACAGTATCTTCA
*Gm*LSD4_R	CCCTTTACACTTGTGTAACATTTTC
*Gm*LSD5_F	TGAGTGTTTTGAAAATTCCCATCT
*Gm*LSD5_R	CACAGAAAAGTAAATGAGTAGTTGAGATTTTAT
*Gm*LSD6_F	CTGTCAATCTAGCTTTGGAAGG
*Gm*LSD6_R	TTTAAAGAAAATCTAGTTAGTGAGGAATAAG
*Gm*LSD7_F	AGCCACAAGTGTGCAATGC
*Gm*LSD7_R	GCTCAGTGGTGCTTGTCGA
*Gm*LSD8_F	GATCAGGTTGGGCAAGTTAAG
*Gm*LSD8_R	GGCTTCCTTTGTTGTACAGACC

ESM Table S1: Summary of the *GmLSD* gene RT-qPCR primer sequences

Specie	Acronym	Locus	EST name
Aquilegia coerulea	Aco1	AcoGoldSmith_v1.015085m	not identified
	Aco2	AcoGoldSmith_v1.019125m	not identified
	Aco3	Aquca_026_00333.1	not identified
	Aco4	Aquca_044_00148.1	1143784
Arabidopsis thaliana	*At*LSD1** - Ath1*	Ath4G20380.1	Expression previously reported by [1]
	*At*LOL1** - Ath2*	Ath1G32540.1	Expression previously reported by [2]
	*At*LOL2** - Ath3*	Ath4G21610.1	4496932
Arabidopsis lyrata	Aly1	Alyrata492829	not identified
	Aly2	Alyrata492714	not identified
	Aly3	Alyrata473445	not identified
Brachypodim distachyon	Bdi1	Bradi1g13970.1	CCYP3855.b1
	Bdi2	Bradi3g14290.1	CCYO15286.g1
	Bdi3	Bradi3g16400.1	CCYP13632.g1
	Bdi4	Bradi4g02040.1	5547_H06_P11Z_034
	Bdi5	Bradi2g43370.1	not identified
Brassica rapa	Bra1	Bra038804	BR083678
	Bra2	Bra020924	4f07
	Bra3	Bra013460	BR118727
	Bra4	Bra020902	not identified
	Bra5	Bra010171	FY423929
Carica papaya	Cpa1	Cpapaya832.2	1613793_5_A15_064
	Cpa2	Cpapaya716.8	1448246_5_K20_070
Capsella rubella	Cru1	Carubv10005824m	not identified
	Cru2	Carubv10007536m	not identified

	Cru3	Carubv10010527m	not identified
Chamydomonas reinhardtii	Cre1	Creinhardtii12g517350t11.g	1031090B07.y2
Citrus clementine	Ccl1	Cclementine09_024795m.g	IC0AAA99DF06RM1
	Ccl2	Cclementine09_023739m	C31706G10EF
	Ccl3	Cclementine09025210m	C02009G12SK
Citrus sinensis	Csi1	Csinensis1g026941m	CGF1004433_H09
	Csi2	Csinensis1g032227m	CSAH-PNP1246E24
	Csi3	Csinensis1g033096m	CS00-C2-003-063-H08-CT.F
Cucumis sativus	Csa1	Csativus048600.1	not identified
	Csa2	Csativus313310.1	not identified
	Csa3	Csativus321740.1	not identified
	Csa4	Csativus126650.1	not identified
Eucaliptus grandi	Egr1	EgrandisE03314.1	GR-SE-001-GO-100-A12
	Egr2	EgrandisA00512.1	GR-SE-001-GO-020-E01
	Egr3	EgrandisD02186.1	GR-TS-001-AF-016-G06
Glycine Max	*GmLSD1*-Gma1*	Glyma08g13630.1	GLNBS20TF
	GmLSD2-Gma2*	Glyma05g30490.1	GLNAG22TF
	GmLSD3-Gma3*	Glyma15g22140.1	GLMCX33TF
	GmLSD4-Gma4*	Glyma09g10010.1	gmrtDrNS01_22-B_M13R_D05_041.s1
	GmLSD5-Gma5*	Glyma05g23710.1	GLLC683TF
	GmLSD6-Gma6*	Glyma17g16600.1	GLMBT61TF
	GmLSD7-Gma7*	Glyma01g40720.1	JGI_ABWZ2556.rev
	GmLSD8-Gma8*	Glyma07g31600.1	Gm_ck8528
Linum usitatisimum	Lus1	Lus10023392	LUSPS1AD_RP_105_L03_14AUG2008_006

	Lus2	Lus10040081	not identified
	Lus3	Lus10035361	LUSES4AD-T3-037_G04_15SEP2009_010
	Lus4	Lus10020766	LUSST4AD-T3-019_E02_15SEP2009_011
	Lus5	Lus10038409	not identified
	Lus6	Lus10008066	not identified
	Lus7	Lus10000782	LUSHE1NG-RP-090_D03_16FEB2007_025
	Lus8	Lus10030969	FLAX-RP-018_H07_13JULY2007_049
Lotus japonicus	Lja1	LJ6G010130	LJGCL48TF
	Lja2	LJ2G026290	BP049974
Malus domesticus	Mdo1	MD06G003140	Mdltb7520E23.g1
	Mdo2	MD12G018170	GO548841.1
Manihot esculenta	Mes1	Mesculenta018557m.g	not identified
	Mes2	Mesculenta018879m.g	CASLR90TF
	Mes3	Mesculenta017533m.g	CV01023A2C07.f1
	Mes4	Mesculenta020122m.g	DB927015
Medicago truncatula	Mtr1	Mtruncatula8g102780.1	457250
	Mtr2	Mtruncatula4g137880.1	MTYEF07TF
Mimulus guttatus	Mgu1	Mguttatus1a014848m	CCBX20509.b1
	Mgu2	Mguttatus1a014809m	CCBX15509.b1
	Mgu3	Mguttatus1a009107m	CCBX18974.b1
Oryza sativa	Osa1	Osativa03g43840.1	CT848680
	*Os*LSD1**- Osa2*	Osativa08g06280.1	Expression previously reported by [3]

	*Os*LOL2** - Osa3*	Osativa08g03610.1	Expression previously reported by [4]
	Osa4	Osativa12g41700.1	OSJNEe11L13.f
	Osa5	Osativa07g28910.1	CI009404.1
Phaseolus vulgaris	Pvu1	Phvulv091016735m	Pv038Dr_M13R_E03.ab1
	Pvu2	Phvulv091005034m	not identified
	Pvu3	Phvulv091014396m	Pv012A_M13R_H11.ab1
	Pvu4	Phvulv091002184m	PVUSE1NG-RP-023_M20_15MAY2009_068
Physcomitrella patens	Ppa1	Ppatens1154V6.1	gd09d09.y1
	Ppa2	Ppatens75166V6.1	CBAW3047.fwd
Populus trichocarpa	Ptr1	Ptrichoc0011s15810.1	UM63TG10
	Ptr2	Ptrichoc0001s43300.1	not identified
	Ptr3	Ptrichoc0001s02650.2	WS0125.BR_H05
	Ptr4	Ptrichoc0004s04310.1	WS02518.B21_C12
	Ptr5	Ptrichoc0003s08890.1	WS02514.B21.1_J03
Prunus pérsica	Ppe1	Ppersica013007m	not identified
	Ppe2	Ppersica013416m	PPN044B03-T7
	Ppe3	Ppersica012408m	not identified
	Ppe4	Ppersica012421m	PPN035D04-T7
Ricinus communis	Rco1	Rcommunis29668m000320	RCSB119TP
	Rco2	Rcommunis29827m002550	LINN_XX_012_J14_LINN12.CR_J1
	Rco3	Rcommunis29851m002411	JB2_39_G07
Selaginella moellinorffii	Smo1	Smoellindorffii|270395	not identified
Setarica itálica	Sit1	Sitalica|Si000307m.g	not identified

	Sit2	Sitalica|Si023407m.g	CFGZ7771.g1
	Sit3	Sitalica|Si014280m.g	not identified
	Sit4	Sitalica|Si014603m.g	CFBX19836.b1
	Sit5	Sitalica|Si037185m.g	CFGZ6426.b1
	Sit6	Sitalica|Si037902m.g	not identified
Sorgum bicolor	Sbi1	Sb08g021100.1	ANR1_14_E09.b1_A002
	Sbi2	Sb03g027750.1	not identified
	Sbi3	Sb07g004050.1	PI1_55_D07.g1_A002
	Sbi4	Sb01g013820.1	POL1_38_H09.g1_A002
	Sbi5	Sb07g002500.1	not identified
Thellungiella halophila	Tha1	Thhalv10008989m	BY832987
	Tha2	Thhalv10001047m	BY813876
	Tha3	Thhalv10026533m	Th729
	Tha4	Thhalv10026412m	not identified
Vitis Vinifera	Vvi1	GSVIVT01003209001	WIN1145.C21_I18
	Vvi2	GSVIVT01013190001	WIN0571.C21_G04
	Vvi3	GSVIVT01037958001	S8B00496
Volvox carteri	Vca1	Vcarteri|57759	CBGZ32500.fwd
Zea mays	Zma1	GRMZM2G055135_T03	ZM_BFb0183G18.r
	Zma2	GRMZM2G089106_T01	MEST128-G01.T3
	Zma3	GRMZM2G060057_T01	ZM_BFc0071L11.r
	Zma4	GRMZM2G173425_T01	ZM_BFc0001K22.r
	Zma5	GRMZM2G114613_T02	ZM_BFc0096N13.r 7

1. Dietrich RA, Richberg MH, Schmidt R, Dean C & Dangl 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

2. Epple P, Mack AA, Morris VR and 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 U S A 100:6831-6836

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

4. Xu C and He C (2007) The rice OsLOL2 gene encodes a zinc finger protein involved in rice growth and disease resistance. Mol Genet Genomics 278:85-94

ESM Table S2: Species, acronym, locus and expressed sequence tag (EST) identified for the 117 *LSD* genes

* Acronym used in the phylogenetic analysis

** Acronym previously used in other studies

CAPÍTULO III

DISCUSSÃO GERAL E PERSPECTIVAS

Os resultados obtidos nesse trabalho contribuem para o aprofundamento dos conhecimentos acerca da família gênica LSD (*Lesion Simulating Disease*). Embora haja muitos trabalhos em *Arabidopsis thaliana*, a identificação da família inteira ainda não havia sido realizada. Nossas análises demonstraram que a família é exclusiva de Viridiplantae. Foi identificado um total de 117 genes LSD, sendo que a expansão da família parece ter ocorrido no clado embriófitas, uma vez que, ao contrário das espécies basais, nas quais apenas um gene *LSD* foi identificado, as espécies pertencentes a esse clado apresentaram múltiplas copias.

Análises das sequências preditas de aminoácidos para os genes *LSD* permitiram a identificação de proteínas com um, dois e três domínios LSD. As proteínas identificadas nos organismos mais basais apresentaram três domínios LSD, indicando ser esta a condição ancestral durante a evolução da família. Esta hipótese foi corroborada pela análise filogenética, na qual a topologia da árvore mostrou que proteínas com um e dois domínios LSD divergiram posteriormente, a partir de um ancestral comum com três domínios. Além disso, as análises de ESTs (*Expressed sequence tag*) mostraram que genes codificantes de proteínas com um dois e três domínios são expressos, sugerindo uma pressão de seleção para a manutenção das três arquiteturas proteicas.

O envolvimento de alguns genes *LSD* em outros processos biológicos além da PCD condicionada por estresses já foi relatada. Pode-se citar a expressão de *BoLSD1* e *BoLSD2* em folhas de brócolis em processo de senescência (Coupe et al., 2004), o papel de *OsLSD1* na diferenciação de calos de arroz (Wang et al., 2005), bem como o envolvimento de *OsLOL2* (*LSD-like 2*) no crescimento de plantas de arroz (Xu e He, 2007) e *BohLOL1* (*LSD-like 1*) no crescimento de bambo (Yeh et al., 2011). Essas evidências, somadas aos nossos resultados (em que três diferentes arquiteturas de proteínas LSD foram identificadas) nos permitem propor que eventos de subfuncionalização e neofuncionalização podem ter ocorrido ao longo da evolução da família LSD. Tais eventos poderiam permitir a aquisição de novas funções para os genes *LSD* duplicados. Nesse cenário, trabalhos adicionais podem ser úteis para explorar o envolvimento de genes *LSD* em processos biológicos dependentes ou não de PCD.

Nossas análises mostraram que os três domínios LSD são altamente conservados e que o terceiro domínio foi identificado nas 117 sequências proteicas. A alta

conservação dos domínios LSD pode ser explicada pelo fato de que estes são responsáveis pela interação de proteínas LSD com outras proteínas envolvidas em PCD, como já foi mostrado para interação de *At*LSD1 com *At*MC1 (Coll *et al.*, 2010), com *At*bZIP10 (Kaminaka *et al.*, 2006) e com *At*GILP (He *et al.*, 2011b). Além disso, He et al. (2011) propuseram que o três domínios LSD da proteína LSD1 de *Pisum sativa* atuam como um sinal de localização nuclear (NLS), que é reconhecido pela importina α, resultando na importação nuclear da proteína *Ps*LSD1.

O presente trabalho teve como principal ênfase a caracterização dos genes *LSD* de soja (*GmLSD*). A análise da estrutura de cada membro *Gm*LSD mostrou que estes se dividem em três grupos, conforme o número de exon e introns. A análise *in silico* da região de 2kb anterior ao códon de iniciação da transcrição permitiu a identificação de inúmeros *cis*-elementos relacionados à resposta a estresses. Entretanto, análises adicionais são necessárias para comprovar a funcionalidade desses *cis*-elementos e para a identificação de *trans*-elementos envolvidos na regulação da transcrição dos genes *GmLSD*.

A análise do padrão de expressão dos genes *GmLSD* em diferentes órgãos da planta indicou que estes variam sua expressão dependendo do órgão. Estes resultados mostram que genes *GmLSD* têm comportamento diferente do que já foi relatado para *AtLSD1*, *AtLOL1* (Epple et al., 2003) e *OsLSD1* (Wang et al., 2005), os quais apresentam expressão constitutiva em diferentes tecidos. Esses resultados indicam que a família LSD pode ter um padrão de expressão diferente dependendo da espécie analisada. O gene *GmLSD8* (único com apenas dois domínios LSD) apresentou expressão ubíqua em todos órgãos analisados, exceto em caules, sugerindo um importante papel no desenvolvimento da planta. Considerando flores antes do processo de fertilização, apenas *GmLSD7* e *GmLSD8* foram expressos, enquanto que em flores fertilizadas, todos *GmLSD* foram expressos. Análises futuras podem esclarecer melhor estes resultados, visto que o envolvimento de genes *LSD* com processos de floração ainda não foi relatado.

A análise do padrão de expressão em plantas de soja inoculadas com *P. pachyrhizi* mostrou que a maioria dos genes *GmLSD* são responsivos ao estresse pelo fungo, indicando que estes podem estar envolvidos na sinalização da PCD condicionada pelo patógeno. Na análise do padrão de expressão por RT-qPCR, os genes *GmLSD1*, *GmLSD4* e *GmLSD8* tiveram sua expressão modulada apenas na variedade resistente à

infecção PI561356, em comparação com plantas controle. No experimento de RNA-seq foi verificada a expressão dos genes *GmLSD6* e *GmLSD8* apenas no genótipo resistente. A diferença de resultados obtidos nos dois experimentos pode ser atribuída a amostragens em momentos distintos do ciclo da doença. No experimento de RT-qPCR, os tempos analisados correspondem às seguintes etapas do processo de infecção, conforme van de Mortel et al. (2007) e Schneider et al. (2011) : a) resposta basal (até 12 horas pós inoculação – hpi), quando ocorre o primeiro pico de expressão gênica, com a formação do apressório e a penetração do fungo na epiderme celular; b) período quiescente (entre 12 e 24 hpi), quando continua o processo de infecção sem que ocorra marcada expressão gênica diferencial e c) colonização do fungo, formação de lesão e uredinia (após 72 hpi), quando ocorre um segundo pico de expressão gênica. As amostras para o experimento de RNA-seq foram coletadas 10 dias após a infecção. Este tempo corresponde ao período no qual é possível visualizar sintomas da doença, com a liberação dos primeiros uredósporos, finalizando o ciclo de infecção do patógeno (Koch *et al.*, 1983; Tremblay *et al.*, 2009). Diante desses resultados, os genes *GmLSD1*, *GmLSD4*, *GmLSD6* e *GmLSD8* devem ser alvo de estudos mais detalhados, pois podem estar relacionados à resposta de resistência observada na variedade resistente PI561356.

Embora trabalhos anteriores tenham mostrado o envolvimento dos genes *LSD* em morte celular condicionada por frio (Huang et al., 2010), hipóxia (Muhlenbock et al., 2007) e excesso de luz (Muhlenbock et al., 2008), o envolvimento destes genes na resposta ao déficit hídrico não havia sido relatada. Nossos dados indicam que a maioria dos genes *GmLSD* responderam ao estresse por seca e alguns foram modulados de forma diferencial (*GmLSD5* em folhas; *GmLSD1* e *GmLSD2* em raízes) apenas na cultivar tolerante EMBRAPA48, sugerindo que estes possam estar envolvidos na resposta de adaptação da planta ao déficit hídrico.

Análises futuras deverão ser realizadas para um melhor entendimento do papel que os genes *GmLSD* desempenham na HR desencadeada por patógeno ou por estresse abiótico. Sabendo que *AtLSD1* e *AtLOL1* atuam de forma antagonista no controle da PCD (Epple et al., 2003), o mesmo pode ocorrer entre membros desta família em soja, uma vez que esta espécie possui oito genes *LSD* e a manutenção de todas as cópias com a mesma função biológica pode não ser vantajosa. Além disso, os genes *GmLSD* podem interagir entre si, formando complexos a fim de controlar positiva ou

negativamente a PCD. Embora interações entre genes *LSD* ainda não tenham sido descritas, esta é uma hipótese interessante a ser explorada.

Estratégias de superexpressão e silenciamento podem ser utilizadas para desvendar o papel de genes *GmLSD* na sinalização de PCD. A obtenção de plantas transgênicas com os genes *GmLSD* superexpressos ou silenciados possibilitará a realização de ensaios de inoculação com patógeno ou químicos sinalizadores de dano celular, permitindo a análise fenotípica das plantas e das alterações em nível de expressão gênica.

Combinada às estratégias de superexpressão/silenciamento, a identificação de proteínas que interajam com os genes *GmLSD* é um aspecto importante a ser explorado. Estas análises certamente serão úteis para um melhor entendimento do papel desempenhado pela família LSD em inúmeros processos biológicos, especialmente no controle de PCD.

Um panorama completo do papel dos genes *LSD* em soja poderá torná-los úteis em programas de melhoramento, visando à obtenção de cultivares mais tolerantes e/ou resistentes a estresses bióticos e abióticos.

REFERÊNCIAS BIBLIOGRÁFICAS DOS CAPÍTULOS I E III

Agarwal, P., Reddy, M.P. e Chikara, J. (2010) WRKY: its structure, evolutionary relationship, DNA-binding selectivity, role in stress tolerance and development of plants. *Mol Biol Rep*. **38**, 3883-3896.

Aroca, R., Porcel, R. e Ruiz-Lozano, J.M. (2011) Regulation of root water uptake under abiotic stress conditions. *J Exp Bot*, **63**, 43-57.

Barros, R. e Borges, E.P. (2007) Doenças da cultura da soja *Soja 2006/2007*.

Barros, R. (2009) Doenças da Cultura da Soja. In *Tecnologia e Produção: Soja e Milho 2008/2009*, pp. 109-122.

Bromfield, K.R. e Melching, J.S. (1982) Sources of specific resistance to soybean rust. *Phytopathology*, **72**.

Chaerle, L., Boever, F., Van Montago, M. e Straeten, V.d. (2001) Thermografic visualization of cell death in Tobacco and Arabidopsis. *Plant Cell and Environment*, **24**, 15-25.

Cheng, Y.W. e Chan, K.L. (1968) The breeding of 'Tainung 3' soybean. *Journal of Taiwan Agricultural Research, Taiwan*, **17**, 30-35.

Coll, N.S., Vercammen, D., Smidler, A., Clover, C., Van Breusegem, F., Dangl, J.L. e Epple, P. (2010) *Arabidopsis* type I metacaspases control cell death. *Science*, **330**, 1393-1397.

Coll, N.S., Epple, P. e Dangl, J.L. (2011) Programmed cell death in the plant immune system. *Cell Death Differ*, **18**, 1247-1256.

Coupe, S.A., Watson, L.M., Ryan, D.J., Pinkney, T.T. e Eason, J.R. (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.

Dietrich, R.A., Delaney, T.P., Uknes, S.J., Ward, E.R., Ryals, J.A. e Dangl, J.L. (1994) *Arabidopsis* mutants simulating disease resistance response. *Cell*, **77**, 565-577.

Dita, M.A., Rispail, N., Prats, E., Rubiales, D. e Singh, K.B. (2006) Biotechnology approaches to overcome biotic and abiotic stress constraints in legumes. *Euphytica* **147**, 1-24.

Dubos, C. e Plomion, C. (2003) Identification of water-deficit responsive genes in maritime pine (*Pinus pinaster* Ait.) roots. *Plant Mol Biol*, **51**, 249-262.

Epple, P., Mack, A.A., Morris, V.R. and Dangl, J.L. (2003) Antagonistic control of oxidative stress-induced cell death in *Arabidopsis* by two related, plant-specific zinc finger proteins. *Proc Natl Acad Sci U S A*, **100**, 6831-6836.

Garcia, A., Calvo, E.S., de Souza Kiihl, R.A., Harada, A., Hiromoto, D.M. e Vieira, L.G. (2008) Molecular mapping of soybean rust (*Phakopsora pachyrhizi*) resistance genes: discovery of a novel locus and alleles. *Theor Appl Genet*, **117**, 545-553.

Gepts, P., Beavis, W.D., Brummer, E.C., Shoemaker, R.C., Stalker, H.T., Weeden, N.F. e Young, N.D. (2005) Legumes as a model plant family. Genomics for food and feed report of the Cross-Legume Advances Through Genomics Conference. *Plant Physiol*, **137**, 1228-1235.

Goellner, K., Loehrer, M., Langenbach, C., Conrath, U., Koch, E. e Schaffrath, U. (2010) *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust. *Mol Plant Pathol*, **11**, 169-177.

Gutterson, N. e Zhang, J.Z. (2004) Genomics applications to biotech traits: a revolution in progress? *Curr Opin Plant Biol*, **7**, 226-230.

Hartman, G.L., Wang, T.C. e Tschanz, A.T. (1991) Soybean rust development and the quantitative relationship between rust severity and soybean yield. *Plant Disease* **75**, 596–600.

Hartman, G.L., Hill, C.B., Twizeyimana, M., Miles, M.R. e Bandyopadhyay, R. (2011) Interaction of soybean and *Phakopsora pachyrhizi*, the cause of soybean rust. *CAB Rev Perspect Agric Vet Sci Nutr Nat Resour*, **6**.

Hartwig, E.E. (1986) Identification of a fourth major gene conferring resistance to soybean rust. *Crop Sci*, **26**, 1135-1136.

He, S., Huang, K., Zhang, X., Yu, X., Huang, P. e An, C. (2011a) The LSD1-Type zinc finger motifs of *Pisum sativa* LSD1 are a novel nuclear localization signal and interact with Importin Alpha. *PLoS One*, **6**, e22131.

He, S., Tan, G., Liu, Q., Huang, K., Ren, J., Zhang, X., Yu, X., Huang, P. e An, C. (2011b) The LSD1-interacting protein GILP is a LITAF domain protein that negatively regulates hypersensitive cell death in Arabidopsis. *PLoS One*, **6**, e18750.

Hidayat, O.O. e Somaatmadia, S. (1977) Screening of soybean breeding lines for resistance to soybean rust (*Phakopsora pachyrhizi* Sydow). *Soybean Rust Newsletter*, **1**, 9-22.

Huang, G.T., Ma, S.L., Bai, L.P., Zhang, L., Ma, H., Jia, P., Liu, J., Zhong, M. e Guo, Z.F. (2012) Signal transduction during cold, salt, and drought stresses in plants. *Mol Biol Rep*, **39**, 969-987.

Huang, X., Li, Y., Zhang, X., Zuo, J. e Yang, S. (2010) The *Arabidopsis* LSD1 gene plays an important role in the regulation of low temperature-dependent cell death. *New Phytol*, **187**, 301-312.

Hunt, M.D., Delaney, T.P., Dietrich, R.A., Weymann, K.B., Dangl, J.L. e Ryals, J.A. (1997) Salicylate-independent lesion formation in *Arabidopsis lsd* mutants. *Mol Plant Microbe Interact*, **10**, 531-536.

Jabs, T., Dietrich, R.A. e Dangl, J.L. (1996) Initiation of runaway cell death in an Arabidopsis mutant by extracellular superoxide. *Science*, **273**, 1853-1856.

Kaminaka, H., Nake, C., Epple, P., Dittgen, J., Schutze, K., Chaban, C., Holt, B.F., 3rd, Merkle, T., Schafer, E., Harter, K. e Dangl, J.L. (2006) bZIP10-LSD1 antagonism modulates basal defense and cell death in *Arabidopsis* following infection. *EMBO J*, **25**, 4400-4411.

Keiša, A., Kānberga, K., Gill, U., Kleinhofs, A. e Rostoks, N. (2008) Cloning and characterization of barley homologues of the *Arabidopsis* LSD1 gene: putative regulators of hypersensitive response *Acta Universitatis Latviensis*, **745**, 87-101.

Kliebenstein, D.J., Dietrich, R.A., Martin, A.C., Last, R.L. e Dangl, J.L. (1999) LSD1 regulates salicylic acid induction of copper zinc superoxide dismutase in *Arabidopsis thaliana*. *Mol Plant Microbe Interact*, **12**, 1022-1026.

Koch, E., Ebrahimnesbat, F. e Hoppe, H.H. (1983) Light and electron-microscopic studies on the development of soybean rust (*Phakopsora pachyrhizi* Syd) in susceptible soybean leaves. *Phytopathology*, **106**, 302-320.

Le, D.T., Nishiyama, R., Watanabe, Y., Mochida, K., Yamaguchi-Shinozaki, K., Shinozaki, K. e Tran, L.S. (2011a) Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress. *DNA Res*.

Le, D.T., Nishiyama, R., Watanabe, Y., Mochida, K., Yamaguchi-Shinozaki, K., Shinozaki, K. e Tran, L.S. (2011b) Genome-wide expression profiling of soybean two-component system genes in soybean root and shoot tissues under dehydration stress. *DNA Res*, **18**, 17-29.

Li, S., Smith, J.R., Ray, J.D. e Frederick, R.D. (2012) Identification of a new soybean rust resistance gene in PI 567102B. *Theor Appl Genet*, **125**, 133-142.

Li, X.P., Tian, A.G., Luo, G.Z., Gong, Z.Z., Zhang, J.S. e Chen, S.Y. (2005) Soybean DRE-binding transcription factors that are responsive to abiotic stresses. *Theor Appl Genet*, **110**, 1355-1362.

Lorrain, S., Vailleau, F., Balague, C. e Roby, D. (2003) Lesion mimic mutants: keys for deciphering cell death and defense pathways in plants? *Trends Plant Sci*, **8**, 263-271.

Mateo, A., Muhlenbock, P., Rusterucci, C., Chang, C.C., Miszalski, Z., Karpinska, B., Parker, J.E., Mullineaux, P.M. and Karpinski, S. (2004) LESION SIMULATING DISEASE 1 is required for acclimation to conditions that promote excess excitation energy. *Plant Physiol*, **136**, 2818-2830.

Miles, M.R., Frederick, R.D. e Hartman, G.L. (2006) Evaluation of soybean germplasm for resistance to *Phakopsora pachyrhizi*. *Plant Health Progress*.

Mittler, R., Vanderauwera, S., Gollery, M. e Van Breusegem, F. (2004) Reactive oxygen gene network of plants. *Trends Plant Sci*, **9**, 490-498.

Moeder, W. e Yoshioka, K. (2008) Lesion mimic mutants: A classical, yet still fundamental approach to study programmed cell death. *Plant Signal Behav*, **3**, 764-767.

Morel, J.B. and Dangl, J.L. (1999) Suppressors of the *arabidopsis lsd5* cell death mutation identify genes involved in regulating disease resistance responses. *Genetics*, **151**, 305-319.

Muhlenbock, P., Plaszczyca, M., Mellerowicz, E. e Karpinski, S. (2007) Lysigenous aerenchyma formation in *Arabidopsis* is controlled by LESION SIMULATING DISEASE1. *Plant Cell*, **19**, 3819-3830.

Muhlenbock, P., Szechynska-Hebda, M., Plaszczyca, M., Baudo, M., Mateo, A., Mullineaux, P.M., Parker, J.E., Karpinska, B. e Karpinski, S. (2008) Chloroplast signaling and LESION SIMULATING DISEASE1 regulate crosstalk between light acclimation and immunity in *Arabidopsis*. *Plant Cell*, **20**, 2339-2356.

Mur, L.A., Kenton, P., Lloyd, A.J., Ougham, H. e Prats, E. (2008) The hypersensitive response; the centenary is upon us but how much do we know? *J Exp Bot*, **59**, 501-520.

Panthee, D.R., Yuan, J.S., Wright, D.L., Marois, J.J., Mailhot, D. e Stewart, C.N., Jr. (2007) Gene expression analysis in soybean in response to the causal agent of Asian soybean rust (*Phakopsora pachyrhizi* Sydow) in an early growth stage. *Funct Integr Genomics*, **7**, 291-301.

Porcel, R. e Ruiz-Lozano, J.M. (2004) Arbuscular mycorrhizal influence on leaf water potential, solute accumulation, and oxidative stress in soybean plants subjected to drought stress. *J Exp Bot*, **55**, 1743-1750.

Reetz, E.R. (2012) Anuário brasileiro da soja 2012. *Editora Gazeta Santa Cruz, 2012.*, 156.

Rusterucci, C., Aviv, D.H., Holt, B.F., 3rd, Dangl, J.L. e Parker, J.E. (2001) The disease resistance signaling components EDS1 and PAD4 are essential regulators of the cell death pathway controlled by LSD1 in *Arabidopsis*. *Plant Cell*, **13**, 2211-2224.

Saibo, N.J., Lourenco, T. e Oliveira, M.M. (2009) Transcription factors and regulation of photosynthetic and related metabolism under environmental stresses. *Ann Bot*, **103**, 609-623.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X.C., Shinozaki, K., Nguyen, H.T., Wing, R.A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R.C. e Jackson, S.A. (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, **463**, 178-183.

Singh, B.B. e Thapliyal, P.N. (1977) Breeding for resistance to soybean rust in India. *FORD, R.E.; SINCLAIR, J.B. (Ed.). Rust of soybean: the problem and research needs. Urbana: University of Illinois*, **(INTSOY. Series, 12)**, 62-65.

Tollefson, J. (2010) Food: The global farm. *Nature*, **466**, 554-556.

Torres, M.A., Jones, J.D. e Dangl, J.L. (2005) Pathogen-induced, NADPH oxidase-derived reactive oxygen intermediates suppress spread of cell death in *Arabidopsis thaliana*. *Nat Genet*, **37**, 1130-1134.

Tran, L.S. e Mochida, K. (2010) Functional genomics of soybean for improvement of productivity in adverse conditions. *Funct Integr Genomics*, **10**, 447-462.

Tremblay, A., Li, S., Scheffler, B.E. e Matthews, B.F. (2009) Laser capture microdissection and expressed sequence tag analysis of uredinia formed by Phakopsora pachyrhizi, the causal agent of Asian soybean rust. *Physiological and Molecular Plant Pathology*, **73**, 163-174.

Twizeyimana, M., Ojiambo, P.S., Haudenshield, J.S., Caetano-Anolle´s, G., Pedley, K.F., Bandyopadhyay, R. e Hartman, G.L. (2011) Genetic structure and diversity of *Phakopsora pachyrhizi* isolates from soyabean. *Plant Pathology* **60**, 719-729.

Vidor, C., Fontoura, J.U.G., Macedo, J., Napoleão, B.A. e Min, T. (2004) Tecnologias de produção de soja região central do Brasil 2004: Doença e Medidas de Controle. .

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

Weymann, K., Hunt, M., Uknes, S., Neuenschwander, U., Lawton, K., Steiner, H.Y. e Ryals, J. (1995) Suppression and restoration of lesion formation in *Arabidopsis lsd* mutants. *Plant Cell*, **7**, 2013-2022.

Xu, C. e He, C. (2007) The rice OsLOL2 gene encodes a zinc finger protein involved in rice growth and disease resistance. *Mol Genet Genomics*, **278**, 85-94.

Yeh, S.H., Lin, C.S., Wu, F.H. e Wang, A.Y. (2011) Analysis of the expression of BohLOL1, which encodes an LSD1-like zinc finger protein in *Bambusa oldhamii*. *Planta*, **234**, 1179-1189.

Zhu, L., Zhang, Y., Zhang, W., Yang, S., Chen, J.Q. e Tian, D. (2009) Patterns of exon-intron architecture variation of genes in eukaryotic genomes. *BMC Genomics*, **10**, 47.

www.biodiesel.gov.br (acesso em 31/07/2011).

www.conab.gov.br (acesso em 31/07/2011).