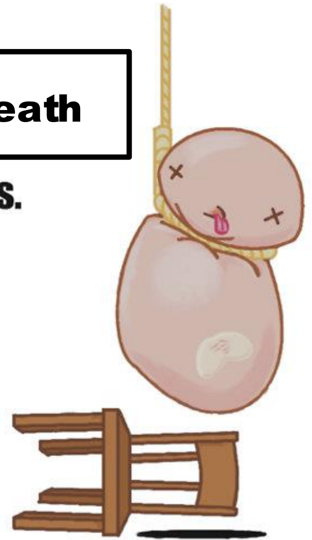
Programmed cell death

Know the signs.



IDENTIFICAÇÃO E CARACTERIZAÇÃO DE FAMÍLIAS GÊNICAS RELACIONADAS À MORTE CELULAR PROGRAMADA EM PLANTAS.

CAROLINE CABREIRA CAGLIARI

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

IDENTIFICAÇÃO E CARACTERIZAÇÃO DE FAMÍLIAS GÊNICAS RELACIONADAS À MORTE CELULAR PROGRAMADA EM PLANTAS.

Caroline Cabreira Cagliari

Porto Alegre - RS
2017

Caroline Cabreira Cagliari

Identificação e caracterização de famílias gênicas relacionadas à morte celular programada em plantas.

Tese 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 Doutor em Ciências - Genética e Biologia Molecular

Orientadora: Dra. Maria Helena Bodanese Zanettini
Co-orientadores: Dr. Alexandro Cagliari e
Dra. Márcia Pinheiro Margis

PORTO ALEGRE-RS
2017

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 bolsa CNPq, apoio financeiro do Projeto Universal CNPq - Chamada 14/2012 e projeto BIOTECSUR II-MCTI.

Dedico essa tese a minha fonte inesgotável de amor e inspiração, minha mãe Judite, meu pai João Pedro (*in memoriam*) e meu companheiro de caminhada, meu marido Alexandro.

“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

É chegado o fim de um longo ciclo da minha formação profissional. Ao longo desses anos de caminhada, muitas experiências foram acrescidas ao meu caminho. Considero que a melhor lição nesse período foi perceber que um trabalho não é feito sozinho e não depende apenas de aplicar os conhecimentos adquiridos. Depende também da solidificação da aprendizagem com base em um propósito maior que vai além da ciência, mas que se mistura com os valores pessoais, se molda na troca de ideias e no objetivo comum.

Muitas pessoas fizeram parte do aporte pessoal necessário para que essa tese fosse possível. Agradeço imensamente a professora Dra. Maria Helena, responsável por me abrir as portas do Departamento de Genética, ainda na Iniciação Cientifica. Registro aqui minha enorme gratidão pelos anos de convivência e aprendizagem. A professora Dra. Márcia, agradeço a disponibilidade constante, as conversas nem sempre acadêmicas e o aceite de embarcar juntamente comigo nesse desafio. Ao professor Dr. Alexandro, agradeço o incentivo, que culminou com a criação de um grupo de pesquisa voltado para o tema dessa tese. Agradeço à confiança em mim depositada, me possibilitando atuar na orientação de alunos de Iniciação Científica, o que me permitiu descobrir o quanto a troca de conhecimento é gratificante.

Aos professores do PPGBM, responsáveis pela transmissão do conhecimento e solidificação do perfil profissional de cada aluno, meu sincero muito obrigado. À professora Dra. Loreta e seu grupo de pesquisa, agradeço de coração a disponibilidade e todas as sugestões, desde meu ingresso no Doutorado. Ao Elmo, incansável na arte de auxiliar e viabilizar qualquer questão oriunda de nós alunos, deixo aqui registrada minha enorme gratidão. À banca, Dra. Maria Lúcia, Dra. Franceli e Dra. Andréia, por terem aceitado contribuir com esse trabalho e participar da conclusão desse ciclo.

Ao NBB, núcleo de pesquisas e fonte de curiosidade residente na UERGS, campus de Santa Cruz do Sul. À Nathália, Bianca e David, agradeço a parceria e a permissão para que eu pudesse transmitir meu conhecimento e assim colocar em prática um dos objetivos traçados para minha vida profissional. Vocês vão longe, certamente!

Aos colegas e amigos conquistados nos anos de laboratório, agradeço a companhia nas horas do chimarrão, nas tardes de pipoca e durante as trocas de ideias nas mais diversas discussões. Ao Lauro (hoje professor Dr. Lauro Bücker), grande incentivador e parceiro desde o início da linha de pesquisa dessa tese. Aos amigos também desbravadores do mundo da filogenética, que me permitiram parcerias da onde surgiram publicações e discussões das mais diversas, Rauber (hoje professor Dr. Rafael Rauber), Douglas (Dr. Douglas Messeder) e a Leila ("pilateira" e quase Dra).

Aos amigos que ficaram na Terra dos Marechais, agradeço as doses de renovação que me foram injetadas a cada visita e a cada reencontro, as quais contribuíram imensamente para a conclusão desse ciclo.

À Deus, que me permitiu caminhar e superar os obstáculos que me foram impostos nesse trajeto. À minha incansável fonte de renovação, minha família, que me permitiu levantar quando os

tombos passaram a fazer parte desse caminho. À minha mãe Judite e seu coração infinitamente bondoso, palavras nunca serão suficientes para definir toda a minha admiração e o meu amor. Obrigado por me permitir enxergar o mundo com positividade e fé em todas as situações e por sempre me amparar com amor e carinho imensurável. Ao meu amado pai, que anda a camperear pelos céus, e aonde estiver, serviu de suporte e orientação nos dias difíceis. Aos meus sobrinhos Gui, Leo, Dudu e Matheus, meus irmãos Jeferson e Adailton e minhas cunhadas Mirian e Silvani, por me serem tão fundamentais sempre, mesmo que à distância. À minha sogra Iraci, pelas palavras de incentivo, pela torcida constante e por compreender a distância muitas vezes inevitável.

Ao meu marido Alexandro, minha gratidão sem tamanho por ser meu amigo e companheiro, pelo incentivo diário, por me impulsionar a recriar meus sonhos e ver neles a razão para tudo isso. Agradeço especialmente por me incentivar a explorar o mundo da didática em sala de aula nos mais diversos níveis, que me trouxe uma fonte imensa de alegria. As nossas guriazinhas Belinha (nossa filhota de asas) e Lucy (nossa filhota canina), alegrias da nossa família. A nossa mais recente fonte de inspiração, nosso serzinho de luz, que com apenas 13 semanas de vida intrauterina, já provoca um mundo de sentimentos e amor nunca antes vividos.

Aos colegas e amigos de trabalho do Colégio Murialdo de Porto Alegre, com quem compartilho o mundo encantador da docência para adolescentes e crianças.

Por fim, agradeço a todos que de uma forma ou de outra ajudaram a construir essa tese e minha trajetória acadêmica e pessoal. Fica aqui registrado meus sinceros agradecimentos.

SUMÁRIO

Resumo 11

Abstract 12

1. Introdução geral 13

1.1 Culturas agrícolas..14

1.2 Ameaças às culturas agrícolas: estresses bióticos e abióticos..............................15

1.3 As respostas a condições de estresse e a morte celular programada....................16

1.4 Explorando a função gênica por meio da genômica...20

2. Objetivo geral...27

Capítulo I – *The phylogeny and evolutionary history of the Lesion Simulating Disease (LSD) gene family in Viridiplantae*……..……………………………28

Capítulo II – *Caspases in plants: metacaspase gene family in plant stress responses*..42

Capítulo III – *Soybean (Glycine max) NF-Y (Nuclear Factor of Y Box) gene family and its potential role under stress conditions and nodulation*………………………...54

Capítulo IV – *Revising the PLAC8 Gene Family: From a Central Role in Differentiation, Proliferation, and Apoptosis in Mammals to a Multifunctional Role in Plant*...………………..……..82

Capítulo V – *GILP Family: A Stress-Responsive Group of Plant Proteins Containing a LITAF Motif*………………………………..113

Discussão geral e perspectivas..139

Bibliografia geral...145

LISTA DE ABREVIATURAS

At: *Arabidopsis thaliana*

*At*bZIP10: do inglês *basic region leucine zipper 10*

*At*MC1: *Arabidopsis thaliana* metacaspase do tipo I

CNA: Confederação da Agricultura e Pecuária do Brasil

CONAB (Companhia Nacional de Abastecimento)

EEE: do inglês *excess excitation energy*

FAO: do inglês *Food and Agriculture organization of the United Nations*

GILP: *GSH-induced LITAF domain protein*

Gm: *Glycine max*

HR: do inglês *hypersensitive response*

LSD: L*esion Simulating Disease 1*

MATOPIBA: região produtiva nos estados do Maranhão, Tocantins, Piauí e Bahia

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

NF-Y: *Nuclear Factor Y*

OMC: Organização Mundial do Comércio

Os: *Oryza sativa*

PCD: do inglês *programmed cell death*

PLAC8: *Placenta* Specific 8

PR: do inglês *pathogenesis-related* ou genes PR

ROS: do inglês *reactive oxygen species*

RESUMO

Previsões indicam que o Brasil deverá assumir a liderança na exportação de produtos agrícolas a partir de 2024. Em 2017, estima-se que a agricultura brasileira poderá ter a maior safra da história. Nesse contexto, investimentos em pesquisas científicas em plantas modelo e em espécies relevantes economicamente são de vital importância a fim de permitir a manutenção do crescimento do setor agrícola em nosso país. Embora o cenário seja positivo e bastante promissor, diversos estresses ambientais afetam a agricultura, ocasionando perdas severas em produtividade e rendimento. Estresses bióticos e abióticos, como seca, alagamento, doenças e patógenos são grandes contribuintes para a diminuição do potencial genético de desenvolvimento e reprodução das culturas agrícolas. Os mecanismos moleculares envolvidos na tolerância/resistência a estresses têm sido intensamente estudados, com grande ênfase nos mecanismos de resposta inerentes aos estresses individuais. Em plantas, um dos processos desencadeados em resposta a estresses é conhecido como Morte Celular Programada (*Programmed cell death*-PCD). A ocorrência de PCD em plantas é anteriormente marcada pela resposta de hipersensibilidade (*hypersensitive response* – HR), uma forma de PCD seguida por autólise. Diversas famílias gênicas são sabidamente envolvidas com o processo de PCD em plantas e foram recentemente propostas como integrantes de uma rede de controle de PCD, conhecida como morteossomo (*deathsome*). Entre essas famílias envolvidas com PCD, muitas são apenas conhecidas superficialmente ou apresentam alguns poucos genes já caracterizados, como as família LSD (*Lesion Simulating Disease),* Metacaspase, NF-Y (*Nuclear Factor Y*), PLAC8 (*Placenta* Specific 8) e GILP (*GSH-induced LITAF domain protein*). Devido à importância dos mecanismos desencadeados em estresses, investigações detalhadas dos genes atuantes em HR e PCD podem contribuir expressivamente para a completa compreensão das rotas de resposta induzidas durante esses processos. A presente tese enquadrou-se no contexto atual de valer-se de diferentes ferramentas de bioinformática, tendo como objetivo final a caracterização *in silico* de cinco famílias gênicas relacionadas com PCD em plantas: LSD, Metacaspase, NF-Y, PLAC8 e GILP. A identificação de todos os genes pertencentes a essas famílias, bem como a descrição detalhada da estrutura gênica e proteica permitiu a determinação de muitos aspectos próprios de cada família, os quais podem explicar seu envolvimento em PCD. Somados a esses dados, aspectos evolutivos de cada família gênica também foram investigados. Nossos resultados servem como base para trabalhos aprofundados considerando genes específicos, além de representar o ponto inicial para a compreensão da função dessas famílias em plantas, especialmente no mecanismo de PCD.

ABSTRACT

Predictions indicate that Brazil should reach the leadership in the exportation of agricultural products until 2024. In 2017, it is estimated that Brazilian agriculture may have the highest harvest in the history. In order to maintain the growth of the agricultural sector in our country, investments in scientific research in model plants and in economically relevant species are very important. Although the scenario is very positive and promising, numerous environmental stresses affect agriculture, causing severe losses in productivity and yield. Biotic and abiotic stresses, such as drought, flooding, diseases and pathogens are major contributors to a decrease in the genetic potential for development and reproduction of agricultural crops. The molecular mechanisms involved in tolerance/resistance to biotic and abiotic stresses have been intensively studied, with focus in the response mechanisms inherent to individual stresses. In plants, one of the processes triggered in response to stresses is known as Programmed Cell Death (PCD). The occurrence of PCD in plants is marked by the previously hypersensitivity response (HR), a form of PCD followed by autolysis. Several gene families are known to be involved with the PCD process in plants and have been recently described as members of a PCD network, known as deathsome. The majority families involved with PCD are poor understood. Some of them present few genes already characterized, such as the Metacaspase, NF-Y (Nuclear Factor Y), PLAC8 (Placenta Specific 8) and GILP (*GSH-induced LITAF domain protein)* family. Due to the importance of the mechanisms triggered in stresses, detailed investigations considering the genes involved in HR and PCD can contribute for a complete understanding of the responses induced during these processes. The present thesis is based on bioinformatics tools, aiming the *in silico* characterization of five gene families related to PCD in plants: LSD, Metacaspase, NF-Y, PLAC8 and GILP. The identification of all genes belonging to these families, as well as a detailed description of gene and protein structure allowed a determination of several aspects own of each family, which can explain their involvement in the PCD. In addition, evolutionary aspects from each family were also investigated. Our results will serve as a basis for future deep studies considering specific genes and represents the starting point for understand the role of these gene families in plants, especially in PCD mechanism.

INTRODUÇÃO GERAL

1. INTRODUÇÃO GERAL

1.1 CULTURAS AGRÍCOLAS

A Organização Mundial do Comércio (OMC) recentemente destacou o Brasil como terceiro maior exportador mundial de grãos, superado apenas pela Comunidade Europeia e Estados Unidos. A Organização das Nações Unidas para Agricultura e Alimentação (FAO), prevê que o Brasil deverá assumir a liderança na exportação de produtos agrícolas a partir de 2024, quando a área plantada será de 69,4 milhões de hectares (http://sna.agr.br/brasil-e-o-terceiro-maior-exportador-mundial-de-produtos-agropecuarios/) (acesso abril/2017).

O avanço da produção agrícola sobre áreas de pecuária degradada, bem como o uso de áreas de abertura (de primeiro plantio) e principalmente o aumento do plantio da segunda safra no Centro-Oeste e no MATOPIBA (região produtiva nos estados do Maranhão, Tocantins, Piauí e Bahia) vêm impulsionando a ampliação da área agrícola, segundo a Confederação da Agricultura e Pecuária do Brasil (CNA), (http://sna.agr.br/brasil-e-o-terceiro-maior-exportador-mundial-de-produtos-agropecuarios/) (acesso abril/2017). Estima-se que em 2017 o Brasil poderá ter a maior safra da história. A estimativa da produção de grãos para a safra 2016/17 é de 215,3 milhões de toneladas. O crescimento em relação à safra anterior deverá ser de 15,3%, ou seja, 28,6 milhões de toneladas. A área total plantada está estimada em 59,7 milhões de hectares, o que representa um crescimento de aproximadamente 2,3%, quando comparada com a safra anterior (http://www.conab.gov.br/OlalaCMS/uploads/arquivos/17_01_11_11_30_39_boletim_graos_janeiro_2017.pdf).

O expressivo aumento da área plantada em nosso país e o consequente impulso que isso representa para nossa economia evidencia a importância de investimentos em pesquisas científicas que contemplem práticas agrícolas superiores, bem como a melhoria das cultivares vegetais utilizadas no plantio. No que se refere às pesquisas, a espécie *Arabidopsis thaliana* é uma das mais utilizadas há mais de 40 anos como modelo entre as dicotiledôneas. Apesar de sua proximidade com outras espécies como nabo, repolho, couve, brócolis e colza, ela não possui importância econômica direta, ficando sua importância restrita a pesquisas nas áreas de genética, bioquímica e fisiologia (Delatorre and Silva,

2008). Além de *Arabidopsis*, outra planta modelo entre as monocotiledôneas é o arroz. Isso se dá devido ao pequeno tamanho de seu genoma - 430 Mbp - (Arumuganathan and Earle, 1991) e a grande facilidade na manipulação genética e molecular (Gale; Devos, 1998). A soja [*Glycine max* (L.) Merrill], além de *Lotus japonicus* e *Medicago truncatula,* também foram recomendadas pela comunidade científica internacional como plantas modelo para estudos genéticos e moleculares (Gepts *et al.*, 2005).

Segundo dados da CONAB (Companhia Nacional de Abastecimento) (www.conab.gov.br), a produção brasileira de soja para a safra 2016/2017 tem projeção de crescimento de 8,7%, atingindo 103,8 milhões de toneladas. Esse crescimento é previsto em função da maior liquidez, rentabilidade e menor custo para a cultura. Atualmente a soja é a principal cultura produzida no Brasil, seguida pelo milho e pelo arroz. Em nível mundial, o Brasil é o segundo maior produtor de soja, atrás apenas dos EUA. A produção de soja representa uma importante alavanca na economia de nosso país, o que destaca fortemente a relevância de estudos envolvendo essa cultura agrícola.

1.2 AMEAÇAS ÀS CULTURAS AGRÍCOLAS: ESTRESSES BIÓTICOS E ABIÓTICOS

Embora o cenário agrícola seja bastante positivo, diferentes estresses ambientais afetam a agricultura, ocasionando perdas severas em produtividade e rendimento. A maioria das plantas cultivadas cresce em ambientes abaixo do adequado devido às alterações climáticas constantes, o que limita o potencial genético de desenvolvimento e reprodução das culturas agrícolas (Bray *et al.*,2000; Rockstrom and Falkenmark, 2000). Tal fato é enfatizado na análise da diferença entre o rendimento máximo e o rendimento médio das culturas (Atkinso and Urwin, 2012). Essa diferença pode ser explicada por condições ambientais não favoráveis, conhecidas como estresses, que levam a alterações fisiológicas potencialmente danosas (Shao *et al.*, 2008). De modo geral, estima-se que perdas em torno de 50% em produtividade são ocasionadas por estresses abióticos (Ahmad *et al.*, 2016), tais como seca, alagamento, químicos orgânicos e inorgânicos, calor/frio, metais pesados e salinidade. Além dos estresses abióticos, estresses bióticos como doenças ocasionadas por pestes e patógenos contribuem negativamente para a diminuição da produtividade. As perdas ocasionadas por estresses abióticos são geralmente maiores do que as perdas decorrentes de estresses bióticos (Ashraf *et al.*, 2012) (Figura 1). Isso se dá devido ao fato de que alterações climáticas (estresses abióticos) podem atingir uma área inteira de uma lavoura, o que não necessariamente acontece com estresses bióticos, que geralmente ficam retidos em uma área/espaço.

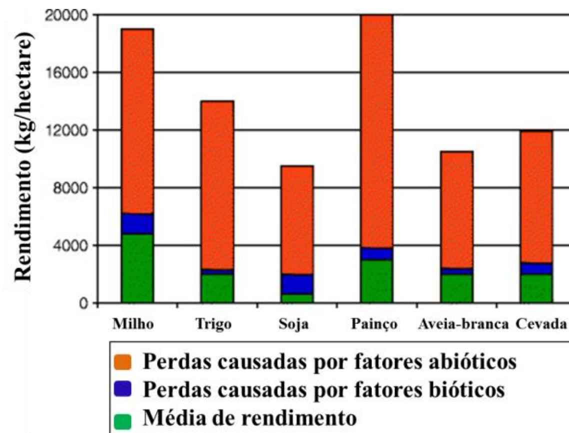


Figura 1: Perdas mundiais ocasionadas em diversas culturas agrícolas, como resultado da ocorrência de estresses. Adaptado de Ashraf *et al.* (2012).

Além dos danos causados por condições de estresse isoladamente, sabe-se que diferentes danos podem ocorrer simultaneamente na planta. Plantas sob condições de estresse abiótico, por exemplo, tornam-se mais suscetíveis aos danos causados por ervas daninhas, insetos e doenças, aumentando consideravelmente as perdas (Dita *et al.*, 2006). Nesse sentido, os mecanismos moleculares envolvidos na tolerância/resistência a estresses bióticos e abióticos têm sido intensamente estudados, com grande ênfase nos mecanismos de resposta inerentes aos estresses individuais (Mengiste *et al.*, 2003; Suzuki *et al.*, 2005; Abuqamar *et al.*, 2009; Atkinso and Urwin, 2012). Estudos submetendo plantas a um estresse ou vários simultaneamente têm destacado conjuntos de genes que são modulados em ambas as condições. Tais genes têm sido propostos como genes de resposta geral a estresses ou representantes de pontos de sobreposição entre vias de sinalização (Atkinson and Urwin, 2012; Atkinson *et al.*, 2013; Bostock *et al.*, 2014; Prasch and Sonnewald, 2014; Rasmussen *et al.*, 2013; Rivero *et al.*, 2013; Suzuki *et al.*, 2014).

1.3 As respostas a condições de estresse e a morte celular programada

O desenvolvimento e a manutenção dos organismos multicelulares dependem de um balanço constante entre a proliferação das células e a morte das mesmas, o que garante a homeostasia celular. Esse programa de deleção de células é um fenômeno biológico comum e essencial desde as fases

iniciais até a fase adulta. Em organismos multicelulares, a indução de processos de morte celular pode ocorrer como resposta a estímulos intra ou extracelulares. De acordo com sua fisiopatologia e características morfológicas, os processos de morte celular podem ser distinguidos em apoptose, autofagia, necrose e senescência (Su *et al.*, 2015). Em plantas, todos esses processos também são identificados (Reape, 2008; Minina *et al.*, 2013; Michaeli *et al.*, 2016), exceto apoptose, que é aparentemente exclusiva de animais (van Doorn, 2011). No entanto, um outro tipo de apoptose relacionada com morte celular programada (*apoptotic-like programmed cell death* – PCD) é relatada em plantas e comumente chamada apenas de PCD (Tabela 1) (van Doorn, 2011). Esta ocorre em uma série de processos que vão desde o desenvolvimento até a senescência (Del Duca *et al.*, 2014). Assim como em animais, a ocorrência de PCD em plantas está intimamente relacionada às respostas de invasões de patógenos e condições de estresse (Reape *et al.*, 2008).

Dois tipos de PCD podem ser identificados em plantas, a que não é seguida por autólise e a seguida por autólise, as quais apresentam características próprias (Figura 2) (van Doorn, 2011). A ocorrência de PCD em plantas como resposta a invasões de patógenos é anteriormente marcada pela resposta de hipersensibilidade (*hypersensitive response* – HR), uma forma de PCD seguida por autólise. A HR é acompanhada pela geração de ROS (*reactive oxygen species*) no apoplasto, cloroplastos e mitocôndrias, o aumento do nível de cálcio celular, encolhimento citoplasmático, condensação da cromatina, vacuolização e rompimento do cloroplasto (Coll *et al.*, 2011). Sua importância esta relacionada à percepção da planta ao ataque do patógeno e consequente comprometimento do crescimento de agentes patogênicos (Rusterucci *et al.*, 2001; Mur *et al.*, 2008).

Embora usualmente a HR seja reportada como um importante marcador de interações bióticas, muitas de suas características, como a morte celular localizada e a indução de genes associados com defesa (*pathogenesis-related* ou genes PR) são compartilhadas com respostas a estresses abióticos, como excesso de excitação energética (*excess excitation energy* – EEE) (Zurbriggen *et al.*, 2010). Devido à sobreposição das características de resposta encontradas em condições de estresse biótico e abiótico, investigações detalhadas dos mecanismos de HR e PCD podem contribuir expressivamente para a completa compreensão dessas rotas de resposta, bem como sobre o conhecimento dos genes atuantes em ambos os processos.

Organismo	Processo	Critérios de definição	Outros critérios
Animais	1)Apoptose	Presença de corpos apoptóticos ou bolhas na superfície da célula. Degradação em outras células depois de fagocitose.	Condensação da cromatina, fragmentação nuclear. Sinais de "*find-me*" e "*eat-me*".
	2) Autofagia	Aumento no número de autofagosomos, autolisosomos e pequenos vacúolos liticos.	
	3) Necrose	Nenhum dos critérios descritos acima.	Dilatação da célula e das organelas. Ruptura da membrana plasmática.
Plantas	1) Seguida por autólise	Rápida liberação do citoplasma depois da ruptura do tonoplasto.	Condensação da cromatina. Aumento do volume vacuolar e diminuição do volume citoplasmático.
	2) Não-seguida por autólise	Ausência de rápida liberação do citoplasma.	Dilatação das organelas. Ausência de aumento do volume vacuolar.

Tabela 1: Diferentes tipos de PCD em animais e plantas e suas características. Adaptado de van Doorn *et al.* (2011).

Diversas famílias gênicas são sabidamente envolvidas com o processo de PCD em plantas. Coll *et al.* (2011), via *screening* de duplo híbrido em levedura, propuseram vários genes, que atuariam em conjunto regulando PCD em *A. thaliana* e que constituiriam o chamado morteossomo (*deathosome*) (Figura 2). Entre os genes propostos como atuantes nessa rede, o gene *LSD1* (*Lesion Simulating Disease 1)* ocupa papel central no controle da PCD induzida por HR (Dietrich *et al.*, 1994; Cabreira *et al.*, 2013) (Figura 2). Alguns mutantes *lsd* 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.* (1997) demonstraram o envolvimento do gene *AtLSD1* na regulação negativa da HR, via um sinal dependente de superóxido. O papel de *AtLSD1* na regulação negativa de PCD reside em controlar a extensão da morte celular, para que essa não se estenda além do limite adequado da HR, o que causaria a destruição total do órgão da planta (Epple *et al.*, 2003; Moeder and 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 de sequências pertencentes às famílias gênicas já conhecidas (Figura 2, genes em preto), oito proteínas de função desconhecida foram descritas como atuantes no morteossomo (Figura 2, genes em cinza). Entre as proteínas com domínio já conhecido esta a proteína metacaspase do tipo I (*At*MC1), que atua regulando positivamente PCD. Coll *et al.* (2010) demonstraram que a interação de *At*LSD1 com *At*MC1 retém esta última inativa no citoplasma, impedindo a propagação de PCD nas células vizinhas à infecção (Figura 3). A interação de *At*LSD1 com o regulador positivo de PCD *At*bZIP10 (*basic region leucine zipper 10*), outra proteína com domínio conhecido, ocorre também via retenção desta no citoplasma, impedindo que a mesmo se desloque para o núcleo da célula e ative PCD

(Kaminaka *et al.*, 2006). Apenas as interações descritas acima foram demonstradas detalhadamente. No que se refere às outras interações presentes no morteossomo, embora estas tenham sido demonstradas como resultados de ensaios de duplo-híbrido em levedura, estudos mais detalhados ainda merecem ser feitos. Além do morteossomo, há um registro na literatura da interação de *At*LSD1 com outro regulador negativo de PCD chamado de *At*GILP (*GSH-induced LITAF domain protein*) (He *et al.*, 2011). Considerando esses resultados e o envolvimento dessas sequências na rede de regulação de PCD, é de especial interesse a investigação desses genes, o que poderá contribuir para o entendimento do mecanismo de resposta a estresses e revelar potenciais genes-alvos para abordagens biotecnológicas.

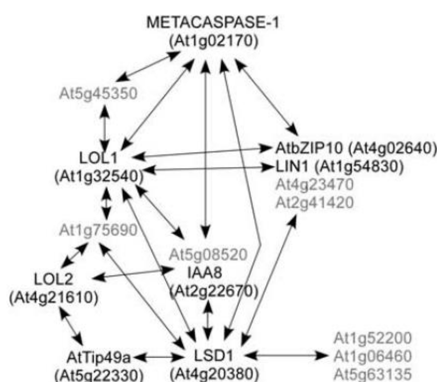


Figura 2: Morteossomo em *A. thaliana*, representado a interação comprovada (via duplo híbrido de levedura) de possíveis reguladores de PCD com *At*LSD1. As proteínas escritas em cinza ainda não tem função conhecida (Coll *et al.*, 2011).

Além das proteínas com alguma função já descrita presentes no morteossomo, estudos referentes à caracterização das famílias gênicas atuantes no processo ainda não foram conduzidos. No que se refere à família LSD, a investigação do perfil de expressão dos genes presentes em soja (*GmLSD1* a *GmLSD8*) foi realizada durante minha dissertação de mestrado (Cabreira et. al, 2013). Ademais, nenhum outro estudo foi realizado visando à caracterização completa da família. Assim como para a família LSD, a ampla caracterização da família gênica metacaspase ainda não foi realizada. Devido ao papel no importante mecanismo de PCD, é de especial interesse a completa caracterização dos genes descritos no morteossomo, bem como das famílias as quais esses genes pertencem. A identificação dos genes parálogos/ortólogos as sequências propostas no morteossomo (incluindo as famílias LSD e metacaspase) poderá contribuir para o entendimento da evolução dessas

famílias, bem como permitir a elucidação do envolvimento destes genes na resposta da planta a estresses bióticos e abióticos.

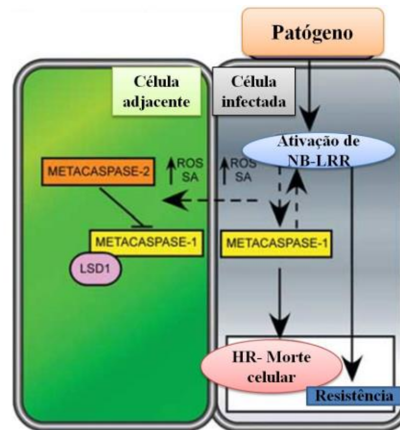


Figura 3: Modelo de atuação de *At*LSD1 e *At*MC1. Após a percepção do ataque de patógenos via NB-LRR na célula inicial da infecção, moléculas sinalizadoras de dano são liberadas como ROS e SA. *At*MC1 atua ativando HR e morte celular, que em ultima análise, leva à resistência. Esses sinais são propagados 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 AtMC1, por sua vez regulando negativamente PCD, tal qual *At*LSD1. Modificado de Coll et al. (2011).

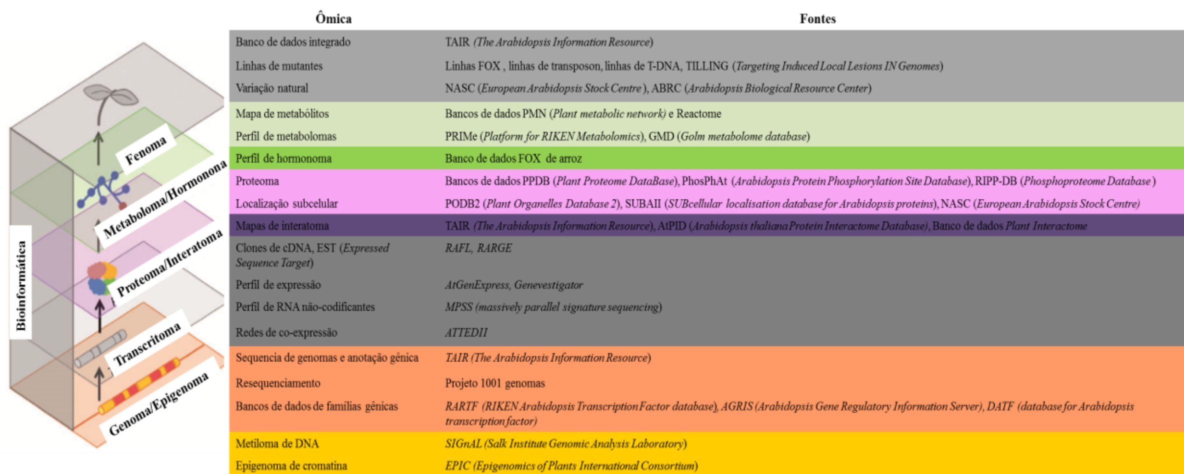
1.4 EXPLORANDO A FUNÇÃO GÊNICA POR MEIO DA GENÔMICA E DA BIOINFORMÁTICA

A genômica é um campo multidisciplinar da ciência que estuda função, evolução e mapeamento de genes e genomas e 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 and Zhang, 2004). Abordagens combinando diversas ferramentas para estudos na área da genômica são fundamentais para uma melhor compreensão do desenvolvimento das plantas e suas respostas a diferentes condições ambientais.

A comparação entre genomas inteiros baseada em alinhamentos de DNA tem sido uma das áreas com grande investimento científico (Miller *et al.*, 2004). Tais comparações aumentam a precisão e profundidade da análise gênica e inferência evolutiva e funcional de espécies modelos para outras espécies sob investigação. A colinearidade entre os genomas é o que embasa o campo da genômica comparativa, permitindo inferir relações entre táxons intimamente relacionados ou não (Paterson *et al.*, 2000). Entre as Brassicaceae, por exemplo, os mapas genéticos mostram colinearidade quase que completa, exceto por pequenas inversões encontradas em *Brassica oleracea* (Lan *et al.*, 2000). Já entre as gramíneas, que inclui muitas culturas alimentares (arroz, trigo, milho, sorgo, cevada, aveia, centeio e painço), a conservação em grandes extensões cromossômicas abrangendo 65 milhões de anos de evolução é uma observação recorrente em diferentes trabalhos (Paterson *et al.*, 2000). A partir da sequência de um gene é possível realizar diversas investigações em bancos de dados e fazer inferências da função a partir da similaridade com sequências conhecidas (Miller *et al.*, 2004). Essa abordagem é o ponto inicial para estudos avançados complementares que incluem a validação experimental dos resultados observados.

Grande aliada dos estudos de genômica comparativa, a bioinformática é essencial para o desenvolvimento de pesquisas que incluem investigação *in silico*. Na segunda década do século 21, as pesquisas baseadas em ômicas (genômica, transcritômica, proteômica, interatômica e metabolômica) permitiram abordar complexos sistemas biológicos globais subjacentes a várias funções da planta (Mochida and Shinozaki, 2011). Essas abordagens tratam cada uma das vias como componentes interligados do sistema celular da planta (Figura 4). Diferentes tipos de conjuntos de dados podem ser gerenciados, como anotação e evolução de genes, investigação de domínios e estrutura gene/proteína, localização subcelular, expressão gênica, análise de metabólitos, modificações pós-traducionais, predição de pequenos RNAs e seus alvos, rede de regulação e interação proteína-proteína (Mochida and Shinozaki, 2011; Bolger *et al.*, 2017).

As análises *in silico* permitidas pela bioinformática são apoiadas no crescente número de bancos de dados disponíveis para *download* ou investigação *online* (Martinez, 2011; Bolger *et al.*, 2017). Na planta modelo *A. thaliana*, por exemplo, no que se refere à anotação de sequências, um expressivo número de genes foram anotados eletronicamente, ou seja, não curados (o que significa que seus dados não são avaliados por um pesquisador) e/ou faltantes de validação experimental (Figura 5) (Bolger *et al.*, 2017). Tais dados demonstram que, embora com crescente dados disponíveis para investigação, a quantidade de informações não exploradas em pesquisa científica (anotados eletronicamente e não anotados) ainda se mantem expressiva.



Ômica	Fontes
Banco de dados integrado	TAIR *(The Arabidopsis Information Resource)*
Linhas de mutantes	Linhas FOX , linhas de transposon, linhas de T-DNA, TILLING (*Targeting Induced Local Lesions IN Genomes*)
Variação natural	NASC (*European Arabidopsis Stock Centre*), ABRC (*Arabidopsis Biological Resource Center*)
Mapa de metabólitos	Bancos de dados PMN (*Plant metabolic network)* e Reactome
Perfil de metabolomas	PRIMe (*Platform for RIKEN Metabolomics*), GMD (*Golm metabolome database*)
Perfil de hormonoma	Banco de dados FOX de arroz
Proteoma	Bancos de dados PPDB (*Plant Proteome DataBase*), PhosPhAt (*Arabidopsis Protein Phosphorylation Site Database*), RIPP-DB (*Phosphoproteome Database*)
Localização subcelular	PODB2 (*Plant Organelles Database 2*), SUBAII (*SUBcellular localisation database for Arabidopsis proteins*), NASC (*European Arabidopsis Stock Centre)*
Mapas de interatoma	TAIR *(The Arabidopsis Information Resource*), AtPID (*Arabidopsis thaliana Protein Interactome Database)*, Banco de dados *Plant Interactome*
Clones de cDNA, EST (*Expressed Sequence Target*)	*RAFL, RARGE*
Perfil de expressão	*AtGenExpress, Genevestigator*
Perfil de RNA não-codificantes	*MPSS (massively parallel signature sequencing*)
Redes de co-expressão	*ATTEDII*
Sequencia de genomas e anotação gênica	*TAIR (The Arabidopsis Information Resource*)
Resequenciamento	Projeto 1001 genomas
Bancos de dados de famílias gênicas	*RARTF (RIKEN Arabidopsis Transcription Factor database), AGRIS (Arabidopsis Gene Regulatory Information Server), DATF (database for Arabidopsis transcription factor)*
Metiloma de DNA	*SIGnAL (Salk Institute Genomic Analysis Laboratory*)
Epigenoma de cromatina	*EPIC (Epigenomics of Plants International Consortium*)

Figura 4: Diferentes áreas ômicas e as fontes de pesquisa disponíveis. Adaptado de Bolger *et al.* (2017).

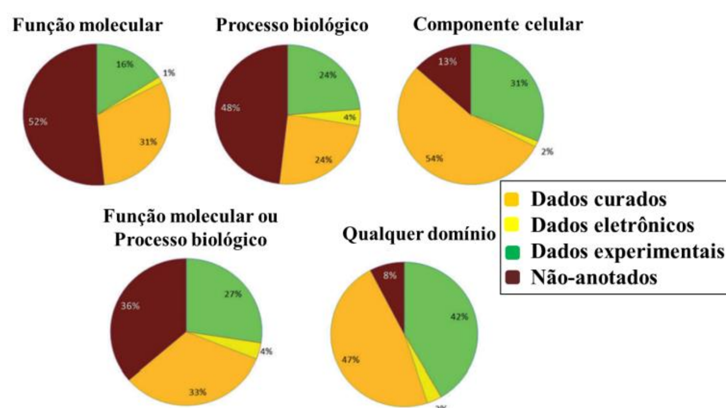


Figura 5: Visão geral do número de genes anotados em *A. thaliana*. Adaptado de Bolger *et al.* (2017).

Nos últimos anos, bancos de dados públicos disponíveis *online* vêm aumento consideravelmente, não apenas para a planta modelo *A. thaliana*, mas também para muitas plantas cultivadas (Tabela 2). Tais bancos contêm informações ditas de longa duração, que se atualizam sistematicamente, de forma que são resultados de uma quantidade maciça de dados gerados pela investigação em vários contextos da biologia vegetal (Dhanapal and Govindaraj, 2015). O perfil público da grande maioria dos bancos garante a máxima acessibilidade e visibilidade a pesquisadores em diferentes campos de interesse, permitindo a formulação de novas hipóteses e a resposta a muitas questões básicas, como estrutura gênica, por exemplo.

Nome da planta	Consórcio/Iniciativa	URL
Alfafa	Consórcio	http://www.alfalfa-genome.org/www/
Algodão	*BGI*	http://www.cottondb.org/wwwroot/cdbhome.php
Arabidopsis thaliana e *Arabidopsis lyrata*	*The Arabidopsis Genome Initiative*	http://www.arabidopsis.org/
Arroz	Consórcio (IRGPS)	http://rgp.dna.affrc.go.jp/E/IRGSP/index.html
Banana	Consórcio	http://www.musagenomics.org/
Batata	Consórcio (PGSC)	http://www.potatogenome.net/index.php/Main_Page
Brachypodium distachyon	Consórcio	http://www.brachypodium.org/
Cacau	Consórcio	http://www.cacaogenomedb.org/
Cevada	Consórcio	http://www.public.iastate.edu/~imagefpc/IBSC%20 Webpage/IBSC%20Template-home.html
Colza	Consórcio (MGPB)	http://www.plantgdb.org/BrGDB/
Feijão	Consórcio	http://www.phytozome.net/commonbean.php
Feijão-da-china	Dado não disponível	http://vigra.comparative-legumes.org/
Gandu	*International Initiative for Pigeonpea Genomics* (IIPG)	http://www.icrisat.org/gt-bt/iipg/Home.html
Grão-de-bico	Consórcio (ICRISAT-BGI)	http://www.icrisat.org/gt-bt/ICGGC/GenomeManuscript.html
Lotus japonicus	Consórcio	http://www.kazusa.or.jp/lotus/
Maconha	Consórcio	http://genome.ccbr.utoronto.ca/index.html?org=C.+sativa&db=canSat3&hgsid=11252
Maça	Consórcio	http://www.rosaceae.org/species/malus/malus_spp
Mamão	Consórcio	http://www.plantgdb.org/CpGDB/
Mamona	TIGR	http://castorbean.jcvi.org/
Melancia	International Watermelon Genomics Initiative	http://www.iwgi.org/
Milho	Consórcio	http://www.maizegdb.org/
Morango	Consórcio	http://www.strawberrygenome.org/
Painço	*Beijing Genomics Institute and the Joint Genomes Institute*	http://www.phytozome.net/foxtailmillet.php
Pepino	*International Cucurbit Genomics Initiative* (ICuGI)	http://www.icugi.org/cgi-bin/ICuGI/index.cgi
Pêssego	*International Peach Genome Initiative*	http://www.rosaceae.org/peach/genome
Physcomitrella patens	*JGI*	http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html
Pinhão-manso	Consórcio	http://www.kazusa.or.jp/jatropha/
Populus trichocarpa	*JGI*	http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html
Soja	*JGI*	https://phytozome.jgi.doe.gov/
Sorgo	*JGI*	http://www.plantgdb.org/SbGDB/
Tamara	Consórcio	http://qatar-weill.cornell.edu/research/datepalmGenome/download.html
Thellungiella parvula	Consórcio	http://www.brassica.info/info/events.php
Tomate	Consórcio	http://solgenomics.net/organism/Solanum_lycopersicum/genome
Trigo	*International Wheat Genome Sequencing Consortium (IWGSC)*	http://www.wheatgenome.org/

Tabela 2: Exemplos de genomas de plantas disponíveis publicamente. Adaptado de Dhanapal and Govindaraj (2015).

O avanço em muitas áreas de estudos biológicos vem sendo fomentado pelas plataformas de bioinformática, especialmente no que diz respeito à investigação de questões genômicas iniciais e a possível fluidez para outras tecnologias de alto rendimento (Figura 6). Abordagens combinatórias, utilizando múltiplas plataformas ômicas e a integração dos seus resultados são estratégias eficazes para auxiliar a esclarecer os sistemas moleculares como um todo, objetivando como última meta a melhoria da produtividade das plantas de cultivo (Dhanapal and Govindaraj, 2015). As abordagens combinatórias são viáveis devido aos diferentes bancos que hospedam dados diversos (Tabela 3). Dessa maneira, o progresso permitido pelas plataformas ômicas na área de pesquisa vegetal e a integração com a ciência animal tornou essas metodologia importantes para a investigação da função gênica e sua associação com alterações fenotípicas.

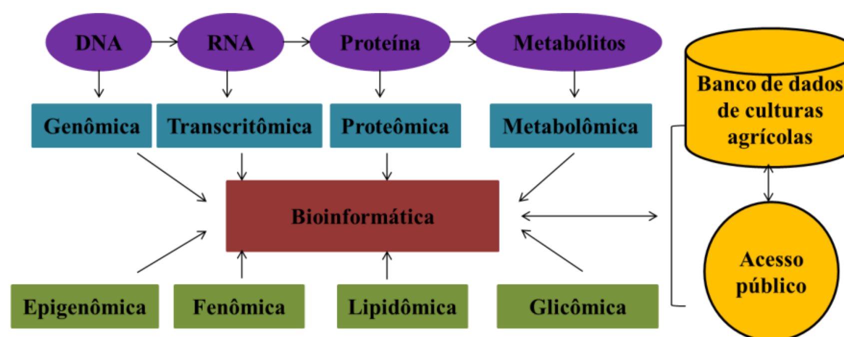


Figura 6: Fluxograma de diversas abordagens combinatórias para pesquisa em ômica. Adaptado de Dhanapal and Govindaraj (2015).

A presente tese enquadra-se no contexto atual de valer-se de diferentes ferramentas de bioinformática, tendo como objetivo final a caracterização *in silico* de algumas famílias gênicas previamente propostas por Coll *et al.* (2011) como constituintes do morteossomo e relacionadas com o processo de PCD. A linha de pesquisa seguida por essa tese representa o desdobramento dos trabalhos conduzidos durante minha dissertação de Mestrado, onde foram iniciados os estudos investigativos sobre a família LSD (elemento central no morteossomo). Durante o Mestrado, buscamos compreender e caracterizar a família gênica LSD, com ênfase na espécie modelo soja. Esse trabalho apresentou a identificação e caracterização dos oitos genes pertencentes à família LSD em soja, mostrando a modulação da expressão gênica em condições de estresse biótico e abiótico desencadeantes de PCD (Cabreira *et al.*, 2013). Assim, nossos dados reforçaram o envolvimento dos genes LSD no controle da PCD e foram pioneiros na caracterização completa da família gênica em uma espécie vegetal.

Como continuidade desse assunto, a presente tese foi desenvolvida a fim de realizar a caracterização completa de algumas famílias de genes relacionados com PCD e/ou pertencentes ao morteossomo. Foram abordados diferentes aspectos para caracterização gênica, incluindo a identificação e investigação da estrutura de genes parálogos e ortólogos, análise evolutiva, localização subcelular, verificação do perfil de expressão e possíveis redes de interação proteína-proteína. A caracterização dessas famílias gênicas visa contribuir com os conhecimentos sobre o mecanismo de resposta das plantas a condições de estresse e fornecer informações importantes para trabalhos futuros mais aprofundados nessa área do conhecimento.

Os resultados obtidos estão divididos em cinco artigos, sendo que cada um é composto por dados referentes a uma família gênica, sendo distribuídos da seguinte maneira:

a) Capítulo 1: esse capítulo é resultado da investigação da família gênica LSD, que inclui *AtLSD1,* gene com papel central no morteossomo. Esse capítulo está apresentado na forma de um artigo científico, publicado em 2015 no periódico *Molecular Genetics and Genomics.*

b) Capítulo 2: esse capítulo é constituído pelo estudo da família gênica Metacaspase, que contém o gene codificante da proteína *At*MC1, que interage diretamente com *At*LSD1 durante PCD. O capítulo está apresentado em forma de artigo científico, publicado em 2015 no periódico *Functional and Integrative Genomics*.

c) Capítulos 3 e 4: esses capítulos são constituídos por resultados de estudo de alguns dos genes com função desconhecida descritos no morteossomo, os quais compõem as famílias gênicas NF-Y e PLAC8, respectivamente.

d) Capítulo 5: esse capítulo é dedicado à caracterização da família gênica composta por genes *GILP*, que não foi incluída no modelo do morteossomo proposto por Coll *et al.* (2011). Entretanto, em trabalho publicado no mesmo ano, foi mostrado que a proteína *At*GILP1 interage com *At*LSD1 e regula negativamente PCD (He *et al.*, 2011).

Nome do banco de dados	Aplicações e dados disponíveis
AgBase	Funcional genômica em plantas cultiváveis.
AutoSNPdb	Identificação de SNPs e sequências de EST em arroz, cevada e Brassica.
BarleyBase	Visualização e análise de dados de microarranjo em plantas.
BBGD	Banco de dados online para análises genômicas em mirtilo.
BIOGEN BASE-CASSAVA	Informações genômicas e fenômicas de mandioca.
CastorDB	Banco de dados contendo diversas informações sobre *Ricinus communis*.
ChromDB	Banco de dados de cromatina.
CR-EST	Fonte de ESTs em plantas cultiváveis.
CSRDB	Banco de dados de pequenos RNAs em cereais.
DEBDOM	Banco de dados do genoma de Banana.
DRASTIC	Informações sobre expressão gênica de espécies de plantas em resposta a patógenos e alterações ambientais.
FLAGdb++	Informações do genoma de *Arabidopsis*.
GabiPD	Banco de dados integrado de ômicas.
GCP	Informações comparativas de genes responsivos a estresses em plantas.
GDR	Banco de dados genômicos para espécies de Rosaceae.
GeneCAT	Análises de co-expressão de genes.
GeneSeqer@PlantGDB	Predição de estrutura gênica em genomas de plantas.
GERMINATE	Informações genotípicas e fenotípicas em genomas de plantas.
GGT	Software para análise e visualização de dados genéticos.
GrainGenes	Informações sobre genomas de pequenos cereais.
Gramene	Banco de dados comparativo de diversas espécies de gramíneas.
MaizeGDB	Informações genéticas e genômicas sobre milho.
MANET	Rede de ancestralidade molecular. Traça a evolução da arquitetura de proteínas em rede de metabólitos.
Medicago	Banco de dados para a mineração de dados na planta modelo *Medicago truncatula* contendo genômica integrada, genética e informações biológicas.
Metacrop e Metacrop 2.0	Banco de dados detalhado do metabolismo de culturas agrícolas.
Narcisse	Conservação do genoma a partir de uma visão de espelho de regiões de sintenia.
NIASGBdb	Informações sobre genética de plantas agrícolas e doenças.
P3db	Fosforilação de proteínas em plantas.
Panzea	Apresenta relações entre genótipos e variações fenotípicas em milho.
Bando de dados de pimenta (*Pepper EST dabatase*)	Dados de transcritoma de pimenta.
PIP	Informações sobre marcadores polimórficos em introns de plantas.
PLACE	Banco de *cis*-elementos em DNA de plantas.
Banco de snoRNA	Dados sobre pequenos RNAs nucleares em plantas.
PlantCARE	Banco de *cis*-elementos em DNA de plantas.
PlantTFDB	Banco de dados de fatores de transcrição em plantas.
PlantTribes	Genômica comparativa de genes e famílias gênicas em diversas plantas.
PlecDom	Informações sobre domínios de Lectina em plantas.
PlnTFDB	Banco de dados de fatores de transcrição em plantas.
PmiRKB	Informações sobre microRNAs em plantas.
PMRD	Informações sobre microRNAs e seus alvos em plantas.
PODB	Coleção de organelas visualizáveis de plantas e protocolos para diversas finalidades.
POGs/PlantRBP	Genômica comparativa em arroz, milho e *Arabidopsis*.
PoMaMo	Informações sobre o genoma da batata.
PREPSuite	Informações sobre edição de RNA em plantas.
PRGDB	Informações sobre genes envolvidos em mecanismos de defesa em plantas.
pssRNAMiner	Servidor para análise de cascatas de regulação de pequenos RNAs em plantas.
RadishBase	Banco de dados para análises genéticas e genômicas de radite.
RoBuST	Banco de dados integrado de culturas de raízes e bulbo

Tabela 3: Bancos de dados disponíveis em diferentes espécies vegetais e suas aplicações. Adaptado de Dhanapal e Govindaraj (2015).

2) OBJETIVO GERAL

O objetivo geral deste estudo foi caracterizar genes pertencentes à rede de controle de PCD descrita como morteossomo, com ênfase especial às famílias gênicas LSD (*Lesion Simulating Disease*), Metacaspase, NF-Y (*Nuclear Factor Y*), PLAC8 (*Placenta* Specific 8) e GILP (*GSH-induced LITAF domain protein*). Para cada família gênica, as questões abordadas foram determinadas conforme os dados disponíveis e as perguntas científicas a serem respondidas. Assim, os objetivos específicos estão descritos nos diferentes manuscritos publicados e/ou submetidos.

CAPÍTULO 1

Mol Genet Genomics
DOI 10.1007/s00438-015-1060-4

CrossMark

ORIGINAL PAPER

The phylogeny and evolutionary history of the *Lesion Simulating Disease* (*LSD*) gene family in Viridiplantae

Caroline Cabreira[1] · Alexandro Cagliari[2] · Lauro Bücker-Neto[1] · Márcia Margis-Pinheiro[1] · Loreta B. de Freitas[1] · Maria Helena Bodanese-Zanettini[1]

Received: 20 November 2014 / Accepted: 3 May 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract The *Lesion Simulating Disease* (LSD) genes encode a family of zinc finger proteins that play a role in programmed cell death (PCD) and other biological processes, such as plant growth and photosynthesis. In the present study, we report the reconstruction of the evolutionary history of the *LSD* gene family in Viridiplantae. Phylogenetic analysis revealed that the monocot and eudicot genes were distributed along the phylogeny, indicating that the expansion of the family occurred prior to the diversification between these clades. Sequences encoding proteins that present one, two, or three LSD domains formed separate groups. The secondary structure of these different LSD proteins presented a similar composition, with the β-sheets being their main component. The evolution by gene duplication was identified only to the genes that contain three LSD domains, which generated proteins with equal structure. Moreover, genes encoding proteins with one or two LSD domains evolved as single-copy genes and did not result from loss or gain in LSD domains. These results were corroborated by synteny analysis among regions containing paralogous/orthologous genes in *Glycine max* and *Populus trichocarpa*. The *Ka/Ks* ratio between paralogous/orthologous genes revealed that a subfunctionalization process possibly could be occurring with the *LSD* genes, explaining the involvement of LSD members in different biological processes, in addition to the negative regulation of PCD. This study presents important novelty in the evolutionary history of the *LSD* family and provides a basis for future research on individual LSD genes and their involvement in important pathway networks in plants.

Keywords *Lesion Simulating Disease* (LSD) · Zinc finger LSD domain · *LSD* family phylogeny · Gene duplication · Synteny · Protein secondary structure

Communicated by B. Yang.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-015-1060-4) contains supplementary material, which is available to authorized users.

✉ Maria Helena Bodanese-Zanettini
maria.zanettini@ufrgs.br; mhbzanettini@yahoo.com.br

[1] Departamento de Genética, Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul (UFRGS), PO Box 15053, Porto Alegre, RS CEP 91501-970, Brazil

[2] Universidade Estadual do Rio Grande do Sul (UERGS), Santa Cruz do Sul, RS, Brazil

Introduction

The sessile nature of plants means that they are able to adjust their metabolic processes according to environmental changes. During the signaling of stress responses, which trigger changes at the transcriptome, cellular, and physiological levels (Atkinson and Urwin 2012), one of the well-characterized reactions is known as the hypersensitive response (HR), which triggers the network of programmed cell death (PCD) (van Doorn et al. 2011). PCD regulates the growth and spread of lesions at the initial site of damage and in the cells that surround this area (Mur et al. 2008; Rusterucci et al. 2001). Efforts to elucidate the genes that mediate the PCD processes have resulted in the identification of the *Lesion Simulating Disease* (LSD) protein family. The LSD family was initially identified in *Arabidopsis thaliana* by the analysis of mutants that spontaneously form necrotic lesions in the absence of pathogenic infection (Dietrich et al. 1994). Genes belonging to this family negatively regulate PCD under stress conditions (Dietrich et al.

Published online: 17 May 2015

1997; Epple et al. 2003) and are exclusive to Viridiplantae (Cabreira et al. 2013). The *A. thaliana LSD1* gene (*AtLSD1*) is the well-characterized member of this family and acts as a cellular hub that makes a central contribution to the oxidative stress response (Kaminaka et al. 2006). This gene enhances ROS (reactive oxygen species)-scavenging capacity in the signaling pathway that is related to oxidative stress (Jabs et al. 1996; Kliebenstein et al. 1999; Mateo et al. 2004; Mullineaux and Baker 2010). Together with *EDS1* (*Enhanced disease susceptibility1*) and *PAD4* (*Phytoalexin deficient4*), *AtLSD1* acts as an ROS/ethylene homeostatic switch, controlling light acclimation and pathogen defense (Muhlenbock et al. 2008). Moreover, *AtLSD1* regulates acclimation to excess excitation energy (Mateo et al. 2004), the response to cold stress (Huang et al. 2010), and lysigenous aerenchyma formation under hypoxic conditions (Muhlenbock et al. 2007). The role of *AtLSD1* in the regulation of cellular signaling homeostasis, photosynthesis, water use efficiency, and seed yield has recently been reported (Wituszynska et al. 2013). Additionally, the gene *OsLSD1* (*Oryza sativa LSD1*) regulates PCD and callus differentiation (Wang et al. 2005); *OsLOL2* (*Oryza sativa LSD1-like 2*) is involved in rice growth and disease resistance (Bhatti et al. 2008; Xu and He 2007); *BohLOL1* (*Bambusa oldhamii LDS1-like 1*) (Yeh et al. 2011) participates in growth and biotic stress response in bamboo, and *GmLSD1* to *GmLSD8* (*Glycine max LSD*) are modulated in response to fungi and dehydration stresses (Cabreira et al. 2013). These studies have demonstrated the involvement of LSD members in important biological processes in plants, especially responses to abiotic and biotic stresses.

The LSD proteins contain exclusively the zinc finger LSD domain CxxCRxxLMYxxGASxVxCxxC (Dietrich et al. 1997). This characteristic allows differentiation between the LSD gene family and the metacaspase gene family, since metacaspase genes present one zinc finger LSD domain and the peptidase_C14 domain. The statement about which domain(s) compose the proteins is important to distinguish between these gene families. The consensus sequence of LSD domains in LSD proteins shows broad conservation (Cabreira et al. 2013) and is required for protein–protein interactions (Coll et al. 2010; He et al. 2011a, b; Kaminaka et al. 2006; Li et al. 2013). The LSD domains contain conserved cysteine residues that allow the formation of the C2C2 arrangement (Cabreira et al. 2013), essential for the nuclear localization of the protein (He et al. 2011a). The changes in LSD protein conformation mediated by the presence of a different number of LSD domains and the evolution mechanism that is involved in the generation of LSD proteins (containing one, two, or three LSD domains) remain unexplored.

In the present study, we report the phylogenetic reconstruction of the *LSD* gene family in Viridiplantae considering the secondary structure of proteins, chromosomal location, duplication pattern, and synteny analysis. These results provide important insights about the evolutionary history of the *LSD* family. These insights could be useful in further studies of PCD mechanisms in plants.

Materials and methods

LSD gene annotation and structure analysis

Previously identified sequences (Cabreira et al. 2013) were used to reconstruct an LSD phylogenetic tree. These sequences correspond to genes identified in *Manihot esculenta*, *Ricinus communis*, *Linum usitatissimum*, *Populus trichocarpa*, *Medicago truncatula*, *Lotus japonicus, Phaseolus vulgaris*, *Glycine max*, *Cucumis sativus, Prunus persica*, *Malus domestica*, *Arabidopsis thaliana, Arabidopsis lyrata, Capsella rubella*, *Brassica rapa*, *Thellungiella halophila*, *Carica papaya, Citrus sinensis*, *Citrus clementina*, *Eucalyptus grandis*, *Vitis vinifera*, *Mimulus guttatus*, *Aquilegia caerulea*, *Sorghum bicolor*, *Zea mays*, *Setaria italica*, *Oryza sativa*, *Brachypodium distachyon*, *Selaginella moellendorffii*, *Physcomitrella patens*, *Chlamydomonas reinhardtii,* and *Volvox carteri*. In reference to our previous annotation, the following four corrections were performed according to the recent Phytozome update (v.9.1, http://phytozome.org): (a) the sequences previously annotated as *Aco1* (*Aquilegia caerulea 1*) and *Aco2* were removed from the analysis (*Aco1* is unavailable in the Phytozome, and *Aco2* corresponds to an alternative transcript of *Aco3),* and the sequences previously named *Aco3* and *Aco4* were renamed as *Aco1* and *Aco2,* respectively; (b) the sequences annotated as *Ppe4* (*Prunus persica 4*) and *Bdi5* (*Brachypodium distachyon 5*) were also removed because they are unavailable in the Phytozome, precluding further analysis. The sequences used for alignments are provided in Supplementary Material 1.

We analyze the intron/exon structure and the intron phase distribution using the Gene Structure Display Server (GSDS) program, developed by the Center of Bioinformatics (CBI), Peking University (Guo et al. 2007). The sequences were accessed using the Phytozome database (June/2013).

Phylogenetic analysis of the *LSD* genes

The complete sequences from the 113 LSD genes were aligned using the Muscle algorithm as implemented in MEGA v.5.05 (Tamura et al. 2011). The sequences were edited manually and back-translated to nucleotide sequences for analyses at the DNA level. Contiguous insertion–deletion events (indels) that had more than one base pair (bp) were treated as single mutations (Simmons and Ochoterena 2000).

A Bayesian inference (BI) was generated using BEAST v. 1.4.7 (Drummond and Rambaut 2007); a run of 20^7 chains was performed, and the trees were sampled 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 TRACER v.1.4 (http://beast.bio.ed.ac.uk/Tracer) was used to check the convergence of the Monte Carlo Markov chains (MCMCs) and for adequate effective sample sizes (EES >200) after the first 10 % of the generations had been deleted as burn-in. The final joint sample was used to estimate the maximum clade credibility tree with the TreeAnnotator program, which is part of the BEAST package. The statistical support for the clades was determined by accessing the Bayesian posterior probability (PP). The trees were visualized using FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Protein secondary structure prediction

We previously identified LSD proteins containing three, two, and one LSD domain(s) (Cabreira et al. 2013). To evaluate whether proteins with a distinct number of LSD domains adopt a different structure, we analyzed the secondary structure of the LSD proteins. The analysis of protein sequences of *Gma*LSD1, *Gma*LSD8, and *Osa*LSD5 (*Oryza sativa* LSD5), which have three, two, and one LSD domain(s), respectively, was conducted using the PSIPRED secondary structure prediction server (http://bioinf.cs.ucl.ac.uk/psipred/) (Buchan et al. 2010; McGuffin et al. 2000). We used the protein that does not belong to the LSD family as a negative control for this analysis. The *Ath*MC1 (*Arabidopsis thaliana* metacaspase 1_At1g02170) contains the LSD1-like zinc finger domain (characteristic of the LSD family) and the peptidase_C14 domain (characteristic of the caspase family). We analyzed the positions of the LSD domains using the SMART database (http://smart.embl-heidelberg.de/).

Chromosomal localization, duplication events, synteny, and *Ka*/*Ks* substitution rate analysis

We analyzed the chromosomal localization and the mechanisms that are involved with the LSD genes evolution. The species previously studied (Cabreira et al. 2013) and with a genome available in the PLAZA v.2.5 (http://bioinformatics.psb.ugent.be/plaza/) database were analyzed using the WGmapping tool, search id HOM000593 (July/2013).

We analyzed 100 kb syntenic regions between orthologous genes in *G. max* LSD *(GmaLSD)* and *Populus trichocarpa LSD (PtrLSD)*. In addition, the syntenic regions that contain *GmaLSD* and *PtrLSD* paralogous genes were considered. We performed the analyses in the Genome Duplication Database (PGDD) (http://chibba.agtec.uga.edu/duplication/) (Lee et al. 2013), which uses the BLASTP tool to search for potential anchors (E <1e−5, top 5 matches) among every possible chromosome pair in multiple genomes. The homolog pairs identified are used as the input in the multiple collinearity scan (MCsan) program, and an E value <1e−10 is used as a significant cutoff (June/2013). We search the putative classification of the genes identified in the Phytozome and the PLAZA databases (July/2013). In addition, we accessed the protein similarity of duplicated genes in Phytozome (August/2013).

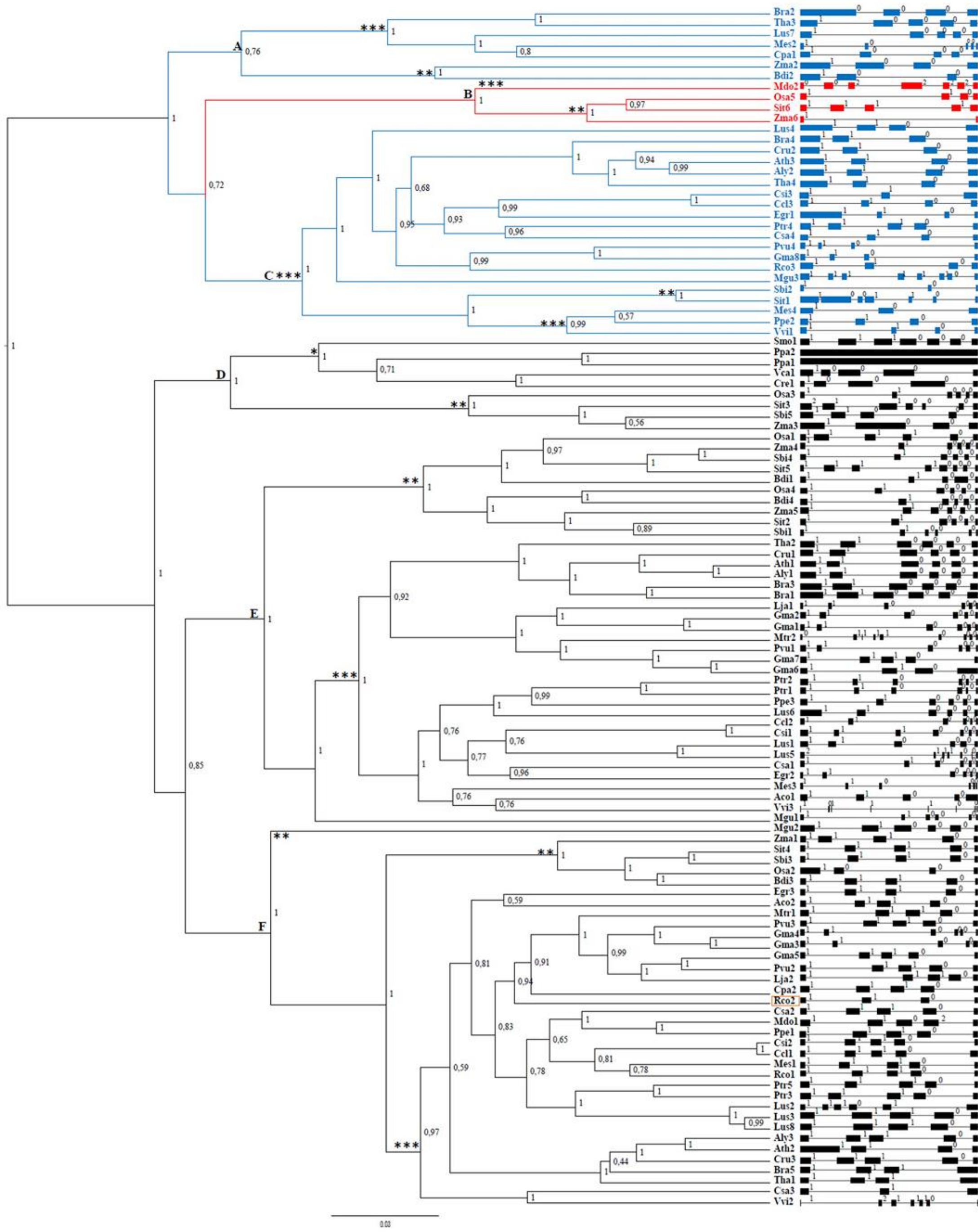
We analyze the ratio of the number of nonsynonymous substitutions per nonsynonymous site (*K*a) to the number of synonymous substitutions per synonymous site (*K*s) in the orthologous and paralogous *GmaLSD* and *PtrLSD* genes using the PGDD database (http://chibba.agtec.uga.edu/duplication/) (June/2013). In this database, the protein sequences were aligned using CLUSTALW as a guide to the CDS (coding DNA sequence) alignments by PAL2NAL (http://abacus.gene.ucl.ac.uk/software/paml.html). Finally, the Nei–Gojobori method was used to generate the *Ka* and *Ks* ratio.

Results

Genes with different *LSD* domain structures form separate groups along the *LSD* gene family evolution

The unrooted phylogenetic tree generated by Bayesian information (BI) was well supported, as show by posterior probability (PP) (Fig. 1). Two major clades of sequences were observed, the first comprising the sequences with one and two LSD domains and the second including the sequences with three LSD domains (Fig. 1). The phylogeny was structured according to the number of LSD domains encoded by the LSD genes. The *Rco2* (*Ricinus communis 2*) gene was the unique exception to this pattern. It encodes two LSD domains and localized in the group of genes that encode proteins with three LSD domains. We do not identify clades comprising only monocot and eudicot genes. These genes were interspersed along the phylogenetic tree, forming several subclades. We found a subclade containing only LSD genes from chlorophyte, lycophyte, and moss.

We compare the exons/introns of the individual LSD genes (Fig. 1). Among the 113 sequences, the number of introns ranged from 0 (Ppa1 and Ppa2) to 9 (Vvi3). In general, phylogenetically related genes exhibited a similar gene intron/exon composition. In the 27 sequences of subclades A and C (two LSD domains; Fig. 1, blue genes), around 63 % contains three introns, except Sbi2, Mes4, and Csi3 (two introns); Ptr4, Lus7, Mes2, Cpa1, and Tha3 (four introns); Sit1 (six introns); and Mgu3 (seven introns). The subclade B genes (one LSD domain; Fig. 1, red genes) contain one to six introns. The subclade D (three LSD domains; Fig. 1, black genes) formed by monocots and


A
B
C
D
E
F
Bra2
Tha3
Lus7
Mes2
Cpa1
Zma2
Bdi2
Mdo2
Osa5
Sit6
Zma6
Lus4
Bra4
Cru2
Ath3
Aly2
Tha4
Csi3
Ccl3
Egr1
Ptr4
Csa4
Pvu4
Gma8
Rco3
Mgu3
Sbi2
Sit1
Mes4
Ppe2
Vvi1
Smo1
Ppa2
Ppa1
Vca1
Cre1
Osa3
Sit3
Sbi5
Zma3
Osa1
Zma4
Sbi4
Sit5
Bdi1
Osa4
Bdi4
Zma5
Sit2
Sbi1
Tha2
Cru1
Ath1
Aly1
Bra3
Bra1
Lja1
Gma2
Gma1
Mtr2
Pvu1
Gma7
Gma6
Ptr2
Ptr1
Ppe3
Lus6
Ccl2
Csi1
Lus1
Lus5
Csa1
Egr2
Mes3
Aco1
Vvi3
Mgu1
Mgu2
Zma1
Sit4
Sbi3
Osa2
Bdi3
Egr3
Aco2
Mtr1
Pvu3
Gma4
Gma3
Gma5
Pvu2
Lja2
Cpa2
Rco2
Csa2
Mdo1
Ppe1
Csi2
Ccl1
Mes1
Rco1
Ptr5
Ptr3
Lus2
Lus3
Lus8
Aly3
Ath2
Cru3
Bra5
Tha1
Csa3
Vvi2

◀ **Fig. 1** Evolutionary relationships among the 113 *LSD* genes based on unrooted tree obtained by Bayesian inference. The PP is shown above the branches. The taxa terminologies are abbreviated using the first letter of the genus and the first two letters of the species named according to (Cabreira et al. 2013), (see Supplementary Table S2). The *black shading* indicates the sequences with three LSD domains. The *blue shading* indicates the sequences with two LSD domains. The *red shading* indicates the sequences with one LSD domain. The eudicot species (*triple asterisk*), the monocot species (*double asterisk*), and the basal organism (*asterisk*) are shown. The right side provides a detailed illustration of the relative intron/exon length and number in each gene. The intron/exon figures were generated using the GSDS program and redrawn. The *orange square* indicates the *Rco2* sequence (color figure online)

chlorophyte, lycophyte, and moss comprised sequences that ranged from zero (Ppa1 and Ppa2) to six introns (Sit 3 and Smo1). In the 37 sequences of subclade E (three LSD domains; Fig. 1, black genes), 76 % contains five introns, except Gma6 (three introns); Bdi1, Gma7, and Aco1 (four introns); Csi1 and Lus5 (six introns); Sit5 (seven introns); and Vvi3 and Mtr2 (eight introns). In the 35 sequences of subclade F (three LSD domains; Fig. 1, black genes), around 72 % contained four introns and the sequences Csa3 (two introns); Rco2, Tha1, Bra5, Ath2, and Osa2 (three introns); Mgu2, Gma4, and Lus2 (five introns); and Vvi2 (six introns) were exceptions.

The product of RNA splicing results from the joining of a 5′ splicing site to a 3′ splicing site (Sharp 1981). The RNA splicing is an important mechanism, which can generate different intron phases and different reading frames. The reading frame is related to the position of the intron within or between codons. Introns between codons are designated phase 0, introns between the first and the second bases of a codon are designated phase 1, and introns between the second and the third bases of a codon are designated phase 2. A correlation between intron phases results in an excess of symmetric exons and symmetric exon sets were reported (Long et al. 1995). We observe that the intron phases were remarkably well conserved in the related genes. In general, considering the subclades that contain sequences with three LSD domains, subclade E had the first and second introns in phase 1, and subclade F had the first, second, and third introns in phase 1. We identified sequences with variable intron phase in subclade D (formed by monocot and chlorophyte, lycophyte, and moss). Among the sequences with two LSD domains, the majority possessed the first and second introns in phase 1.

Distinct number of *LSD* domains and the maintenance of secondary structure in *LSD* proteins

The prediction of secondary structure, which constitutes the second level of protein structure, provides the location of β-sheets or/and α-helices within a protein or protein family. To predict which structure (β-sheets or α-helices) is predominant in the chain of LSD proteins, we analyzed the amino acid sequences of three different LSD proteins. In general, the predictions were accurate and reliable, based on the confidence of prediction value represented in Fig. 2 (blue bars). We observed that β-sheets are the main component of the secondary structures of *Gma*LSD1 (three LSD domains), *Gma*LSD8 (two LSD domains), and *Osa*LSD5 (one LSD domains) (see Fig. 2 and Supplementary Material 2 for details). As exemplified by soybean and rice, these observations were confirmed in several species such as *M. esculenta*, *R. communis*, *C. papaya, Z. mays,* and *S. bicolor* (Supplementary Material 3). Therefore, the number of LSD domains did not alter the secondary structure of LSD proteins. *At*MC1 (negative control) revealed a different protein structure than that of the LSD proteins. It has both β-sheets and α-helices, but the α-helices are the main component. We identified two β-sheets in regions that contain the zinc finger LSD domains in LSD proteins, but only one β-sheet was found in this region in the *At*MC1 control.

Genome distribution and expansion of *LSD* genes

Figure 3 illustrates the chromosomal location of LSD genes and the putative mechanism of their duplication. The complete results are presented in Supplementary Table S1. We found LSD genes distributed on various chromosomes throughout the genomes analyzed.

Single LSD isoforms (without a corresponding paralogous gene) were more often found than genes subjected to block and tandem duplication processes. This was observed for the single LSD gene in *Chlamydomonas reinhardtii* and for the two copies of the LSD gene in chromosome 7 and 4 for *S. bicolor* and *Z. mays*, respectively. Nevertheless, we observed that the block and tandem duplication contributed substantially to LSD family expansion in species such as *P. trichocarpa* and *G. max*. These species have passed through polyploidization events.

We identified that the duplication processes were restricted to genes with three LSD domains (Supplementary Table S1). All genes that contained two or one LSD domains did not have a matching paralogous gene.

Although some gene copies were generated from tandem or block duplication, we found that some genes appeared as part of both processes, as the genes in chromosome 3 and 12 of *O. sativa* ssp. *japonica*. These data indicate that these genes are included in a set of tandem-duplicated genes that are present in a block-duplicated region.

Syntenic analysis of *LSD* orthologous/paralogous genes in *Glycine max* and *Populus trichocarpa*

We analyze a 100-kb syntenic region between LSD orthologous/paralogous genes. Supplementary Table S2 shows

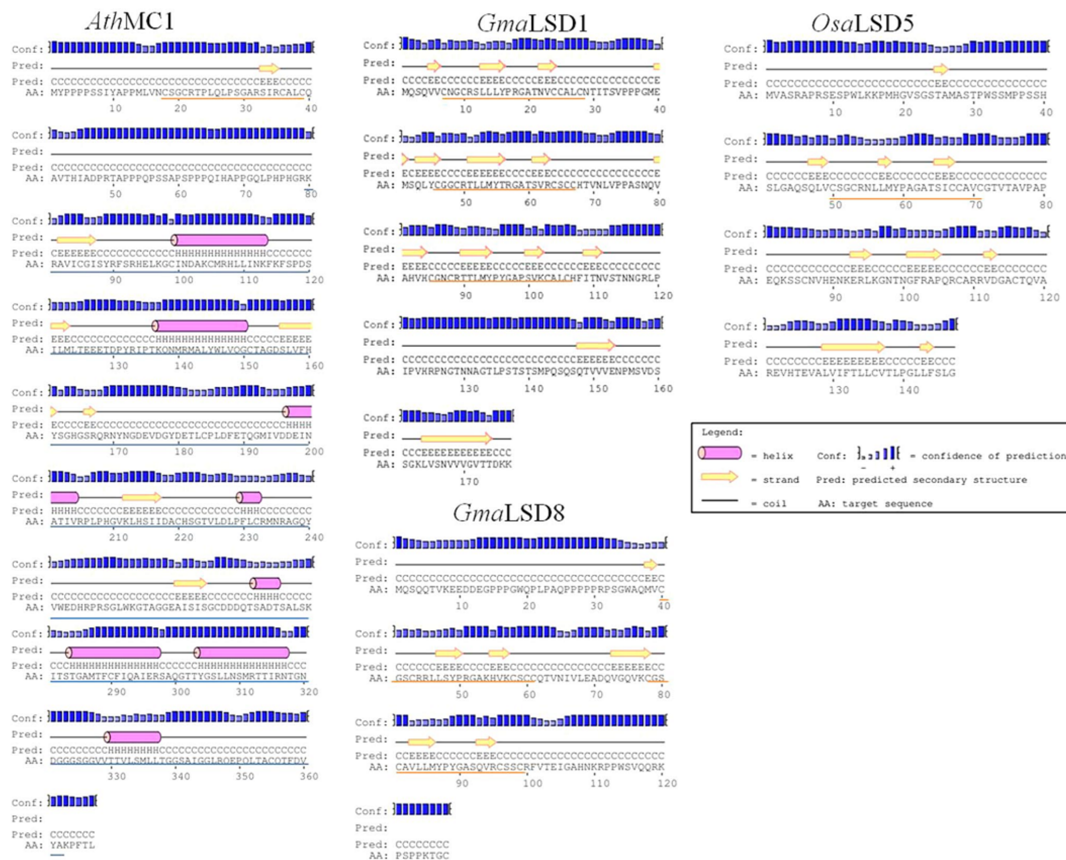


Fig. 2 Secondary structure of LSD proteins. *Gma*LSD1 (three LSD domains), *Gma*LSD8 (two LSD domains), and *Osa*LSD5 (one LSD domain) were analyzed using the PSIPRED program. *Ath*MC1 was used as a control for the analysis. The *orange line* under the amino acid sequence highlights the LSD domains, and the *blue line* highlights the C14 domain (color figure online)

the protein family classification for all genes per syntenic region based on the Phytozome and PLAZA databases. In addition, we analyze the *Ka*/*Ks* values.

We chose two related species (soybean and poplar) with the entire genomes assembled and annotated and with a relatively large number of *LSD* members. Poplar and soybean have experienced different numbers of genome duplications: poplar underwent one complete genome duplication after the genome triplication event (γ) and soybean passed through two duplication events (Lin and Paterson 2011). Therefore, this analysis allows a comparison between syntenic regions after a different number of genome duplications.

Figure 4 illustrates the syntenic region between orthologous genes in soybean (*GmaLSD)* compared with poplar (*PtrLSD)* genes. Poplar and soybean frequently contain the same protein families distributed in their syntenic regions, indicating a common origin for the chromosomes. We observed *Ka*/*Ks* ratios in orthologous genes that suggest a negative selection. Only one exception to this pattern was observed in the syntenic regions between *GmaLSD3* and *PtrLSD1* and *GmaLSD3* and *PtrLSD2*. In these regions, the *Ka*/*Ks* ratio of Glyma15g22100 and Ptrichoc_0011s15820 and Glyma15g22100 and Ptrichoc001s43290 (belonging to the nitrate, formate, and iron dehydrogenase family) suggests neutral selection.

We observed three different types of block duplication between poplar and soybean (Supplementary Table S2). A small duplication block was found in the syntenic regions that contain *GmaLSD1* and *PtrLSD1* (78.0 % protein similarity), *GmaLSD1* and *PtrLSD2* (75.1 % protein similarity), *GmaLSD2* and *PtrLSD1* (69.3 % protein similarity), *GmaLSD3* and *PtrLSD1* (67.2 % protein similarity), *GmaLSD3* and *PtrLSD2* (59.9 % protein

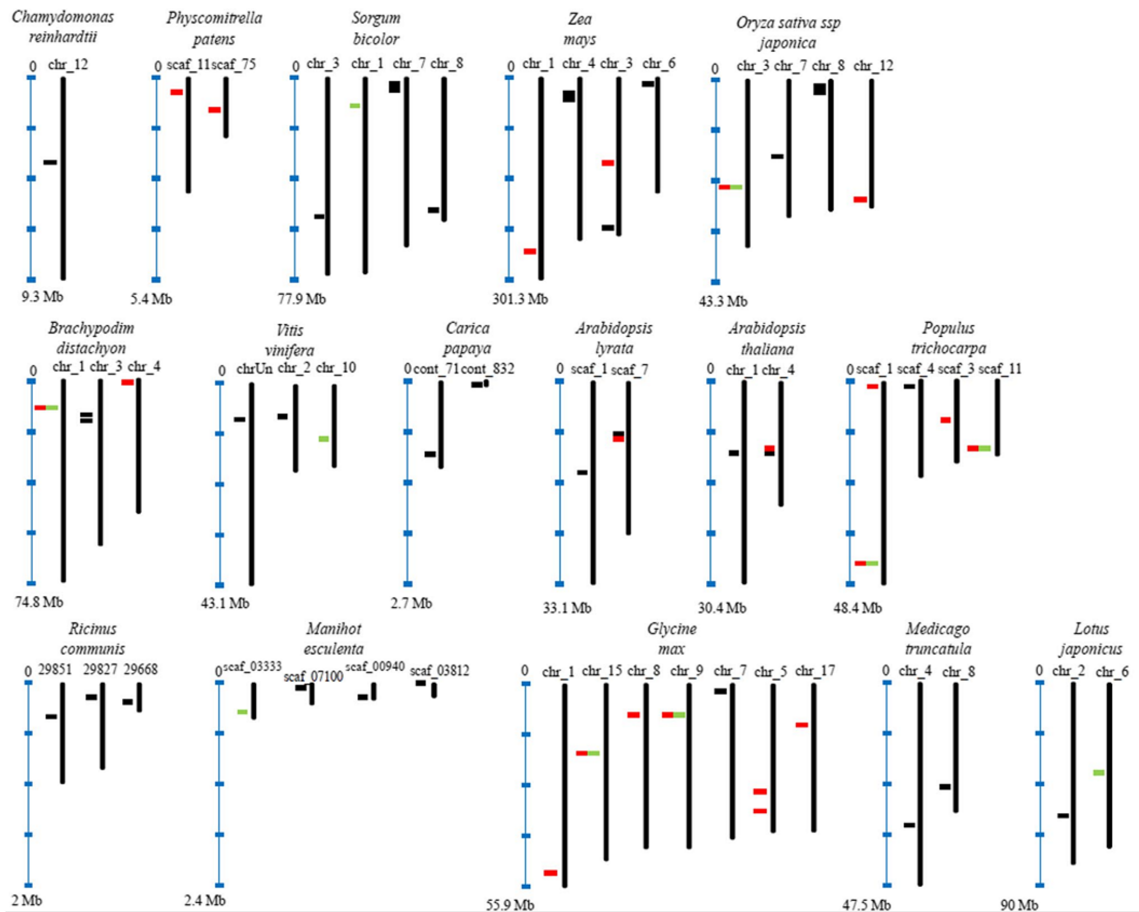


Fig. 3 Analysis of chromosomal location and the mechanism of *LSD* gene generation in the available species using the WGmapping tool in the PLAZA database. For each species, the individual scale in Mb is shown. The *black vertical line* represents the chromosomes, scaffolds, or contigs and shows their number at the *top*. The *black*, *red*, and *green horizontal lines* represent the single-copy genes, the block, and the tandem-duplicated genes, respectively (color figure online)

similarity), *GmaLSD4* and *PtrLSD1* (75.3 % protein similarity), *GmaLSD4* and *PtrLSD2* (73.0 % protein similarity), *GmaLSD5* and *PtrLSD3* (82.8 % protein similarity), and *GmaLSD5* and *PtrLSD5* (82.1 % protein similarity). A large duplication block was observed in the syntenic region that contains *GmaLSD6* and *PtrLSD3* (75.3 % protein similarity) and *GmaLSD7* and *PtrLSD3* (77.6 % protein similarity, data not shown). A huge duplication block was observed in the syntenic regions that possess *GmaLSD6* and *PtrLSD5* (75.3 % protein similarity), *GmaLSD7* and *PtrLSD5* (77.6 % protein similarity), and *GmaLSD8* and *PtrLSD4* (58.6 % protein similarity).

Figure 5 shows the syntenic regions between LSD paralogous genes in soybean and poplar. In accordance with the syntenic regions between the orthologous genes, we observed a high degree of conservation in the protein families. For the majority of the genes, the *Ka*/*Ks* ratio also suggested negative selection (Supplementary Table S2). Two exceptions were observed: in the syntenic region that contains *GmaLSD3* and *GmaLSD4* and in the region that contains Glyma09g09990 and Glyma15g22100 (nitrate, fromate, and iron dehydrogenase family). These exceptions could suggest neutral selection. The same pattern was observed for Glyma01g40750 and Glyma17g16560 (E3 ubiquitin ligase family) in the syntenic region that contains *GmaLSD6* and *GmaLSD7*.

We verified two different types of duplication block in LSD paralogous genes (Supplementary Table S2). A huge duplication block was found between *GmaLSD1* and *GmaLSD2* (99.4 % protein similarity), *GmaLSD3* and *GmaLSD4* (87.8 % protein similarity), *GmaLSD5* and *GmaLSD6* (86.9 % protein similarity), *GmaLSD6* and *GmaLSD7* (78.9 % protein similarity), and *PtrLSD3* and

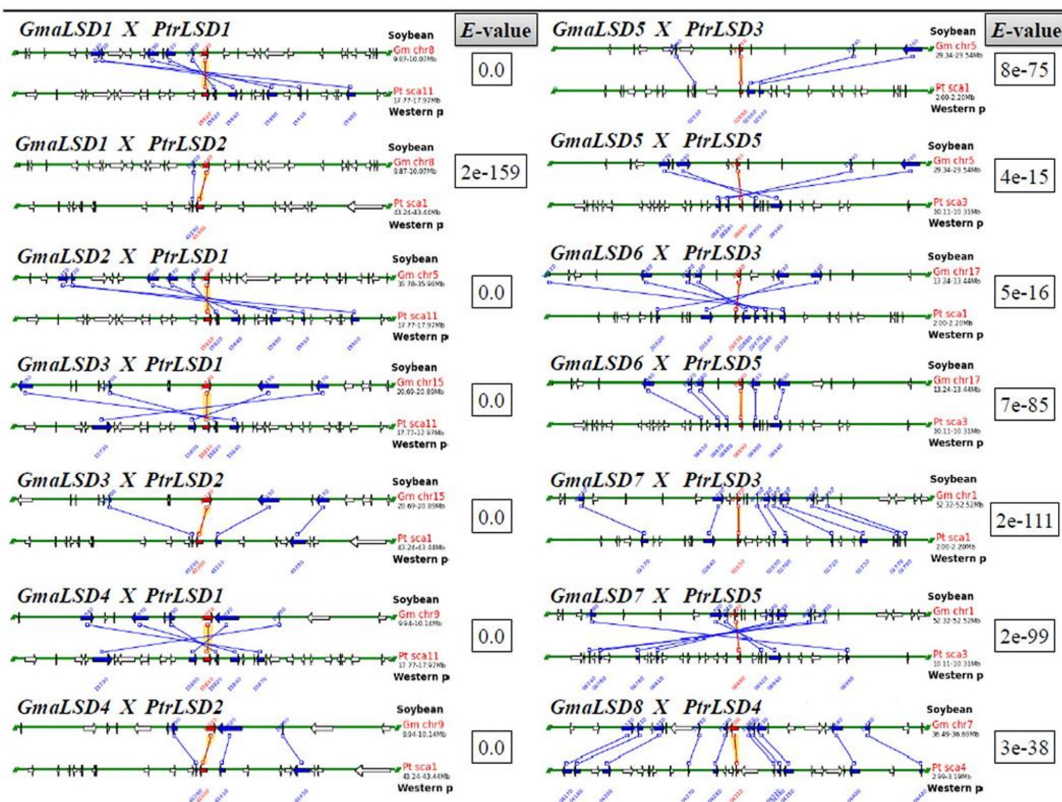


Fig. 4 Syntenic regions between *GmaLSD* and *PtrLSD* orthologous genes analyzed in the PGDD database. The *green horizontal line* represents the chromosome. The *blue* and *red vertical line* and *arrows* represent the duplicated orthologous genes; the *red line* represents the sequence that was used as the search query (indicated at the *top* of each of the syntenic regions). The *white arrow* indicates that there were no duplicated genes. The gene code that was used in the PGDD is located next to the *arrows*. The *e* value for each syntenic region is shown (color figure online)

PtrLSD5 (78.2 % protein similarity). A large duplication block was observed in the syntenic region that contains *GmaLSD5* and *GmaLSD7* (80 % protein similarity) and *PtrLSD1* and *PtrLSD2* (88.3 % protein similarity).

We observed that the number of LSD domain was maintained between paralogous/orthologous genes (Figs. 4, 5). Therefore, the syntenic regions showed that duplicated LSD genes originate proteins with equal structure of domains.

Discussion

LSD gene diversification occurred before the separation of the monocots and eudicots

The phylogenetic analysis performed in this study shows that the distribution of the *LSD* genes among the Viridiplantae species reflects the number of LSD domains that they encode and that this distribution is well supported (Fig. 1). The proposed scenario based on the unrooted tree suggests that all genes evolved during the same time and emphasizes the idea that the common ancestor of each of these genes preceded the separation of monocots and eudicots. Additionally, the incongruence observed in the localization of Rco2 (with two LSD domains) could be explained by the finding that while other proteins with two LSD domains comprise the second and third domains (i.e., the core and *C*-terminal domains), Rco2 is exclusively formed by 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 is closer to proteins that show three LSD domains.

Structural divergence between exon/intron within gene families is also a mechanism for their evolution (Li et al. 2009). The differences observed in the exon–intron

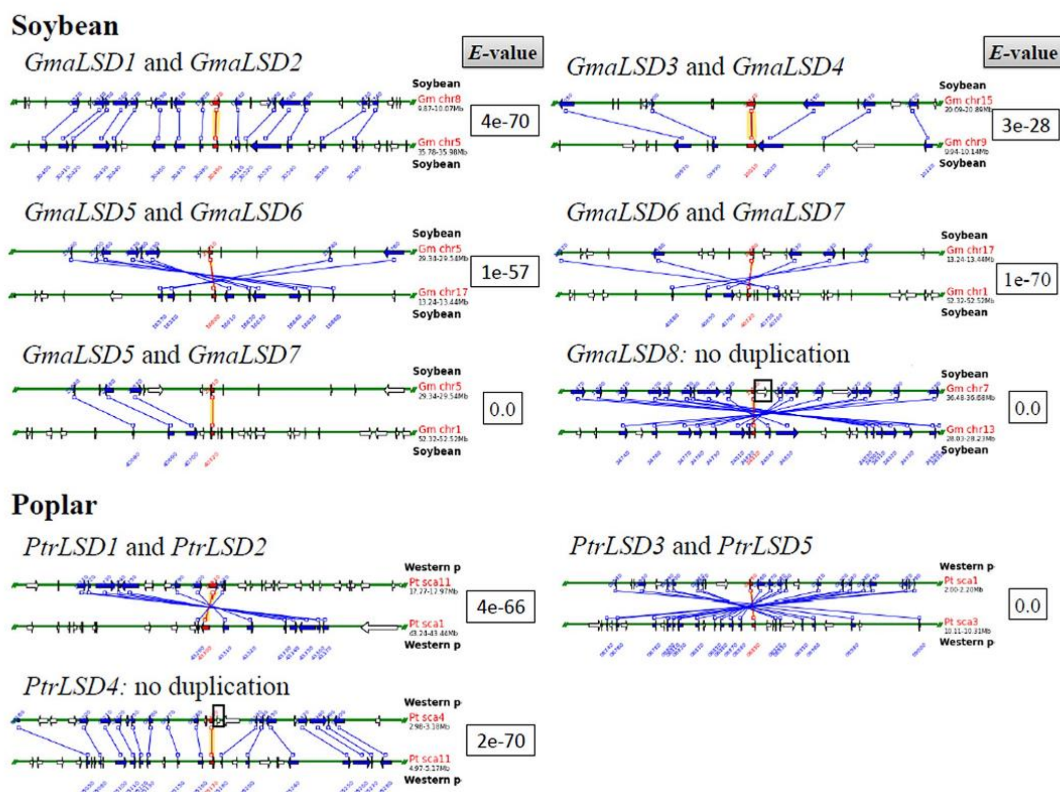


Fig. 5 Syntenic regions between the *GmaLSD* and *PtrLSD* paralogous genes analyzed in the PGDD database. The *green horizontal line* represents the chromosome. The *blue* and *red vertical line* and *arrows* represent the duplicated paralogous genes; the *red line* represents the sequence that was used as the search query (indicated at the *top* of each of the syntenic regions). To illustrate the lack of duplication of *GmaLSD8* and *PtrLSD4,* Glyma07g31590 and Ptrichoc0004s04280 were used as query sequences. The *white arrow* indicates that there were duplicated genes. The gene code that was used in PGDD is located next to the *arrows*. The *e* value for each syntenic region is shown. The *black squares* represent the lack of duplicated *GmaLSD8* and *PtrLSD4* genes (color figure online)

structure of paralogous/orthologous genes could be explained by several mechanisms, such as exon/intron gain/loss, point mutation, insertion/deletion, exonization (a process in which an intronic or intergenic sequence becomes exonic), and pseudoexonization (the opposite process of exonization) in a DNA sequence (Xu et al. 2011). Although related paralogous/orthologous genes contain exon/intron that is conserved, events of exon/intron modifications might have occurred during the evolution of the LSD family. We observed modifications in the number of intron to specific genes, as observed in the subcluster that is contained in subclade E (between the Tha2 and Gma6 genes) (Fig. 1). The genes that belong to this subcluster contain five introns, except Mtr2, Gma7, and Gma6.

Genes that showed modifications in their intron number had also alterations in their intron phase. This pattern is remarkable in the subclade that is formed between the Tha2 and Gma6 genes (contained in subclade E). In these, Mtr2 and Gma7 showed modifications, but remained clustered with related genes. We conclude that, despite the variation, they still resemble their corresponding paralogous/orthologous genes.

In summary, we verify that the analysis of intron/exon structure of LSD genes corroborates the phylogenetic analysis. Therefore, the combination of these results suggests that all the LSD genes might have a common ancestor.

β-sheet composition in *LSD* proteins is maintained independent of the number of *LSD* domains

Protein secondary structure carries information about the structural arrangements that occur in a protein. To date,

data involving the proteins of the *LSD* family are associated with protein–protein interactions (Coll et al. 2010; He et al. 2011b; Kaminaka et al. 2006; Li et al. 2013), the role of the zinc finger LSD domains (He et al. 2011a), and their conservation in protein sequences (Cabreira et al. 2013). These studies revealed that zinc finger LSD domains are fundamental for the functionality of these proteins.

β-sheets are formed by the pairing of multiple β-strands that are held together by typical patterns of hydrogen bonds that run in parallel or antiparallel (essential to maintain protein stability), which involve interactions between residues that are often separated by large distances along the primary sequence of proteins (Cheng and Baldi 2005). We showed that β-sheets are the main component in the assembly of these LSD proteins (Fig. 2; see also Supplementary Material 2 and Supplementary Material 3 to an overview of these results). Three sequential β-sheets (indicated by the yellow arrow) generally overlap the LSD domain position (indicated by the orange line) (Fig. 2). Regardless of the number of LSD domains (exemplified as *Gm*LSD1, *Gm*LSD8, and *Osa*LSD5), the structures of β-sheets are maintained. Thus, the high conservation of LSD domains (Cabreira et al. 2013) and β-sheets that overlap with their positions points to a possible similar protein secondary structure in all the LSD family. The results of previous studies and the present data suggest an important role for LSD domains in LSD protein structure and activity.

The majority of *LSD* genes are single-copy genes

In gene families, gene duplication events are fundamental for the families' evolution because duplicated genes provide the raw materials for the generation of new gene functions (Yin et al. 2013). Duplication of genomic content can occur by many independent mechanisms, such as tandem duplication (local duplications that involve one or two genes), block/segmental duplications (duplications of subchromosomal-length regions), and whole-genome duplications (WGD) originated by polyploidy events (Flagel and Wendel 2009).

The analysis of the distribution of *LSD* genes reveals that they are located on various chromosomes, and several members were found on duplicated chromosomal segments (Fig. 3). Similar results were observed for the annexin gene family in Viridiplantae (Jami et al. 2012).

Block duplication process is more frequent in plants because most of them are polyploids, and numerous duplicated chromosomal blocks are maintained within their genomes (Yin et al. 2013). Several plant species, such as *A. thaliana* (Simillion et al. 2002), rice (Yu et al. 2005), and soybean (Schmutz et al. 2010), experienced WGD during their evolution. This phenomenon is particularly prominent in eudicot plants, which share a γ event (Lin and Paterson 2011).

The genes originated by the tandem duplication process do not present a matching paralog (Fig. 3; Supplementary Table S1), suggesting that the duplicated copies are quickly lost. The exceptions to this scenario are the tandem-duplicated genes in poplar (*Ptr*LSD1 and *Ptr*LSD2) and soybean (*Gma*LSD3 and *Gma*LSD4), and species that had undergone genome duplications during their evolutionary histories (Lin and Paterson 2011). Therefore, we suggested that the majority of the LSD genes are single-copy genes that were duplicated, but maintained throughout evolution in an ancestral genome. These events resulted in a current scenario that consists of a large number of orthologous LSD genes.

A common origin of *LSD* genes is particularly interesting for sequences that possess one and two LSD domains. Although certain sequences with three LSD domains are subject to duplications, the sequences with one or two LSD domains do not show duplicated genes (Supplementary Table S1). Synteny analysis of *Gma*LSD8 and *Ptr*LSD4 (both of which present two LSD domains) supports this result (Fig. 5). These genes are located in a very large duplication block and do not have matching paralogs. This result shows that they are descendants from an ancestral genome and that their duplication did not occur after the split between these genomes or that the duplicated copy was lost during evolution.

Synteny analysis demonstrates high conservation between the related species poplar (*Populus trichocarpa*) and soybean (*Glycine max*)

Synteny is an inference that two or more chromosomes or segments are derived from a common ancestor (Lyons et al. 2008). The analysis of physical synteny may be used to explain the events that a genome has undergone prior to achieving its current structural form. These genomic events can include the evolutionary repositioning of genes responsible for shared phenotypes between two related species (McClean et al. 2010).

Although several breakpoints occur in the syntenic regions between poplar and soybean, a significant synteny is still maintained (Fig. 4). These breakpoints can be explained by the different number of duplications in these species. Thus, syntenic regions containing *LSD* genes identified as intraspecies (*GmaLSD* × *GmaLSD* and *PtrLSD* × *PtrLSD*) and interspecies (*GmaLSD* × *PtrLSD*) emphasize that the *LSD* genes originated from a common ancestor.

LSD genes with the same number of LSD domains formed syntenic regions (Figs. 4, 5). As exemplified by soybean and poplar, this observation was confirmed in several species such as *M. esculenta*, *R. communis*, *C. papaya,* and *S. bicolor* (data not shown). Therefore, these results show that the *LSD* genes that encode proteins with one,

two, and three LSD domains are not the result of loss or gain in LSD domains. We conclude that the duplicated copies retain the original number of domains.

Although duplicated genes can undergo neofunctionalization (when one copy acquires a novel function) or subfunctionalization (in which both copies are mutated and adopt complementary functions), pseudogenization is the most common fate of the copies after the duplication process (Cagliari et al. 2011). The *Ka/Ks* ratio generates insights into the processes that drive the changes at the molecular level. A *Ka/Ks* ratio >1 shows that neofunctionalization through positive selection acted during the sequence divergence, and a *Ka/Ks* ratio <1 suggests that subfunctionalization through negative selection operated during the divergence (Fan et al. 2013; Zhang et al. 2002).

The results for the homologous *Rxp* regions in soybean suggest that smaller *Ks* values (0.16, 0.17) are associated with recent duplication regions and that high *Ks* values (0.62–0.67) are associated with ancient polyploidy events (Kim et al. 2009). Based on these data, we observed that the syntenic regions with *GmaLSD1* and *GmaLSD2*, *GmLSD3* and *GmaLSD4*, and *GmaLSD5* and *GmaLSD6* can be associated with recent duplication regions (low *Ks* values). Moreover, the syntenic regions with *GmaLSD6* and *GmaLSD7* and *GmaLSD5* and *GmaLSD7* generally exhibit high *Ks* values, suggesting an association with an ancient polyploidy event. Therefore, these results suggest that the *GmaLSD* genes could have been derived from different duplication events.

Because gene duplication makes a major contribution to the evolution of new gene functions, duplicated genes can experience a substantial relaxation of selection compared with unduplicated genes (Kondrashov et al. 2002). An evolutionary model of gene duplication proposed by Kondrashov et al. (2002) holds that an apparent early phase of relaxed constraint or even near-neutrality that is followed by a negative selection period occurs at the first stages after gene duplication. The model proposed that the later stages of this process most likely provide a long-term advantage by enabling the creation of new functions. In terms of the syntenic regions analyzed, the *Ka/Ks* ratio shows that both paralogous and orthologous sequences appear to be evolving under similar levels of negative selection (Supplementary Table S2). This finding indicates that a subfunctionalization process probably could be occurring in LSD genes. In fact, we observed different expression patterns of *GmaLSD* genes in various organs and plants subjected to *Phakopsora pachyrhizi* infection and dehydration (Cabreira et al. 2013). Thus, the differential expression together with the low *Ka/Ks* supports a possible subfunctionalization of these genes. Previous studies have shown that the prevalence of variable expression in soybean after polyploidization has revealed that several paralogous genes undergo subfunctionalization (Fan et al. 2013; Roulin et al. 2013). Accordingly, our results reinforce the theory of the retention of duplicate genes through subfunctionalization, which may constitute a transitional step to neofunctionalization (Roulin et al. 2013).

Figure 6 summarizes and provides a comprehensive overview of the results. In summary, we state that the diversification of the *LSD* family precedes the separation of monocots and eudicots. We observed that the different number of LSD domains does not alter the protein secondary structure, which suggests a similar conformation in all LSD proteins. Proteins that are composed of one, two, and three domains are not a result of a loss or gain in LSD domains. Duplicated LSD genes generated copies with equal numbers of LSD domains. The majority of the *LSD* genes are maintained as a single copy in the different species and are orthologous in several genomes. Together, these results suggest the existence of a common ancestor of the LSD proteins. Finally, although the majority of members are related to the negative regulation of PCD, LSD members probably could be involved in other

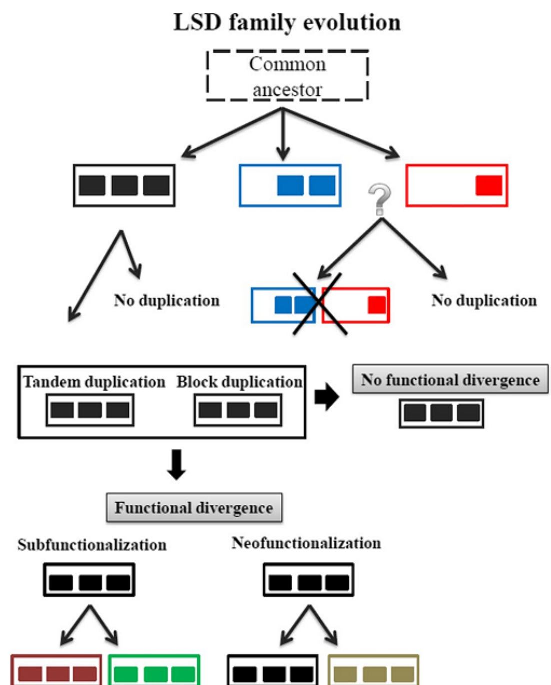


Fig. 6 Model of possible LSD family evolution. The *rectangles* represent genes encoding proteins with three LSD domains (*black*), two LSD domains (*blue*), and one LSD domain (*red*). The *dark pink* and *green rectangles* represent the mutated copy with complementary functions. The *black rectangle* shows the copy that keeps the original function, and the *brown rectangle* indicates the copy that acquires a novel function (color figure online)

biological processes, due the occurrence of functionalization processes.

Acknowledgment This study was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul (PPGBM/UFRGS), GENOSOJA/CNPq, and BIOTECSUR II/MCT.

Conflict of interest The authors declare that they have no conflicts of interest.

Compliance with ethical standards This article does not contain any studies with human participants or animals performed by any of the authors.

References

Atkinson NJ, Urwin PE (2012) The interaction of the plant biotic and abiotic stresses: from genes to the field. J Exp Bot 63:3523–3543

Bhatti KH, Xu C, Wu J, He C (2008) Overexpression of rice OsLOL2 gene confers disease resistance in tobacco to *Pseudomonas syringae* pv. tabaci. Prog Nat Sci 18:807–812

Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K, Jones DT (2010) Protein annotation and modelling servers at University College London. Nucleic Acids Res 38:W563–W568

Cabreira C, Cagliari A, Bucker-Neto L, Wiebke-Strohm B, de Freitas LB, Marcelino-Guimaraes FC, Nepomuceno AL, Margis-Pinheiro MM, Bodanese-Zanettini MH (2013) The *Lesion Simulating Disease* (LSD) gene family as a variable in soybean response to *Phakopsora pachyrhizi* infection and dehydration. Funct Integr Genomics 13(3):323–338

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 F (ed) Gene duplication, vol 1. InTech, Rijeka, pp 1–27

Cheng J, Baldi P (2005) Three-stage prediction of protein beta-sheets by neural networks, alignments and graph algorithms. Bioinformatics 21(Suppl 1):i75–i84

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(6009):1393–1397

Dietrich RA, Delaney TP, Uknes SJ, Ward ER, Ryals JA, Dangl JL (1994) *Arabidopsis* mutants simulating disease resistance response. Cell 77(4):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(5):685–694

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

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(11):6831–6836

Fan C, Wang X, Hu R, Wang Y, Xiao C, Jiang Y, Zhang X, Zheng C, Fu YF (2013) The pattern of Phosphate transporter 1 genes evolutionary divergence in *Glycine max* L. BMC Plant Biol 13:48

Flagel LE, Wendel JF (2009) Gene duplication and evolutionary novelty in plants. New Phytol 183(3):557–564

Guo AY, Zhu QH, Chen X, Luo JC (2007) GSDS: a gene structure display server. Yi Chuan 29(8):1023–1026

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(7):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(4):e18750

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(2):301–312

Jabs T, Dietrich RA, Dangl JL (1996) Initiation of runaway cell death in an Arabidopsis mutant by extracellular superoxide. Science 273(5283):1853–1856

Jami SK, Clark GB, Ayele BT, Ashe P, Kirti PB (2012) Genome-wide comparative analysis of annexin superfamily in plants. PLoS One 7(11):e47801

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(18):4400–4411

Kim KD, Shin JH, Van K, Kim DH, Lee SH (2009) Dynamic rearrangements determine genome organization and useful traits in soybean. Plant Physiol 151(3):1066–1076

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(11):1022–1026

Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV (2002) Selection in the evolution of gene duplications. Genome Biol 3(2):RESEARCH0008

Lee TH, Tang H, Wang X, Paterson AH (2013) PGDD: a database of gene and genome duplication in plants. Nucleic Acids Res 41:D1152–D1158

Li W, Liu B, Yu L, Feng D, Wang H, Wang J (2009) Phylogenetic analysis, structural evolution and functional divergence of the 12-oxo-phytodienoate acid reductase gene family in plants. BMC Evol Biol 9:90

Li Y, Chen L, Mu J, Zuo J (2013) *Lesion Simulating Disease1* interacts with catalases to regulate hypersensitive cell death in *Arabidopsis*. Plant Physiol 163(2):1059–1070

Lin L, Paterson AH (2011) Size variation in homologous segments across divergent plant genomes. Mob Genet Elements 1(2):92–96

Long M, de Souza SJ, Gilbert W (1995) Intron phase correlations and the evolution of intron/exon structure of genes. Proc Natl Acad Sci USA 92:1249512499

Lyons E, Pedersen B, Kane J, Alam M, Ming R, Tang H, Wang X, Bowers J, Paterson A, Lisch DMF (2008) Finding and comparing syntenic regions among *Arabidopsis* and the Outgroups Papaya, Poplar, and Grape: CoGe with Rosids. Bioinformatics 148:1772–1781

Mateo A, Muhlenbock P, Rusterucci C, Chang CC, Miszalski Z, Karpinska B, Parker JE, Mullineaux PM, Karpinski S (2004) *Lesion Simulating Disease1* is required for acclimation to conditions that promote excess excitation energy. Plant Physiol 136(1):2818–2830

McClean PE, Mamidi S, McConnell M, Chikara S, Lee R (2010) Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. BMC Genom 11:184

McGuffin LJ, Bryson K, Jones D (2000) The PSIPRED protein structure prediction server. Bioinformatics 16(4):404–405

Muhlenbock P, Plaszczyca M, Mellerowicz E, Karpinski S (2007) Lysigenous aerenchyma formation in *Arabidopsis* is controlled by *Lesion Simulating Disease1*. Plant Cell 19(11):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(9):2339–2356

Mur LA, Kenton P, Lloyd AJ, Ougham H, Prats E (2008) The hypersensitive response; the centenary is upon us but how much do we know? J Exp Bot 59(3):501–520

Roulin A, Auer PL, Libault M, Schlueter J, Farmer A, May G, Stacey G, Doerge RW, Jackson SA (2013) The fate of duplicated genes in a polyploid plant genome. Plant J 73:143–153

Rusterucci C, Aviv DH, Holt BF III, Dangl JL, Parker JE (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(10):2211–2224

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(7278):178–183

Sharp PA (1981) Speculations on RNA splicing. Cell 23:643–646

Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, Van de Peer Y (2002) The hidden duplication past of *Arabidopsis thaliana*. Proc Natl Acad Sci USA 99(21):13627–13632

Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic analyses. Syst Biol 49(2):369–381

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

van Doorn WG, Beers EP, Dangl JL, Franklin-Tong VE, Gallois P, Hara-Nishimura I, Jones AM, Kawai-Yamada M, Lam E, Mundy J, Mur LAJ, Petersen M, Smertenko A, Taliansky M, Van Breusegem F, Wolpert T, Woltering E, Zhivotovsky B, Bozhkov PV (2011) Morphological classification of plant cell deaths. Cell Death Differ 18:1241–1246

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(5):375–384

Wituszynska W, Slesak I, Vanderauwera S, Szechynska-Hebda M, Kornas A, Van Der Kelen K, Muhlenbock P, Karpinska B, Mackowski S, Van Breusegem F, Karpinski S (2013) *Lesion Simulating Disease1*, enhanced disease susceptibility1, and phytoalexin deficient 4 conditionally regulate cellular signaling homeostasis, photosynthesis, water use efficiency, and seed yield in *Arabidopsis*. Plant Physiol 161(4):1795–1805

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(1):85–94

Xu G, Guo C, Shan H, Kong H (2011) Divergence of duplicate genes in exon-intron structure. Proc Natl Acad Sci USA 109(4):187–1192

Yin G, Xu H, Xiao S, Qin Y, Li Y, Yan Y, Hu Y (2013) The large soybean (*Glycine max*) WRKY TF family expanded by segmental duplication events and subsequent divergent selection among subgroups. BMC Plant Biol 13(1):148

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, Zhang J, Zhang Y, Li R, Xu Z, Li X, Zheng H, Cong L, Lin L, Yin J, Geng J, Li G, Shi J, Liu J, Lv H, Li J, Deng Y, Ran L, Shi X, Wang X, Wu Q, Li C, Ren X, Li D, Liu D, Zhang X, Ji Z, Zhao W, Sun Y, Zhang Z, Bao J, Han Y, Dong L, Ji J, Chen P, Wu S, Xiao Y, Bu D, Tan J, Yang L, Ye C, Xu J, Zhou Y, Yu Y, Zhang B, Zhuang S, Wei H, Liu B, Lei M, Yu H, Li Y, Xu H, Wei S, He X, Fang L, Huang X, Su Z, Tong W, Tong Z, Ye J, Wang L, Lei T, Chen C, Chen H, Huang H, Zhang F, Li N, Zhao C, Huang Y, Li L, Xi Y, Qi Q, Li W, Hu W, Tian X, Jiao Y, Liang X, Jin J, Gao L, Zheng W, Hao B, Liu S, Wang W, Yuan L, Cao M, McDermott J, Samudrala R, Wong GK, Yang H (2005) The genomes of *Oryza sativa*: a history of duplications. PLoS Biol 3(2):e38

Zhang L, Vision TJ, Gaut BS (2002) Patterns of nucleotide substitution among simultaneously duplicated gene pairs in *Arabidopsis thaliana*. Mol Biol Evol 19(9):464–1473

CAPÍTULO 2

Funct Integr Genomics
DOI 10.1007/s10142-015-0459-7

REVIEW

Caspases in plants: metacaspase gene family in plant stress responses

David Fagundes[1] · Bianca Bohn[1] · Caroline Cabreira[2] · Fábio Leipelt[1] · Nathalia Dias[1] · Maria H. Bodanese-Zanettini[2] · Alexandro Cagliari[1]

Received: 22 April 2015 / Revised: 22 July 2015 / Accepted: 24 July 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Programmed cell death (PCD) is an ordered cell suicide that removes unwanted or damaged cells, playing a role in defense to environmental stresses and pathogen invasion. PCD is component of the life cycle of plants, occurring throughout development from embryogenesis to the death. Metacaspases are cysteine proteases present in plants, fungi, and protists. In certain plant–pathogen interactions, the PCD seems to be mediated by metacaspases. We adopted a comparative genomic approach to identify genes coding for the metacaspases in Viridiplantae. We observed that the metacaspase was divided into types I and II, based on their protein structure. The type I has a metacaspase domain at the *C*-terminus region, presenting or not a zinc finger motif in the *N*-terminus region and a prodomain rich in proline. Metacaspase type II does not feature the prodomain and the zinc finger, but has a linker between caspase-like catalytic domains of 20 kDa (p20) and 10 kDa (p10). A high conservation was observed in the zinc finger domain (type I proteins) and in p20 and p10 subunits (types I and II proteins). The phylogeny showed that the metacaspases are divided into three principal groups: type I with and without zinc finger domain and type II metacaspases. The algae and moss are presented as outgroup, suggesting that these three classes of metacaspases originated in the early stages of Viridiplantae, being the absence of the zinc finger domain the ancient condition. The study of metacaspase can clarify their assignment and involvement in plant PCD mechanisms.

Keywords Programmed cell death · Metacaspase gene family · Biotic and abiotic stress responses

David Fagundes, Bianca Bohn and Caroline Cabreira contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10142-015-0459-7) contains supplementary material, which is available to authorized users.

✉ Alexandro Cagliari
alexandro-cagliari@uergs.edu.br

David Fagundes
dsanttos007@hotmail.com

Bianca Bohn
biabohn@hotmail.com

Caroline Cabreira
carol.cabreira@yahoo.com.br

Fábio Leipelt
fabioleipelt@hotmail.com

Nathalia Dias
nat112006@yahoo.com.br

Maria H. Bodanese-Zanettini
mhbzanettini@yahoo.com.br

[1] Universidade Estadual do Rio Grande do Sul (UERGS), CEP 96816-50 Santa Cruz do Sul, RS, Brazil

[2] Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

Published online: 16 August 2015

Programmed cell death (PCD)

Programmed cell death (PCD) is a genetically programmed and highly ordered cell suicide process that removes unwanted or damaged cells from both eukaryotic and prokaryotic organisms. PCD is recognized as a key mechanism in organ development and in tissue homeostasis (De Pinto et al. 2012). PCD also plays a major role in defense responses against certain kinds of environmental stresses, development and survival against pathogen invasion, and other external stimuli. PCD acts to remove any mutated, infected, or damaged cells from surrounding healthy tissue in both plant and animal systems (Baskett 2012; Greenberg and Yao 2004; Gunawardena 2008; Lord and Gunawardena 2012; Pennell and Lamb 1997; Potten and Wilson 2004; Ranganath and Nagashree 2001; Rogers 2005; Sanmartin et al. 2005; Zhang et al. 2013).

The term PCD was used for the first time to describe a cell death process controlling tadpole tail development (Tata 1966). PCD-like processes have been reported in several unicellular groups, suggesting that some components of the PCD machinery existed early in the evolution of the eukaryotic lineage (Nedelcu 2009).

PCD shares several morphological and biochemical features with apoptosis in animals (Greenberg 1996; Levine et al. 1996; Woltering et al. 2002). Therefore, it is thought that some of the components of the cell death machinery are conserved in both apoptosis and PCD (Baskett 2012). However, unlike apoptosis in animal cells, relatively little is known about PCD in plants, even though its role in plant development and physiology is well established (Beers and McDowell 2001; Coll et al. 2011; Jones 2001; Lam 2004; Love et al. 2008; Williams and Dickman 2008). The lack of plant PCD knowledge has led to speculative comparisons with animal cell death (Dickman and Reed 2004; Lord and Gunawardena 2012; Ranganath and Nagashree 2001). However, plant PCD is not considered to be synonymous with apoptosis since different forms of plant PCD have been found to exhibit only some of the morphological and molecular hallmarks of apoptosis (Cacas 2010; Danon et al. 2004; De Pinto et al. 2012). For instance, in plant PCD in cell culture models, cell condensation has been observed, but it is not accompanied by cell fragmentation and formation of the cellular apoptotic bodies that characterize apoptosis. Morphologically, a key difference between the PCD of plants and apoptosis in animals is the absence of engulfment by neighboring cells in plants (Lam 2004). These distinctive features observed are likely due to the presence of the cell wall and the absence of phagocytes in plants (De Pinto et al. 2012).

PCD in animals requires the cysteine-dependent aspartyl protease activity of caspase enzymes. Caspases represent a family of proteases, which have cysteine in their active site and are able to induce cleavage in correspondence to aspartate residues (De Pinto et al. 2012; Thornberry and Lazebnik 1998). Despite their importance in metazoans and the ubiquitous occurrence of PCD in eukaryotes, caspases are not conserved in the genomes of yeasts, protozoans, and plants (Lam and Zhang 2012). However, caspase-like enzymatic activity has been found to be essential in many forms of plant PCD (Baskett 2012; del Pozo and Lam 1998; Woltering et al. 2002). In this scenario, metacaspases have emerged as the best candidates to perform this role, because they contain a caspase-specific catalytic dyad of histidine and cysteine, as well as a conserved caspase-like secondary structure (Suarez et al. 2004).

PCD in plants

The current classification of PCD in plants was based on morphological criteria, being divided into two broad categories: vacuolar cell death and necrosis (Lord and Gunawardena 2012; van Doorn et al. 2011). Vacuolar cell death is characterized by the removal of cell content through a combination of autophagy-like processes and the release of hydrolases from collapsed vacuoles. The early rupture of the plasmatic membrane, shrinkage of the protoplast, and absence of vacuolar cell death features characterizes necrotic cell death (Lord and Gunawardena 2012).

PCD is an essential part of the life cycle of plants, occurring throughout plant development from embryogenesis to the death of the whole plant (Lord and Gunawardena 2012; Rogers 2005; Staal and Dixelius 2007). From seed germination until seed production, developmental cell death is manifested (De Pinto et al. 2012).

Within plant systems, PCD falls within two broad categories: environmentally induced and developmentally regulated (Coll et al. 2010). The induction of PCD is required for many growth and developmental processes, such as the development of endosperm and aleurone cells in cereals or seed storage tissues (Fath et al. 2000; Lombardi et al. 2010; Young and Gallie 2000), differentiation of tracheary elements (Fukuda 2000; Kwon et al. 2010), female gametophyte differentiation (Wu and Cheun 2000), leaf abscission and whole plant senescence (De Pinto et al. 2012; Lee et al. 2007a, b), deletion of the embryonic suspensor (Bozhkov et al. 2005a, b; Giuliani et al. 2002; Rogers 2005), and anther dehiscence (Rogers 2005; Senatore et al. 2009; Wilson et al. 2011). PCD is also required for xylem differentiation (Fukuda 1997), leaf senescence (De Michele et al. 2009; Lim et al. 2007), flower senescence (Rogers 2006), and leaf morphogenesis (Gunawardena et al. 2004; Lord and Gunawardena 2011; Lord and Gunawardena 2012; Wright et al. 2009), leaf and flower senescence, elimination of reproductive organs in unisexual flowers, pollen rejection in the self-incompatibility response, fruit dehiscence, and pod shattering (van Doorn and Woltering 2005; Vercammen et al. 2007).

In contrast, environmentally induced PCD is mostly a result of external factors including heat shock (Balk et al. 2003; Lord and Gunawardena 2011; Zhang et al. 2009), pathogen infection (Lam et al. 2001; Mur et al. 2008), and oxygen treatment (Gunawardena et al. 2001).

Metacaspases in viridiplantae

True caspases have not yet been described in plants (Lord and Gunawardena 2012). Nevertheless, two broad groups of caspase-like proteases have been identified in plant systems: cysteine endopeptidases (Rojo et al. 2004) and serine endopeptidases (Coffeen and Wolpert 2004).

Cysteine endopeptidases are further divided into two groups, vacuolar processing enzymes (Sanmartin et al. 2005) and metacaspases (Bozhkov and Jansson 2007; Bozhkov et al. 2010; Lord and Gunawardena 2012; Sanmartin et al. 2005).

Metacaspase is a family of cysteine proteases described in plants (Vercammen et al. 2004), fungi (Madeo et al. 2002), and protists (Lee et al. 2007a, b; Mottram et al. 2003; Szallies et al. 2002) based on homology with caspase-like domains (Rezanezhad et al. 2011; Uren et al. 2000). Instead of cleaving substrate with D residue at P1 position (referred to the *N*-terminus direction from the cleaved bond) as caspases in animals, metacaspase enzymes prefer R or K at the cleavage site (Lee et al. 2008; Ojha et al. 2010; Vercammen et al. 2007; Watanabe and Lam 2005). Together with the eukaryotic caspases, legumains, paracaspases, separases, and the bacterial clostripains and gingipains, they belong to the clan CD of cysteine proteases. Clan CD includes organisms that utilize a catalytic His-Cys dyad for their activity (Vercammen et al. 2006).

Plant metacaspases are classified into type I and type II based on overall structure and the level of sequence similarity (Baskett 2012; Trzyna et al. 2008; Uren et al. 2000). Both type I and type II metacaspases have a putative conserved caspase-like catalytic domain composed of 20 kDa (p20) and 10 kDa (p10) subunits (Ojha et al. 2010), which contain the catalytic amino acid dyad histidine/cysteine. The catalytic histidine lies in the H(Y/F) SGHG sequence and the catalytic cysteine in the active-site pentapeptide DXCHS (where X is A or S) sequence (Suarez et al. 2004; Zhang et al. 2013).

Type I, but not type II, metacaspases exhibit an *N*-terminus extension that usually contains a zinc-finger motif as well as a proline-rich stretch and may or may not contain a glutamine-rich region (Lam and Zhang 2012).

Type II metacaspases have only been identified in plants and have no obvious prodomain, but harbor a linker region of 160–180 amino acids between the p20 and p10 subunits (Ojha et al. 2010). It is hypothesized that this longer linker may interfere with proper dimer formation (Lam and Zhang 2012).

It has been hypothesized that type I represents the ancient form of the metacaspase family and that the evolution of type II had occurred before the emergence of multicellular plants from their photosynthetic, unicellular ancestors (Lam and Zhang 2012). It is speculated that eukaryotic metacaspases originate possibly from a horizontal gene transfer between the mitochondrial endosymbionts (α-proteobacteria) and the early eukaryotes (Koonin and Aravind 2002). Moreover, metacaspase-like proteins are present not exclusively in α-proteobacteria but also in all Bacteria groups, such as cyanobacteria, the known ancestors of plant chloroplasts (Vercammen et al. 2007).

The distribution of the caspase-like protease family demonstrated that while caspases and paracaspases are, so far, limited to metazoans and *Dictyostelium*, respectively, metacaspases are highly conserved in plants and fungi. This distribution suggests that metacaspases are likely the most closely representative of the eukaryote ancestral protease (Uren et al. 2000).

The increasing availability of genome sequences from various eukaryotic groups provides an opportunity to explore the degree of conservation of gene families throughout evolution. We adopted a high-throughput comparative genomic approach to conduct a broad survey of fully sequenced Viridiplantae genomes, in order to identify the presence of homologous genes coding for the metacaspase gene family (Supplementary Table 1). Metacaspase gene and protein sequences were obtained through blast searches (blastp, blastx, and tblastx) against the Protein and Genome databases with the default parameters provided at the Phytozome Web site (http://www.phytozome.net). We used the nucleotide and protein sequences from the well-described metacaspase genes of the model plant *Arabidopsis thaliana* as queries for the blast searches.

Compared with the genomes of protozoa and fungi with a single or a few type I metacaspase genes and no type II metacaspase genes (Tsiatsiani et al. 2011), genomes of higher plants encode larger metacaspase families including both type I and type II members (Table 1).

The high number of metacaspases in plant genomes would indicate that members of this protease family might have evolved specialized functions after gene amplification (Lam and Zhang 2012). However, there is no evidence indicating either specialization or redundancy in the physiological functions of different metacaspase genes expressed in the same organism (Tsiatsiani et al. 2011).

We observed that the metacaspase genes identified could be divided into two classes, type I and type II, based on their protein structure (Fig. 1). The type I metacaspases could present or not present a prodomain rich in proline, were characterized by a zinc finger motif in the *N*-terminus region, and necessarily had a metacaspase domain at the *C*-terminus region. Metacaspase type II does not have the prodomain and the zinc finger motif, but has an insert (linker) between the p20 and p10 subunits (Fig. 1).

Table 1 Metacaspases divided into type I* (without the zinc-finger prodomain), type I (with the zinc-finger prodomain), or type II groups

Group	Species	Type I *	Type I	Type II	Total
Eudicot	*Manihot esculenta*	9	3	1	13
	Ricinus communis	6	2	1	9
	Linum usitatissimum	8	2	2	12
	Populus trichocarpa	5	7	2	14
	Medicago truncatula	5	0	2	7
	Phaseolus vulgaris	6	0	3	9
	Glycine max	9	2	5	16
	Cucumis sativus	2	3	1	6
	Prunus persica	4	2	1	7
	Malus domestica	6	2	1	9
	Fragaria vesca	3	2	1	6
	Arabidopsis thaliana	1	2	6	9
	Arabidopsis lyrata	3	1	6	10
	Capsella rubella	2	2	6	10
	Brassica rapa Chiifu-FPsc	9	3	8	20
	Thellungiella halophila	2	2	4	8
	Carica papaya	3	3	1	7
	Gossypium raimondii	13	3	1	17
	Theobroma cacao	6	3	1	10
	Citrus sinensis	6	2	1	9
	Citrus clementina	8	1	1	10
	Eucalyptus grandis	4	2	1	7
	Vitis vinifera	3	2	1	6
	Solanum tuberosum	3	3	1	7
	Solanum lycopersicum	4	3	1	8
	Mimulus guttatus v2.0	8	1	9	18
	Aquilegia coeruela	1	2	2	5
Monocot	*Sorghum bicolor v2.1*	4	2	7	13
	Zea mayz	7	3	3	13
	Setaria italica	5	2	4	11
	Panicum virgatum	7	2	7	16
	Panicum hallii v0.5	4	2	5	11
	Oryza sativa	4	1	3	8
	Brachypodium distachyon	4	2	4	10
Lycophyta	*Selaginella moelledorffii*	0	1	3	4
Bryophyta	*Physicomitrella patens v3.0 e v1.6*	2	1	4	7
Chlorophyta	*Chlamydomonas reinhardtii*	1	0	1	2
	Volvox carteri	1	0	1	2
	Coccomyxa subellipsoidea C-169	3	0	1	3
	Micromonas pusilla CCMP1545	0	0	1	1
	Micromonas pusilla RCC299	1	0	0	1
	Ostreococcus leucimarinus	1	0	0	1
	Total	183	76	113	372

Motif analysis using the default parameters of the MEME software (http://meme.nbcr.net/meme/) demonstrated the conservation degree of p20 and p10 subunits identified in both type I and type II metacaspase proteins (Fig. 2a, b). Motif analysis also demonstrated that the zinc finger domain is well conserved in all type I proteins (Fig. 3).

In order to generate new insights about the evolution of the metacaspase family in Viridiplantae, some representative

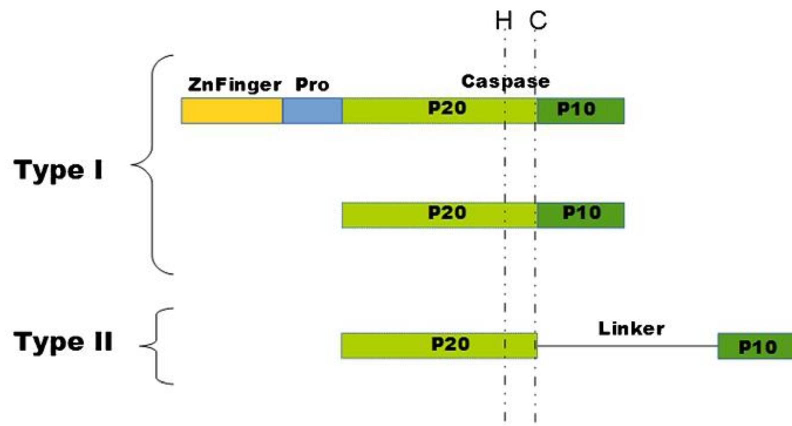


Fig. 1 Metacaspases are divided into two classes, type I and type II, based on their protein structure. The type I metacaspases could present or not present a prodomain rich in proline, include a zinc finger motif in the *N*-terminus region, and necessarily have a metacaspase domain at the *C*-terminus region. Type II metacaspases do not have the prodomain and the zinc finger motif, but feature an insert (linker) between the p20 and p10 subunits. The catalytic amino acid regions containing histidine (**h**) and cysteine (**c**) residues are shown as *dotted lines*

species of the identified genes were submitted for phylogenetic analysis. The phylogenetic analysis was reconstructed following protein sequence alignments using the conserved metacaspase domain from representatives of the identified proteins, through a Bayesian approach using Beast software (Drummond and Rambaut 2007). The mixed amino acid substitution model plus gamma and invariant sites were used in a run of 20,000,000 generations with two Metropolis-Coupled Monte Carlo Markov chains (MCMCMC) that were run in parallel (starting each from a random tree). Markov chains were sampled every 100 generations, and the first 10 % of the trees was discarded as burn-in. The remaining trees were used to compute the majority rule consensus tree and the posterior probability of clades (Fig. 4).

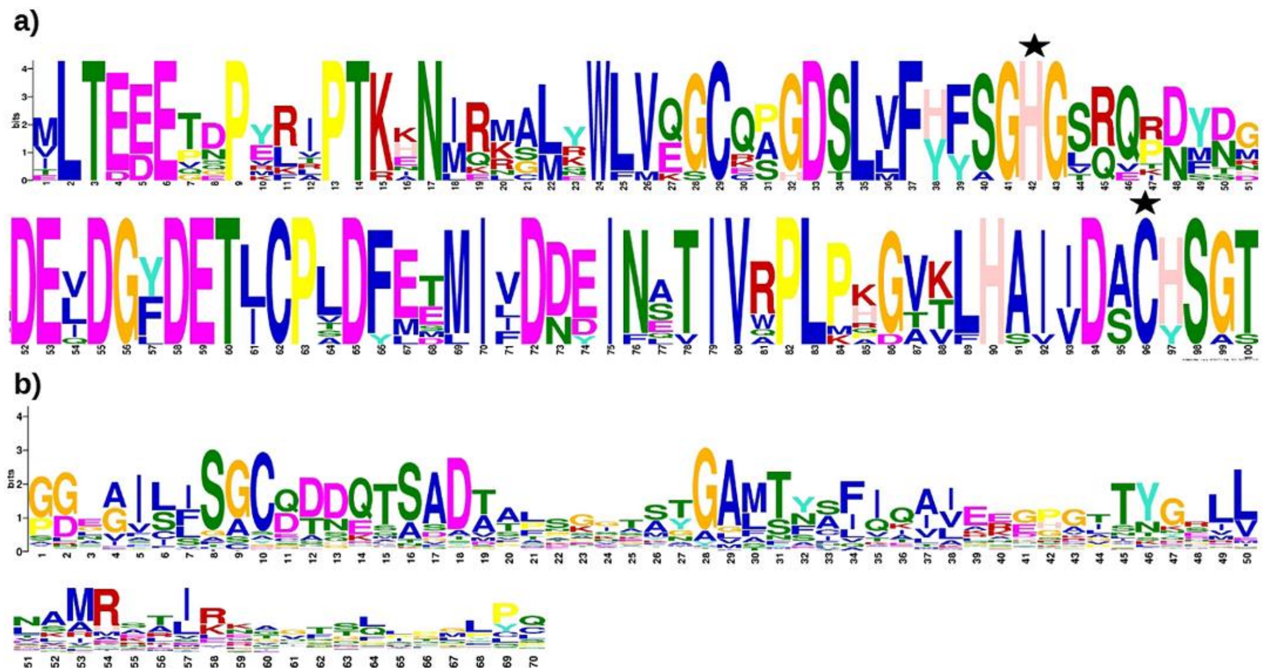


Fig. 2 Analysis of consensus sequence conservation in the metacaspase gene family. The analysis of 370 metacaspase genes was performed using the MEME platform (http://meme.nbcr.net/meme/), using default parameters. The total height of each cell indicates the conserved sequence at each position. The height of each letter is proportional to the corresponding relative frequency. The amino acids are colored according to their chemical properties: *blue* for most hydrophobic residues (A, C, F, I, L, V, and M); *green* for polar, noncharged, nonaliphatic residues (N, Q, S, T); *red* for positively charged residues (K and R); *orange* for glycine (G); *rose* for histidine (H); *yellow* for proline (P); and *turquoise* for tyrosine (Y). *Black star* indicates the conservation of cysteine and histidine residues. **a** P20 and **b** P10 metascapase conserved domains

Fig. 3 Analysis of consensus sequence conservation of the zinc finger domain in the metacaspase type I gene family. The analysis of the zinc finger domain in the metacaspase type I gene family was performed using the MEME platform (http://meme.nbcr.net/meme/). The total height of each cell indicates the conserved sequence at each position. The height of each letter is proportional to the corresponding relative frequency. The amino acids are colored according to their chemical properties: *blue* for most hydrophobic residues (A, C, F, I, L, V and M); green for polar, non-charged, non-aliphatic residues (N, Q, S, T); red for positively-charged residues (K and R); *orange* for glycine (G); *rose* for histidine (H); *yellow* for proline (P); and *turquoise* for tyrosine (Y)

The phylogeny showed that the metacaspases are divided into three classes (with exceptions): type I with (green names, Fig. 4) and without (pink names, Fig. 4) zinc finger domain and type II metacaspases (blue names, Fig. 4). The ancient representatives (algae and moss, indicated with asterisks in Fig. 4) are presented as outgroup inside the principal clusters formed by metacaspase genes (type I with and without zinc finger and type II). This observation suggests that these three classes of metacaspases originated in the early stages of Viridiplantae evolution (Fig. 4). We also observed that some monocot and eudicot representatives formed subclusters inside the three classes, indicating that duplication events might have occurred after the monocot and eudicot divergence (Fig. 4). We hypothesized that the acquisition of the zinc finger domain seems to have occurred later during the metacaspase gene family evolution. The absence of the zinc finger domain seems to be a property of the ancient condition based on the fact that no zinc finger domains were observed in algae representatives.

The role of the metacaspase gene family in plant stress responses

Plant pathogen attacks can induce PCD at the infection sites through a mechanism known as hypersensitive response (HR). HR is an integral part of plant immune systems and one of the most dramatic manifestations of PCD (Bozhkov and Lam 2011). During HR, plants attempt to block the invasion of biotrophic pathogens, leading to localized cell death at the site of infection (Ameisen 2002; Jones 2001; Vercammen et al. 2007). As a result, the plant isolates the pathogen in an inhospitable environment, thus limiting pathogen spread and avoiding greater damage and even the death of the entire plant (De Pinto et al. 2012; Greenberg 1996) (Fig. 5).

In plants, R genes are involved extensively in the recognition of pathogens, functioning as controlling adaptors in a plant apoptosome-like structure (Hoeberichts and Woltering 2003). This apoptosome-like structure would be activated by the pathogen avirulence signals, and R gene products could then help in the activation of caspase-like proteases involved with PCD (Fig. 5) (Lord and Gunawardena 2012).

Following a pathogen attack, one of the first host responses is the production of high levels of reactive oxygen species (ROS) (Hamann et al. 2008) (Fig. 5). During PCD, a strong interplay occurs between ROS and other signaling molecules, redox metabolites, lipid messengers, or plant hormones, the result of which is the modification of the biological response to altered ROS levels and cell fate determination (De Pinto et al. 2012). Oxidative stress-induced PCD, therefore, represents a dramatic example of a metacaspase-dependent process conserved from protozoa to plants (Tsiatsiani et al. 2011).

Different phytohormones are involved in modulating HR cell death under various conditions. Salicylic acid (SA) is a well-known mediator of systemic acquired resistance (SAR) in plants. Nitric oxide (NO) cooperates with SA to induce HR cell death and activate defense mechanisms (Lam 2004). Jasmonic acid (JA) and ethylene regulate cell death under stress conditions and during development. The effects of these phytohormones on PCD are believed to be stress-specific. JA, for instance, negatively regulates cell death in *A. thaliana* under oxidative stress by ozone treatment, but it might work as a positive factor to promote cell death by the fungal toxin fumonisin B1 (Lam 2004).

Several genes involved in HR have been identified in different species (Cabreira et al. 2013; Liao et al. 2014; Wu et al. 2014; Eck et al. 2014; Kumar et al. 2014). In certain plant–pathogen interactions, the development of PCD seems to be mediated by metacaspases (Fig. 5) (Ahmad et al. 2012; Hoeberichts et al. 2003; Uren et al. 2000). Metacaspases play important roles in different species in response to various pathogens and elicitors, and during various abiotic stresses

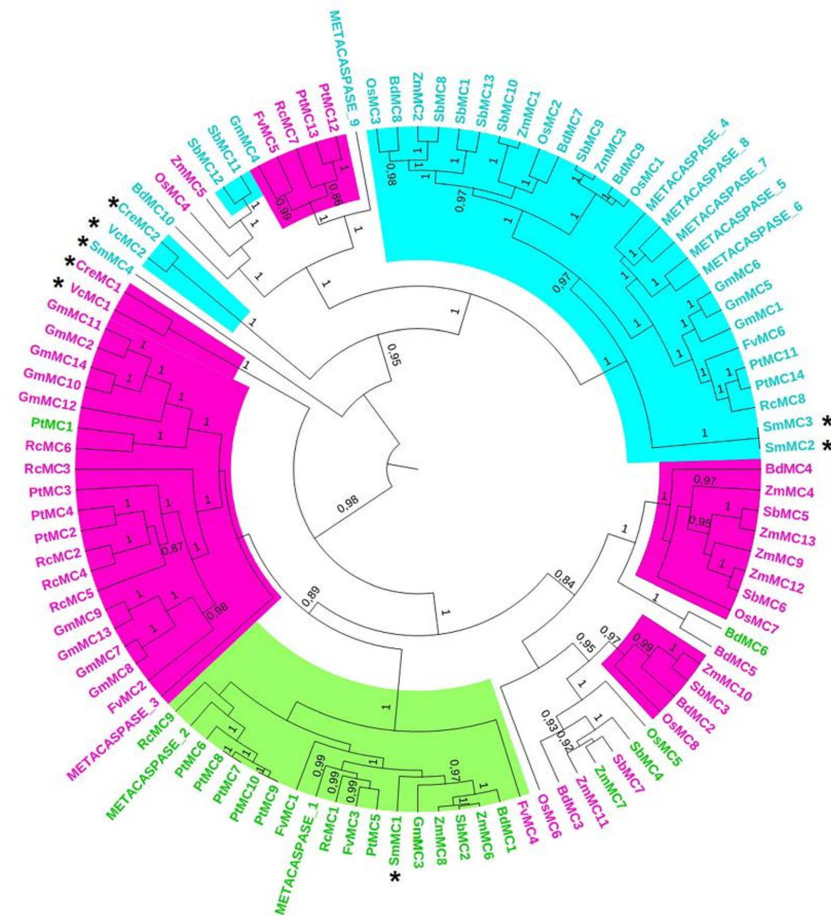


Fig. 4 Phylogenetic analysis of the metacaspase gene family. The Bayesian phylogenetic analysis was reconstructed following protein sequence alignments using a Bayesian approach by employing Beast software. The metacaspase genes belonging to type I with and without zinc finger domain are colored in *green and pink*, respectively. The metacaspase genes belonging to type II are colored in *blue*. The *asterisks* indicate ancient species (algae and moss representatives). *The numbers in the branches of the tree* indicate the posterior probability values. The acronym of metacaspase genes were named using the first letter of the genus followed by the first letter of the species: AtMC (*Arabidopsis thaliana*) Metacaspase; RcMC (*Ricinus comunnis*); PtMC (*Populus trichocarpa*); GmMC (*Glycine max*); FvMC (*Fragaria vesca*); ZmMC (*Zea mayz*); OsMC (*Oriza sativa*); BdMC (*Brachypodium distachyon*); SmMC (*Selaginella moelledorffii*); CreMC (*Chlamydomonas reinhardtii*); VcMC (*Volvox carteri*); SbMC (*Sorghum bicolor*)

(Coll et al. 2010; Hao et al. 2007; He et al. 2008; Sanmartin et al. 2005; Vercammen et al. 2007; Watanabe and Lam 2011; Zhang et al. 2013).

Evidence suggests that metacaspases are involved in facilitating the development of the morphological features typical of the various classes of PCD, through proteolytic activation of other metacaspases and degradative enzymes (Ahmad et al. 2012; van Doorn et al. 2011). The involvement of metacaspases in cell death is developed in response to ROS, ultraviolet light, herbicide-induced stress, and senescing flowers (Ahmad et al. 2012; He et al. 2008; Vercammen et al. 2007; Watanabe and Lam 2011).

In *A. thaliana*, several metacaspase genes (*AtMC*) are involved in stress responses. A statistically significant reduction in cell death phenotypes in response to pathogen inoculation was observed in knockout mutants of *Arabidopsis* type II metacaspase genes (Baskett 2012; He et al. 2008). Type I metacaspase *AtMC1* has been found to be a positive regulator of hypersensitive cell death (Coll et al. 2010; De Pinto et al. 2012). Watanabe and Lam (2011) showed that *AtMC4* is a positive regulator of PCD brought on by abiotic and biotic stress (Watanabe and Lam 2011). *AtMC4* and *AtMC5* are rapidly induced in infection with bacterial pathogens (Baskett 2012; Watanabe and Lam 2005).

The importance of the type II metacaspases was demonstrated in the resistance of *Nicotiana benthamiana* against *Colletotrichum destructivum* (Hao et al. 2007). The authors have shown that virus-induced gene silencing of a type II *Nicotiana benthamiana* metacaspase (*NbMCA1*) enhanced lesion development following infection with *Colletotrichum*

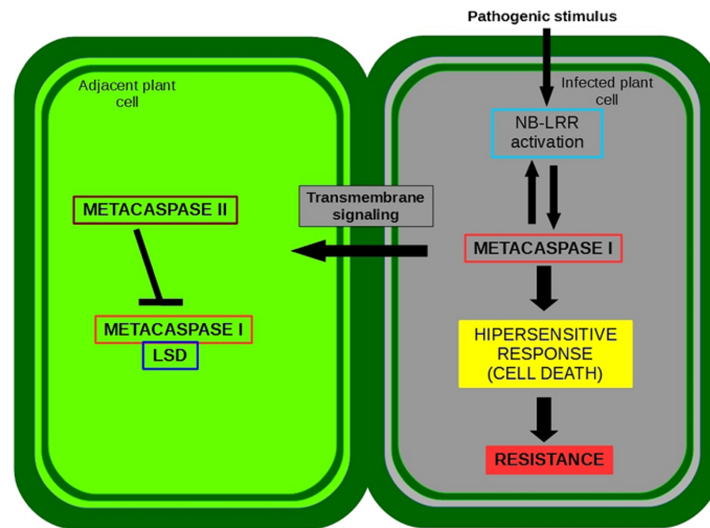


Fig. 5 In plants under stress conditions, metacaspase type I positively regulates hypersensitive response cell death by NB-LRR (nucleotide binding domain) that recognizes pathogenic agents at the site of infection. LSD1 negatively regulates the spread of cell death in cells adjacent to the site of infection, presumably by binding to type I metacaspase, rendering it inactive in the cytoplasm. Metacaspase type II negatively regulates the function of metacaspase type I through an unknown mechanism. Adapted from Coll et al. (2011)

destructivum, a fungal pathogen of the cowpea, but not the bacterial *Pseudomonas syringae* pv. *tomato* (Baskett 2012).

Rice (*Oryza sativa*) metacaspase (*OsMC*) genes are also expressed under abiotic and biotic stresses. *OsMC5* is expressed in *Magnaporthe oryzae*-infected resistant plants, *OsMC1* in plants under cold stress, *OsMC6* and *OsMC8* in plants under drought stress, and *OsMC6* is expressed when rice leaves were damaged by beef armyworms (Dona et al. 2013; Gupta et al. 2012; Ning et al. 2002; Wang and Zhang 2014).

Other studies have pointed to metacaspase involvement in plant PCD responses, notably in Norway spruce (*Picea abies*) (De Pinto et al. 2012; Suarez et al. 2004) and tomato (*Esculentum personicum*) (Baskett 2012). Knocking down type II metacaspases in *P. abies* abolishes somatic embryogenesis related to PCD (Bozhkov et al. 2005a, b; Suarez et al. 2004; Vercammen et al. 2006). Tomato type II metacaspase (LeMCA1) was upregulated following inoculation with the fungal pathogen *Botrytis cinerea* (Baskett 2012; Hoeberichts et al. 2003; Zhang et al. 2013).

Are metacaspases true caspases?

The assignment of metacaspases and caspases to the same family is controversial. First, metacaspases possess substrate specificity (cleavage at arginine or lysine) different from that of caspases (cleavage at aspartic acid). Therefore, metacaspases may not be responsible for the PCD caspase activity detected in nonanimal representatives (Asplund-Samuelsson et al. 2012). Second, P1 preference of metacaspases is basic whereas that of caspases is acidic (Gonzalez et al. 2007; Vercammen et al. 2004). Third, previous phylogenetic analyses of clan CD peptidases have shown that caspases and metacaspases constitute separate groups, suggesting different functions for both kinds of proteins (Cambra et al. 2010).

On the other hand, although a major divergence of the amino acid specificity of caspases and metacaspases may have occurred during evolution, their target proteins should fall into similar functional groups if the role of caspases and metacaspases was conserved (Carmona-Gutierrez et al. 2010). Caspases and metacaspases do fulfill the criteria of homology (common cellular program, PCD, and common substrates, at least in part) and represent variants of the same enzyme that has varied in evolution, in particular regarding the cleavage site: whereas metacaspases exist in virtually all eukaryotic organisms lacking caspases, the presence of caspases excludes that of metacaspases. In this light, caspases and their aspartate specificity may constitute a secondary development in just one branch of eukaryotes that includes higher animals. Thus, different lines of evidence reinforce the rational concept that caspases evolved from metacaspases (Carmona-Gutierrez et al. 2010).

In fact, although metacaspases have been involved in responses to stress (Belenghi et al. 2007; Bidle and Falkowski 2004), the role of metacaspases in PCD remains enigmatic (Cambra et al. 2010). Nuclear translocation of *N. spruce* metacaspase during cell disassembly in the embryo-suspensors and in the dying epimastigotes of *Trypanosoma cruzi*, associated with the identification of Tudor staphylococcal nuclease as a common substrate for both *N. spruce* metacaspase mcII-Pa and human caspase-3, suggests that metacaspases can execute PCD-like effector caspases (Tsiatsiani et al. 2011). In this scenario, the study of plant metacaspase genes can clarify their involvement in PCD mechanisms.

References

Ahmad R, Zuily-Fodil Y, Passaquet C, Bethenod O, Roche R, Repellin A (2012) Ozone and aging up-regulate type II metacaspase gene expression and global metacaspase activity in the leaves of field-grown maize (Zea mays L.) plants. Chemosphere 87:789–795. doi:10.1016/j.chemosphere.2011.12.081

Ameisen JC (2002) On the origin, evolution, and nature of programmed cell death: a timeline of four billion years. Cell Death Differ 9:367–393. doi:10.1038/sj/cdd/4400950

Asplund-Samuelsson J, Bergman B, Larsson J (2012) Prokaryotic caspase homologs: phylogenetic patterns and functional characteristics reveal considerable diversity. PLoS One 7:e49888. doi:10.1371/journal.pone.0049888

Balk J, Chew SK, Leaver CJ, McCabe PF (2003) The intermembrane space of plant mitochondria contains a DNase activity that may be involved in programmed cell death. Plant J 34:573–583

Baskett JA (2012) A type II metacaspase interacts with rps1-k-2 in soybean and analysis of the soybean metacaspase gene family Iowa State University

Beers EP, McDowell JM (2001) Regulation and execution of programmed cell death in response to pathogens, stress and developmental cues. Curr Opin Plant Biol 4:561–567

Belenghi B, Romero-Puertas MC, Vercammen D, Brackenier A, Inze D, Delledonne M, Van Breusegem F (2007) Metacaspase activity of Arabidopsis thaliana is regulated by S-nitrosylation of a critical cysteine residue. J Biol Chem 282:1352–1358. doi:10.1074/jbc.M608931200

Bidle KD, Falkowski PG (2004) Cell death in planktonic, photosynthetic microorganisms. Nat Rev Microbiol 2:643–655. doi:10.1038/nrmicro956

Bozhkov P, Jansson C (2007) Autophagy and cell-death proteases in plants: two wheels of a funeral cart. Autophagy 3:136–138

Bozhkov PV, Filonova LH, Suarez MF (2005a) Programmed cell death in plant embryogenesis. Curr Top Dev Biol 67:135–179. doi:10.1016/S0070-2153(05)67004-4

Bozhkov PV, Lam E (2011) Green death: revealing programmed cell death in plants. Cell Death Differ 18:1239–1240. doi:10.1038/cdd.2011.86

Bozhkov PV, Smertenko AP, Zhivotovsky B (2010) Aspasing out metacaspases and caspases: proteases of many trades. Sci Signal 3:pe48. doi:10.1126/scisignal.3152pe48

Bozhkov PV, Suarez MF, Filonova LH, Daniel G, Zamyatnin AA Jr, Rodriguez-Nieto S, Zhivotovsky B, Smertenko A (2005b) Cysteine protease mcII-Pa executes programmed cell death during plant embryogenesis. Proc Natl Acad Sci U S A 102:14463–14468. doi:10.1073/pnas.0506948102

Cabreira C, Cagliari A, Bucker-Neto L, Wiebke-Strohm B, de Freitas LB, Marcelino-Guimaraes FC, Nepomuceno AL, Margis-Pinheiro MM, Bodanese-Zanettini MH (2013) The Lesion Simulating Disease (LSD) gene family as a variable in soybean response to Phakopsora pachyrhizi infection and dehydration. Funct Integr Genomics 13(3):323–338

Cacas JL (2010) Devil inside: does plant programmed cell death involve the endomembrane system? Plant Cell Environ 33:1453–1473. doi:10.1111/j.1365-3040.2010.02117.x

Cambra I, Garcia FJ, Martinez M (2010) Clan CD of cysteine peptidases as an example of evolutionary divergences in related protein families across plant clades. Gene 449:59–69. doi:10.1016/j.gene.2009.09.003

Carmona-Gutierrez D, Frohlich KU, Kroemer G, Madeo F (2010) Metacaspases are caspases. Doubt no more. Cell Death Differ 17:377–378. doi:10.1038/cdd.2009.198

Coffeen WC, Wolpert TJ (2004) Purification and characterization of serine proteases that exhibit caspase-like activity and are associated with programmed cell death in Avena sativa. Plant Cell 16:857–873. doi:10.1105/tpc.017947

Coll NS, Epple P, Dangl JL (2011) Programmed cell death in the plant immune system. Cell Death Differ 18:1247–1256. doi:10.1038/cdd.2011.37

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. doi:10.1126/science.1194980

Danon A, Rotari VI, Gordon A, Mailhac N, Gallois P (2004) Ultraviolet-C overexposure induces programmed cell death in Arabidopsis, which is mediated by caspase-like activities and which can be suppressed by caspase inhibitors, p35 and Defender against Apoptotic Death. J Biol Chem 279:779–787. doi:10.1074/jbc.M304468200

De Michele R, Formentin E, Lo Schiavo F (2009) Legume leaf senescence: a transcriptional analysis. Plant Signal Behav 4:319–320

De Pinto MC, Locato V, De Gara L (2012) Redox regulation in plant programmed cell death. Plant Cell Environ 35:234–244. doi:10.1111/j.1365-3040.2011.02387.x

del Pozo O, Lam E (1998) Caspases and programmed cell death in the hypersensitive response of plants to pathogens. Curr Biol 8:R896

Dickman MB, Reed, J.C. (2004) Paradigms of programmed cell death in animals and plantsProgrammed Cell Death in Plants. Blackwell Publishing,, pp. 26-43

Dona M, Macovei A, Fae M, Carbonera D, Balestrazzi A (2013) Plant hormone signaling and modulation of DNA repair under stressful conditions. Plant Cell Rep 32:1043–1052. doi:10.1007/s00299-013-1410-9

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214. doi:10.1186/1471-2148-7-214

Eck LV, Davidson RM, Wu S, Zhao BY, Botha AM, Leach JE, Lapitan NL (2014) The transcriptional network of WRKY53 in cereals links oxidative responses to biotic and abiotic stress inputs. Funct Integr Genomics 14(2):351–362

Fath A, Bethke P, Lonsdale J, Meza-Romero R, Jones R (2000) Programmed cell death in cereal aleurone. Plant Mol Biol 44:255–266

Fukuda H (1997) Tracheary Element Differentiation. Plant Cell 9:1147–1156. doi:10.1105/tpc.9.7.1147

Fukuda H (2000) Programmed cell death of tracheary elements as a paradigm in plants. Plant Mol Biol 44:245–253

Giuliani C, Consonni G, Gavazzi G, Colombo M, Dolfini S (2002) Programmed cell death during embryogenesis in maize. Ann Bot 90:287–292

Gonzalez IJ, Desponds C, Schaff C, Mottram JC, Fasel N (2007) Leishmania major metacaspase can replace yeast metacaspase in programmed cell death and has arginine-specific cysteine peptidase activity. Int J Parasitol 37:161–172. doi:10.1016/j.ijpara.2006.10.004

Greenberg JT (1996) Programmed cell death: a way of life for plants. Proc Natl Acad Sci U S A 93:12094–12097

Greenberg JT, Yao N (2004) The role and regulation of programmed cell death in plant-pathogen interactions. Cell Microbiol 6:201–211

Gunawardena AH (2008) Programmed cell death and tissue remodelling in plants. J Exp Bot 59:445–451. doi:10.1093/jxb/erm189

Gunawardena AH, Greenwood JS, Dengler NG (2004) Programmed cell death remodels lace plant leaf shape during development. Plant Cell 16:60–73. doi:10.1105/tpc.016188

Gunawardena AH, Pearce DM, Jackson MB, Hawes CR, Evans DE (2001) Characterisation of programmed cell death during aerenchyma formation induced by ethylene or hypoxia in roots of maize (Zea mays L.). Planta 212:205–214

Gupta SK, Rai AK, Kanwar SS, Chand D, Singh NK, Sharma TR (2012) The single functional blast resistance gene Pi54 activates a complex defence mechanism in rice. J Exp Bot 63:757–772. doi:10.1093/jxb/err297

Hamann A, Brust D, Osiewacz HD (2008) Apoptosis pathways in fungal growth, development and ageing. Trends Microbiol 16:276–283. doi:10.1016/j.tim.2008.03.003

Hao L, Goodwin PH, Hsiang T (2007) Expression of a metacaspase gene of Nicotiana benthamiana after inoculation with Colletotrichum destructivum or Pseudomonas syringae pv. tomato, and the effect of silencing the gene on the host response. Plant Cell Rep 26:1879–1888. doi:10.1007/s00299-007-0387-7

He R, Drury GE, Rotari VI, Gordon A, Willer M, Farzaneh T, Woltering EJ, Gallois P (2008) Metacaspase-8 modulates programmed cell death induced by ultraviolet light and H2O2 in Arabidopsis. J Biol Chem 283:774–783. doi:10.1074/jbc.M704185200

Hoeberichts FA, ten Have A, Woltering EJ (2003) A tomato metacaspase gene is upregulated during programmed cell death in Botrytis cinerea-infected leaves. Planta 217:517–522. doi:10.1007/s00425-003-1049-9

Hoeberichts FA, Woltering EJ (2003) Multiple mediators of plant programmed cell death: interplay of conserved cell death mechanisms and plant-specific regulators. Bioessays 25:47–57. doi:10.1002/bies.10175

Jones AM (2001) Programmed cell death in development and defense. Plant Physiol 125:94–97

Koonin EV, Aravind L (2002) Origin and evolution of eukaryotic apoptosis: the bacterial connection. Cell Death Differ 9:394–404. doi:10.1038/sj/cdd/4400991

Kumar A, Bimolata W, Kannan M, Kirti PB, Qureshi IA, Ghazi IA (2014) Comparative proteomics reveals differential induction of both biotic and abiotic stress response associated proteins in rice during Xanthomonas oryzae pv. oryzae infection. Funct Integr Genomics 15(4):425–437

Kwon SI, Cho HJ, Park OK (2010) Role of Arabidopsis RabG3b and autophagy in tracheary element differentiation. Autophagy 6:1187–1189. doi:10.4161/auto.6.8.13429

Lam E (2004) Controlled cell death, plant survival and development. Nat Rev Mol Cell Biol 5:305–315. doi:10.1038/nrm1358

Lam E, Kato N, Lawton M (2001) Programmed cell death, mitochondria and the plant hypersensitive response. Nature 411:848–853. doi:10.1038/35081184

Lam E, Zhang Y (2012) Regulating the reapers: activating metacaspases for programmed cell death. Trends Plant Sci 17:487–494. doi:10.1016/j.tplants.2012.05.003

Lee KP, Kim C, Landgraf F, Apel K (2007a) EXECUTER1- and EXECUTER2-dependent transfer of stress-related signals from the plastid to the nucleus of Arabidopsis thaliana. Proc Natl Acad Sci U S A 104:10270–10275. doi:10.1073/pnas.0702061104

Lee N, Gannavaram S, Selvapandiyan A, Debrabant A (2007b) Characterization of metacaspases with trypsin-like activity and their putative role in programmed cell death in the protozoan parasite Leishmania. Eukaryot Cell 6:1745–1757. doi:10.1128/EC.00123-07

Lee RE, Puente LG, Kaern M, Megeney LA (2008) A non-death role of the yeast metacaspase: Yca1p alters cell cycle dynamics. PLoS One 3:e2956. doi:10.1371/journal.pone.0002956

Levine A, Pennell RI, Alvarez ME, Palmer R, Lamb C (1996) Calcium-mediated apoptosis in a plant hypersensitive disease resistance response. Curr Biol 6:427–437

Liao W, Ji L, Wang J, Chen Z, Ye M, Ma H, An X (2014) Identification of glutathione S-transferase genes responding to pathogen infestation in Populus tomentosa. Funct Integr Genomics 14(3):517–529

Lim PO, Kim HJ, Nam HG (2007) Leaf senescence. Annu Rev Plant Biol 58:115–136. doi:10.1146/annurev.arplant.57.032905.105316

Lombardi L, Ceccarelli N, Picciarelli P, Sorce C, Lorenzi R (2010) Nitric oxide and hydrogen peroxide involvement during programmed cell death of Sechium edule nucellus. Physiol Plant 140:89–102. doi:10.1111/j.1399-3054.2010.01381.x

Lord CE, Gunawardena AH (2011) Environmentally induced programmed cell death in leaf protoplasts of Aponogeton madagascariensis. Planta 233:407–421. doi:10.1007/s00425-010-1304-9

Lord CE, Gunawardena AH (2012) Programmed cell death in C. elegans, mammals and plants. Eur J Cell Biol 91:603–613. doi:10.1016/j.ejcb.2012.02.002

Love AJ, Milner JJ, Sadanandom A (2008) Timing is everything: regulatory overlap in plant cell death. Trends Plant Sci 13:589–595. doi:10.1016/j.tplants.2008.08.006

Madeo F, Herker E, Maldener C, Wissing S, Lachelt S, Herlan M, Fehr M, Lauber K, Sigrist SJ, Wesselborg S, Frohlich KU (2002) A caspase-related protease regulates apoptosis in yeast. Mol Cell 9:911–917

Mottram JC, Helms MJ, Coombs GH, Sajid M (2003) Clan CD cysteine peptidases of parasitic protozoa. Trends Parasitol 19:182–187

Mur LA, Kenton P, Lloyd AJ, Ougham H, Prats E (2008) The hypersensitive response; the centenary is upon us but how much do we know? J Exp Bot 59:501–520. doi:10.1093/jxb/erm239

Nedelcu AM (2009) Comparative genomics of phylogenetically diverse unicellular eukaryotes provide new insights into the genetic basis for the evolution of the programmed cell death machinery. J Mol Evol 68:256–268. doi:10.1007/s00239-009-9201-1

Ning SB, Song YC, Damme Pv P (2002) Characterization of the early stages of programmed cell death in maize root cells by using comet assay and the combination of cell electrophoresis with annexin binding. Electrophoresis 23:2096–2102. doi:10.1002/1522-2683(200207)23:13<2096::AID-ELPS2096>3.0.CO;2-V

Ojha M, Cattaneo A, Hugh S, Pawlowski J, Cox JA (2010) Structure, expression and function of Allomyces arbuscula CDP II (metacaspase) gene. Gene 457:25–34. doi:10.1016/j.gene.2010.02.014

Pennell RI, Lamb C (1997) Programmed Cell Death in Plants. Plant Cell 9:1157–1168. doi:10.1105/tpc.9.7.1157

Potten C, Wilson J (2004) Apoptosis - the life and death of cells. University Press, Cambridge, Cambridge

Ranganath RM, Nagashree NR (2001) Role of programmed cell death in development. Int Rev Cytol 202:159–242

Rezanezhad H, Menegon M, Sarkari B, Hatam GR, Severini C (2011) Characterization of the metacaspase 1 gene in Plasmodium vivax field isolates from southern Iran and Italian imported cases. Acta Trop 119:57–60. doi:10.1016/j.actatropica.2011.03.010

Rogers HJ (2005) Curr Top Dev Biol 71:225–261. doi:10.1016/S0070-2153(05)71007-3

Rogers HJ (2006) Programmed cell death in floral organs: how and why do flowers die? Ann Bot 97:309–315. doi:10.1093/aob/mcj051

Rojo E, Martin R, Carter C, Zouhar J, Pan S, Plotnikova J, Jin H, Paneque M, Sanchez-Serrano JJ, Baker B, Ausubel FM, Raikhel NV (2004) VPEgamma exhibits a caspase-like activity that contributes to defense against pathogens. Curr Biol 14:1897–1906. doi:10.1016/j.cub.2004.09.056

Sanmartin M, Jaroszewski L, Raikhel NV, Rojo E (2005) Caspases. Regulating death since the origin of life. Plant Physiol 137:841–847. doi:10.1104/pp. 104.058552

Senatore A, Trobacher CP, Greenwood JS (2009) Ricinosomes predict programmed cell death leading to anther dehiscence in tomato. Plant Physiol 149:775–790. doi:10.1104/pp. 108.132720

Staal J, Dixelius C (2007) Tracing the ancient origins of plant innate immunity. Trends Plant Sci 12:334–342. doi:10.1016/j.tplants.2007.06.014

Suarez MF, Filonova LH, Smertenko A, Savenkov EI, Clapham DH, von Arnold S, Zhivotovsky B, Bozhkov PV (2004) Metacaspase-dependent programmed cell death is essential for plant embryogenesis. Curr Biol 14:R339–R340. doi:10.1016/j.cub.2004.04.019

Szallies A, Kubata BK, Duszenko M (2002) A metacaspase of Trypanosoma brucei causes loss of respiration competence and

clonal death in the yeast Saccharomyces cerevisiae. FEBS Lett 517: 144–150
Tata JR (1966) Requirement for RNA and protein synthesis for induced regression of the tadpole tail in organ culture. Dev Biol 13:77–94
Thornberry NA, Lazebnik Y (1998) Caspases: enemies within. Science 281:1312–1316
Trzyna WC, Legras XD, Cordingley JS (2008) A type-1 metacaspase from Acanthamoeba castellanii. Microbiol Res 163:414–423. doi:10.1016/j.micres.2006.06.017
Tsiatsiani L, Van Breusegem F, Gallois P, Zavialov A, Lam E, Bozhkov PV (2011) Metacaspases. Cell Death Differ 18:1279–1288. doi:10.1038/cdd.2011.66
Uren AG, O'Rourke K, Aravind LA, Pisabarro MT, Seshagiri S, Koonin EV, Dixit VM (2000) Identification of paracaspases and metacaspases: two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma. Mol Cell 6:961–967
van Doorn WG, Beers EP, Dangl JL, Franklin-Tong VE, Gallois P, Hara-Nishimura I, Jones AM, Kawai-Yamada M, Lam E, Mundy J, Mur LA, Petersen M, Smertenko A, Taliansky M, Van Breusegem F, Wolpert T, Woltering E, Zhivotovsky B, Bozhkov PV (2011) Morphological classification of plant cell deaths. Cell Death Differ 18:1241–1246. doi:10.1038/cdd.2011.36
van Doorn WG, Woltering EJ (2005) Many ways to exit? Cell death categories in plants. Trends Plant Sci 10:117–122. doi:10.1016/j.tplants.2005.01.006
Vercammen D, Belenghi B, van de Cotte B, Beunens T, Gavigan JA, De Rycke R, Brackenier A, Inze D, Harris JL, Van Breusegem F (2006) Serpin1 of Arabidopsis thaliana is a suicide inhibitor for metacaspase 9. J Mol Biol 364:625–636. doi:10.1016/j.jmb.2006.09.010
Vercammen D, Declercq W, Vandenabeele P, Van Breusegem F (2007) Are metacaspases caspases? J Cell Biol 179:375–380. doi:10.1083/jcb.200705193
Vercammen D, van de Cotte B, De Jaeger G, Eeckhout D, Casteels P, Vandepoele K, Vandenberghe I, Van Beeumen J, Inze D, Van Breusegem F (2004) Type II metacaspases Atmc4 and Atmc9 of Arabidopsis thaliana cleave substrates after arginine and lysine. J Biol Chem 279:45329–45336. doi:10.1074/jbc.M406329200
Wang L, Zhang H (2014) Genomewide survey and characterization of metacaspase gene family in rice (Oryza sativa). J Genet 93:93–102
Watanabe N, Lam E (2005) Two Arabidopsis metacaspases AtMCP1b and AtMCP2b are arginine/lysine-specific cysteine proteases and activate apoptosis-like cell death in yeast. J Biol Chem 280: 14691–14699. doi:10.1074/jbc.M413527200
Watanabe N, Lam E (2011) Arabidopsis metacaspase 2d is a positive mediator of cell death induced during biotic and abiotic stresses. Plant J 66:969–982. doi:10.1111/j.1365-313X.2011.04554.x
Williams B, Dickman M (2008) Plant programmed cell death: can't live with it; can't live without it. Mol Plant Pathol 9:531–544. doi:10.1111/j.1364-3703.2008.00473.x
Wilson ZA, Song J, Taylor B, Yang C (2011) The final split: the regulation of anther dehiscence. J Exp Bot 62:1633–1649. doi:10.1093/jxb/err014
Woltering EJ, van der Bent A, Hoeberichts FA (2002) Do plant caspases exist? Plant Physiol 130:1764–1769. doi:10.1104/pp. 006338
Wright H, van Doorn WG, Gunawardena AH (2009) In vivo study of developmental programmed cell death using the lace plant (Aponogeton madagascariensis; Aponogetonaceae) leaf model system. Am J Bot 96:865–876. doi:10.3732/ajb.0800343
Wu HM, Cheun AY (2000) Programmed cell death in plant reproduction. Plant Mol Biol 44:267–281
Wu J, Zhang Y, Yin L, Qu J, Lu J (2014) Linkage of cold acclimation and disease resistance through plant–pathogen interaction pathway in Vitis amurensis grapevine. Funct Integr Genomics 14(4):741–755
Young TE, Gallie DR (2000) Programmed cell death during endosperm development. Plant Mol Biol 44:283–301
Zhang C, Gong P, Wei R, Li S, Zhang X, Yu Y, Wang Y (2013) The metacaspase gene family of Vitis vinifera L.: characterization and differential expression during ovule abortion in stenospermocarpic seedless grapes. Gene 528:267–276. doi:10.1016/j.gene.2013.06.062
Zhang L, Li Y, Xing D, Gao C (2009) Characterization of mitochondrial dynamics and subcellular localization of ROS reveal that HsfA2 alleviates oxidative damage caused by heat stress in Arabidopsis. J Exp Bot 60:2073–2091. doi:10.1093/jxb/erp078

CAPÍTULO 3

Artigo em revisão no Periódico *Plant Gene*

Soybean *(Glycine max)* NF-Y (Nuclear Factor of Y Box) gene family and its potential role under stress conditions and nodulation

Cabreira-Cagliari, C.[1]*; Bohn, B.[2]*; Dias, N. C. F[2]; Fagundes, D. G. S[2]; Margis-Pinheiro, M[1]; Bodanese-Zanettini, M.H.[1]; Cagliari, A[2].

[1]Programa de Pós Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul – UFRGS, Porto Alegre – RS, Brazil

[2]Universidade Estadual do Rio Grande do Sul – UERGS, Santa Cruz do Sul – RS, Brazil

* These authors contributed equally.

Corresponding author: Alexandro Cagliari, e-mail: alexandro-cagliari@uergs.edu.br, phone: +55-51-3715-6926, address: Universidade Estadual do Rio Grande do Sul - UERGS, CEP 96816-501, Santa Cruz do Sul, RS, Brazil.

Abstract

The Nuclear Factor of Y box (NF-Y) is a transcription factor composed of three subunits (NF-YA, NF-YB, and NF-YC) involved in the regulation of genes related to physiological processes. In plants, NF-Y genes act in development, in addition to biotic and abiotic stresses response. Regarding *Glycine max* (*Gm*), the NF-Y genes are poorly investigated. In this important crop, the characterization of genes related to stresses represents an important material for improvement of cultivars. In this study, *GmNF-Y* genes were investigated through an extensive data mining using several bioinformatics approaches. We identify a total of 21 genes of *GmNF-YA*, 23 to *GmNF-YB*, and 24 to *GmNF-YC* subunits. During evolution, *GmNF-Y* genes arose by block duplication and subfuctionalization processes could be occurring in this family. *In silico* expression analyses indicate that *GmNF-Y* genes are not ubiquitously expressed in organs; their expression are instead organ-dependent. In several biotic and abiotic stresses, the majority of *GmNF-Y* genes are modulated, especially during *Phakopsora pachyrhizi* infection, representing a new characterization for these genes. In addition, in root hair inoculated with *Bradyrhizobium japonicum,* some genes are modulated. Our results have showed an overview of the NF-Y family in soybean, through the characterization of *Gm*NF-Y genes and their expression pattern, suggesting their involvement in plant defense responses.

Key words: NF-Y transcription factor, gene evolution, soybean nodulation, biotic and abiotic stresses.

Introduction

The Nuclear Factor of Y box (NF-Y) is a eukaryotic conserved oligotrimeric transcription factor involved in the regulation of genes related to different physiological processes. NF-Y is composed of three subunits: NF-YA, NF-YB, and NF-YC. Each subunit is required for DNA binding, subunit association, and transcriptional regulation (Stephenson et al. 2007). The three NF-Y subunits interact, forming a heterocomplex in plants (Baudin et al. 2015) that can bind CCAAT DNA motifs, controlling the expression of target genes. NF-YA performs specific recognition and binding to the CCAAT-box sequence in the promoter of target genes. NF-YB and NF-YC promote chromatin accessibility to NF-YA and to additional transcription factors involved with the transcriptional regulation process (Baudin et al. 2015; Nardini et al. 2013; Oldfield et al. 2014).

It has been previously showed that NF-Y genes are involved with the transcriptional control of gametogenesis, embryogenesis, nodule development and rhizobia symbiosis, seed germination, development and desiccation, grain filling, bud dormancy, hormonal signaling and flowering time, root architecture and elongation, fruit ripening, blue light responses, photosynthesis, endoplasmic reticulum stress response, and drought tolerance (Baudin et al. 2015; Fornari et al. 2013; Hwang et al. 2016; Kim et al. 2016; Kumimoto et al. 2010; Li et al. 2016; Liu and Howell 2010; Mei et al. 2015; Mu et al. 2013; Nelson et al. 2007; Quach et al. 2015; Soyano et al. 2013; Stephenson et al. 2010; Stephenson et al. 2011; Sun et al. 2014; Xu et al. 2014; Yadav et al. 2015; Zanetti et al. 2010). These results clarify that the characterization of NF-Y genes in several species could contributed to the understanding of the mechanisms involved with the genetic regulation in plants.

While most eukaryotic genomes possess only one or two genes encoding each NF-Y subunit, in vascular plants each subunit is encoded by gene families (Riechmann and Ratcliffe

2000). In *Physcomitrella patens*, 2 NF-YA, 9 NF-B, and 12 NF-YC proteins were identified (Zhang et al. 2015a). In the emerging monocot model plant *Brachypodium distachyon*, there are 7 NF-YA, 17 NF-B, and 12 NF-YC proteins (Cao et al. 2011). In rice, there are at least 10 NF-YA, 11 NF-B, and 7 NF-YC genes (Thirumurugan et al. 2008), while in *Arabidopsis* there are 10 genes coding for NF-YA subunit, 13 for the NF-B subunit, and 13 for the NF-YC subunit (Gusmaroli et al. 2002).

Soybean (*Glycine max*) is one of the most important crops in the world. Different types of stresses severely restrict soybean productivity. The identification of genes related to stresses regulation is a fundamental issue, representing the raw material for improvement of soybean cultivars. Regarding the NF-Y family, it is poorly investigate in soybean. Here, we carried out an extensive data mining to identify NF-Y genes (*GmNF-Y*) in soybean genome. We characterized their duplication and evolution, and gene structures. *In silico* expression analyses of *GmNF-Y* in organs and stresses conditions were investigated. Our results contribute to the knowledge about the *GmNF-Y* and lay a foundation for deep characterization of this gene family in soybean.

Material and Methods

NF-Y gene annotation and domains characterization

The nucleotide and protein sequences of the *Arabidopsis thaliana* genes were used as queries to BLAST searches in NCBI (http://www.ncbi.nlm.nih.gov), Phytozome (http://www.phytozome.org/), and Gramene (http://www.gramene.org) databases. Genes with at least 25% of homology with the queries were examined to the presence of the NF-Y domains (Stephenson et al. 2007) using SMART (http:// smart.embl-heidelberg.de/) and InterProScan Signature (http://smart.embl-heidelberg.de/). The exon-intron of each *GmNF-Y* gene was investigated using GSDS (http://gsds.cbi.pku.edu.cn/). GmNF-Y domain conservation was investigated using MEME (http://meme.sdsc.edu/meme/).

Duplication events, synteny, and *Ka/Ks* substitution rate analysis

We have analyzed the mechanisms that are involved with the evolution of *GmNF-Y* genes using the WGmapping tool in the PLAZA v 3.0 database (http://bioinformatics.psb.ugent.be/plaza/). The physical co-localization of NF-Y genes in soybean genome was analyzed in the Genome Duplication Database (PGDD) (http://chibba.

agtec.uga.edu/duplication/), considering 100 kb syntenic region between paralogous *GmNF-Y*. This analysis uses the BLASTP tool to search for potential anchors (*E* <1e−5, top 5 matches) among every possible chromosome pair in multiple genomes. The homolog pairs identified are used as the input in the multiple collinearity scan (MCsan) program, and an *E* value <1e−10 is considered a significant cut-off. The putative classification of the genes identified in the syntenic regions was accessed using Phytozome and the PLAZA databases.

The ratio of the number of nonsynonymous substitutions per nonsynonymous site (*K*a) to the number of synonymous substitutions per synonymous site (*K*s) in paralogous *GmNF-Y* genes was analyzed using the PGDD database (http://chibba.agtec.uga.edu/duplication/). In this database, the protein sequences were aligned using CLUSTALW as a guide to the CDS (coding DNA sequence) alignments by PAL2NAL (http://abacus. gene.ucl.ac.uk/software/paml.html). Finally, the Nei-Gojobori method was applied to generate the *Ka* and *Ks* ratio.

Expression analysis of *GmNF-Y* genes

The expression pattern of *GmNF-Y* genes was analyzed using BAR (The Botany Array Resource, http://bar.utoronto.ca/) and Genevestigator (In the BAR database, gene expression was analyzed in leaves, flowers, green pods, SAM (Shoot Apical Meristem), roots, root tips, and nodules. An additional expression analysis was performed in roots inoculated with *Bradyrhizobium japonicum*. Gene expression was monitored in soybean stripped roots and root hairs in different hours after inoculation (HAI). Using Genevestigator database (https://genevestigator.com/gv/), the expression of *GmNF-Y* genes was investigated under different stress conditions, for different perturbations (biotic, chemical, temperature and stress), and in different soybean tissues (inflorescence, roots, seedling and shoots).

Results

GmNF-Y gene annotation and protein *in silico* characterization

Using BLAST approaches, a total of 21 genes encoding *GmNF-YA*, 23 *GmNF-YB*, and 24 *GmNF-YC* subunits in soybean genome were identified (Figure 1 and ESM1 - Electronic

Supplementary Material 1). These genes were distributed in several locus, among 2 and 20 chromosomes. A wide difference in the length of *GmNF-Y* genes was observed, ranging to 500 bp up to 6000 bp. Analysis of exon-intron structure showed that the number of exons in *GmNF-YA* genes range of four up to six. In *GmNF-YB* and *GmNF-YC* genes, even the majority of genes possess one exon, the exon number vary of one up to five in *GmNF-YB,* and one up to six in *GmNF-YC*. Larges 5´UTR compared to the total gene length were observed in some sequences, as GmNF-YA9, GmNF-YA10, GmNF-YA 11, GmNF-YA14, GmNF-YB15, GmNF-YC8, GmNF-YC9, GmNF-YC11, GmNF-YC12, GmNF-YC13 and GmNF-YC15. In addition, wide 3´UTR were also observed in GmNF-YB14, GmNF-YB19 and GmNF-YC21. The consensus analyses of the *GmNF-Y* domains showed a high degree of amino acid conservation, especially for *GmNF-YA* and *GmNF-YC* subunits (Figure 3).

NF-Y genes are located in a duplicated region

Block duplication processes generated all *GmNF-Y* genes encoding subunits A and B (Supplementary material 1). Regarding subunit C, *GmNF-Y13*, *GmNF-Y15,* and *GmNF-Y24* were originated by tandem duplication, and *GmNF-Y18* by tandem and block duplication. Considering the analysis of syntenic regions containing the duplicated NF-Y genes, *GmNF-YA4, GmNF-YB4* and *GmNF-YC10* exemplify the results (Figure 2 and Supplementary material 2). A small duplication block between *GmNF-YB4* and *GmNF-YB2*; a large duplication blocks between *GmNF-YA4* and *GmNF-YA5, GmNF-YB4* and *GmNF-YB3,* and between *GmNF-YB4* and *GmNF-YB1,* and a huge duplication between *GmNF-YB4* and *GmNF-YB11, GmNF-YB4* and *GmNF-YB5*, and between *GmNF-YC10* and *GmNF-YC20* was found.

To analyze the processes that have driven the changes at the molecular level, the *Ka/Ks* values between *GmNF-Y* syntenic genes were evaluated. The results showed that all *Ka/Ks* values were <1, indicating that a subfuctionalization process could be acting on these paralogous sequences (Supplementary material 2).

In silico* expression of *GmNF-Y

The results considering *in silico* expression of *GmNF-Y* in each organ and stress conditions is showed in Supplementary figures 1-3. All NF-Y genes available in BAR were presented. Figure 3 exemplify the results for all subunits.

Regarding subunit A, we observed that in green pods, *GmNF-YA2* was upregulated, while *GmNF-YA8* and *GmNF-YA12* were downregulated. In nodules, *GmNF-YA4, 5, 6, 9,* and *11* were highly upregulated, while *GmNF-YA1* was downregulated. In roots, *GmNF-YA2* was downregulated. In root tips, *GmNF-YA3, 7, 8, 9, 10, 12,* and *18* were downregulated. *GmNF-YA5* was downregulated in flowers. Regarding the NF-B subunits, *GmNF-YB4* was up- and *GmNF-YB19* downregulated in nodules. In roots, *GmNF-YB4* was up- and *GmNF-YB16* was downregulated. In root tips, *GmNF-YB1, 2, 3, 4, 5,* and *17* were downregulated, while *GmNF-YB21* was upregulated. In leaves, *GmNF-YB16* was up-, while *GmNF-YB17* was downregulated. The expression of genes corresponding to subunit C demonstrated that *GmNF-YC3* was upregulated, while *GmNF-YC9* and *GmNF-YC10* were downregulated in nodules. In roots, *GmNF-YC4* and *GmNF-YC11* were upregulated. In root tips, *GmNF-YC3* and *GmNF-YC7* were downregulated, while *GmNF-YC20* was upregulated. *GmNF-YC3, 20* and *21* were upregulated in SAM.

In root hairs of plants growing in soil colonized with *Bradyrhizobium japonicum* (see example in Figure 3 and all results in supplementary figures 1-3), *GmNF-YA2* (12 HAI) and *GmNF-YA17* were upregulated (12 and 24 HAI). *GmNF-YA6* was downregulated in stripped roots (48 HAI). Regarding subunit B genes, *GmNF-YB17* was highly upregulated in root hairs (12 HAI), and downregulated in stripped roots (48 HAI) and root hairs (48 HAI) (Figure 3). Considering subunit C, *GmNF-YC4, 7, 9, 19,* and *20* were downregulated, while *GmNF-YC11* was highly upregulated in stripped roots (48 HAI). *GmNF-YC11* (48 HAI), *GmNF-YC19* (12 HAI), and *GmNF-YC*20 (24 HAI) were downregulated in root hairs (Figure 3).

Under stress conditions, the majority of *GmNF-YA* genes were modulated, being highly up- or downregulated in inflorescences (Figure 3). *GmNF-YB4* was highly modulated under biotic stress conditions. *GmNF-YC10* was highly induced after *P. pachyrhizi* inoculation. In roots, seedlings, and shoots, the majority of *GmNF-YA* and some *GmNF-YB* genes were modulated under biotic stress, especially under *P. pachyrhizi* fungi infection. For elicitor, nutrient, chemical, and temperature stress conditions, the majority of *GmNF-Y* genes were typically downregulated in all tissues.

Discussion

The gene families size reflects the number of duplicated genes. As observed to several gene families, considering the NF-Y family, soybean presents a high number of genes in comparison with other species, as *Arabidopsis* (36 NF-Y genes) (Gusmaroli et al. 2002) and rice (28 NF-Y genes) (Thirumurugan et al. 2008). Even with variation in length, the structural analysis of *GmNF-Y* genes by exon-intron composition showed that some genes present a very similar structure, as observed in all *GmNF-YA* genes, in which the majority presents five exons. The composition of *GmNF-YA18* suggests an exon gain, since it is the only one with six exons, while the composition of *GmNF-YA6*, 9, 10 and 11 suggests the loss of one exon. In relation to *GmNF-YB* and *GmNF-YC,* the majority genes possess only one exon, maybe representing the ancestral structure in these subunits. As the *GmNF-Y* genes are located in duplicated regions, appointed a common origin of paralogous genes, the acquisition of exons may have occurred during the evolution.

Block duplication events happen frequently in plants, because most of them are polyploids and retain duplicated chromosomal blocks. We found the majority of the *GmNF-Y* genes are located in duplicated blocks (Supplementary material 1 and 2, Figure 2), suggesting that segmental duplication contributed significantly to the family expansion. Similar results were observed in annexin (Jami et al. 2012) and LSD (*Lesion Simulating disease*) family in Viridiplantae (Cabreira et al. 2015). Interestingly, genes corresponding to subunit C were not duplicated or the duplicated copy was lost, as observed for *GmNF-YC14*, *GmNF-YC17,* and *GmNF-YC23* (ESM 3). In fact, several genes from other families present in the syntenic regions had no correspondent paralogous (Figure 2, white arrows), suggesting that the duplicated genes were lost. Multiple episodes of unequal crossovers might lead to increases/decreases in gene copy. Tandem duplication often results from unequal crossing-over. The occurrence of tandem duplication was observed in some genes encoding subunit C, including *GmNF-YC13*, *GmNF-YC15,* and *GmNF-YC24*. An unequal crossover mechanism should have driven the evolution in this subunit.

Soybean experienced whole genome duplication during their evolution. The high *Ks* values observed in the majority of the syntenic regions indicate that the duplicated blocks are a result of ancient polyploidy events (Kim et al. 2009). During evolution, duplicated genes can undergo

pseudogenization, neofunctionalization or subfunctionalization (Cagliari et al. 2011). A *Ka/Ks* ratio >1 indicates a neofunctionalization process (through positive selection), and a *Ka/Ks* ratio <1 suggests a subfunctionalization process (through negative selection) (Fan et al. 2013). In this study, the *Ka/Ks* ratio <1 suggests a subfunctionalization process can be acting in *GmNF-Y* sequences. In soybean, it has been shown that several paralogous genes are undergoing subfunctionalization (Fan et al. 2013). The different expression patterns of *GmNF-Y* genes in different tissues and stresses, associated with the *Ka/Ks* ratio, reinforce the occurrence of subfunctionalization.

It is know that members of plant NF-Y family are involved with several processes, as development and biotic stress. NF-YB2 confers drought tolerance and leads to improved corn yields in transgenic *Z. mays* (Ma et al. 2015a; Ma et al. 2015b; Nelson et al. 2007). In *A. thaliana*, the overexpression of *NF-YA5* and *NF-YB1* reduces water loss in leaves and improves drought tolerance. In addition, ectopic expression of NF-YC from *Amaranthus hypochondriacus* and NF-YB from *Picea wilsonii* confers resistance to water deficiency in *Arabidopsis* (Palmeros-Suarez et al. 2015; Zhang et al. 2015b). In foxtail millet, *NF-YA1* and *NF-YB8* were highly activated in leaves and/or roots by drought and salt stresses and were induced under abscisic acid and H_2O_2 treatments (Feng et al. 2015). In rice, the overexpression of *NF-YA7* improved drought tolerance (Lee et al. 2015) while the overexpression of *HAP2E* confers resistance to pathogens, salinity, and drought (Alam et al. 2015).

It was previously described that *NF-YA3* of soybean is induced by various stress treatments, and their overexpression in *Arabidopsis* reduced leaf water loss and enhanced drought tolerance (Ni et al. 2013). Otherwise, the GmNF-Y family is poorly explored. Our study characterized for the first time that *GmNF-Y* genes are not ubiquitously expressed in soybean organs; their expression are instead organ-dependent. Our results demonstrate for the first time that *GmNF-Y* genes were modulated under *P. pachyrhizi* fungi, representing a new characterization for these genes (Figure 3). *Phakopsora pachyrhizi* fungi is one of the most important pathogens that affect soybean production, responsible for the development of Asian Soybean Rust (ASR) disease. The pathogen attacks leaves, stems, and pods and may defoliate soybean plants in a few days, leading to drastic crop losses. Identifying genes involved in susceptible or resistant response and characterizing their individual roles are key steps for engineering soybean resistant soybean plants. Thus, additional experiments are necessary to better understand the function of these genes in response to ASR.

An interaction between soybean plant and symbiotic soil bacteria, started by the infection of the plant root hair cells by symbiont, results in the nodulation process. In this study, especially after soil bacteria inoculation, the expression of some genes was induced in nodule (Figure 3), indicating these genes could play an important role in soybean nodule association besides plant/pathogen interaction. Recently, a phylogenetically conserved group of NF-Y, which interacts to control nodulation, was described (Baudin et al. 2015). MtNF-YB16 and MtNF-YC1/MtNF-YC2 interact with MtNF-YA1 and MtNF-YA2 to form NF-Y trimers in yeast and *in planta*. Moreover, a similar trimer was formed in common bean (Baudin et al. 2015). A number of independent NF-Y subunits had been reported to participate in nitrogen-fixing rhizobia symbiosis. This includes the *Medicago truncatula* NF-YA1 and A2 (Combier et al. 2006; Laloum et al. 2014; Laporte et al. 2014), NF-YA1 and NF-YB1 from *Lotus japonicus* (Soyano et al. 2013), and NF-YC1 from common bean (Zanetti et al. 2010). In soybean, nodulation and nodule activity are strongly reduced under stress conditions (Sinclair TR 2007). Thus, it is important to note that the modulation of NF-Y genes in both conditions (stresses and nodulation) is an interestingly characteristic to be deep explored, clarifying the importance of this family to soybean improvement.

Conclusion

Altogether, the results presented here have demonstrated that segmental duplication contributed significantly to the expansion of this gene family, and the subfuctionalization process can drive the evolution of duplicated genes. Several *GmNF-Y* members are modulated in soybean nodules and under different stresses. The modulation of *GmNF-Y* expression, especially in response to ASR, suggested their involved in the plant stress response. Future experiments might be important to clarify the relationship between *GmNF-Y* and the transcription regulation of physiological processes that culminate with the resistance/tolerance of soybean plants to biotic/abiotic stress conditions.

The authors declare that they have no conflicts of interest.

This work was supported by grants from the CNPq.

References

Alam MM, Tanaka T, Nakamura H, Ichikawa H, Kobayashi K, Yaeno T, Yamaoka N, Shimomoto K, Takayama K, Nishina H, Nishiguchi M (2015) Overexpression of a rice heme activator protein gene (OsHAP2E) confers resistance to pathogens, salinity and drought, and increases photosynthesis and tiller number. Plant Biotechnol J 13: 85-96

Ballif J, Endo S, Kotani M, MacAdam J, Wu Y (2011) Over-expression of HAP3b enhances primary root elongation in Arabidopsis. Plant Physiol Biochem 49: 579-583

Baudin M, Laloum T, Lepage A, Ripodas C, Ariel F, Frances L, Crespi M, Gamas P, Blanco FA, Zanetti ME, de Carvalho-Niebel F, Niebel A (2015) A Phylogenetically Conserved Group of Nuclear Factor-Y Transcription Factors Interact to Control Nodulation in Legumes. Plant physiology 169: 2761-2773

Cabreira C, Cagliari A, Bucker-Neto L, Margis-Pinheiro M, de Freitas LB, Bodanese-Zanettini MH (2015) The phylogeny and evolutionary history of the Lesion Simulating Disease (LSD) gene family in Viridiplantae. Mol Genet Genomics 290: 2107-2119

Cagliari A, Turchetto-Zolet AC, Korbes AP, Maraschin Fdos S, Margis R, Margis-Pinheiro M (2014) New insights on the evolution of Leafy cotyledon1 (LEC1) type genes in vascular plants. Genomics 103: 380-387

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: 1-26

Cagliari AT-Z, A. C.; Maraschin, F. S.; Loss, G.; Margis, R. and Margis-Pinheiro, M. (2011) The Evolutionary History of CBF Transcription Factors: Gene Duplication of CCAAT – Binding Factors NF-Y in Plants. Gene duplication. Edited by Felix Friedberg

Cannon SB, Mitra A, Baumgarten A, Young ND, May G (2004) The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. BMC Plant Biol 4: 10

Cao S, Kumimoto RW, Siriwardana CL, Risinger JR, Holt BF, 3rd (2011) Identification and characterization of NF-Y transcription factor families in the monocot model plant Brachypodium distachyon. PLoS One 6: e21805

Combier JP, Frugier F, de Billy F, Boualem A, El-Yahyaoui F, Moreau S, Vernie T, Ott T, Gamas P, Crespi M, Niebel A (2006) MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in Medicago truncatula. Genes Dev 20: 3084-3088

Currie RA (1998) Biochemical characterization of the NF-Y transcription factor complex during B lymphocyte development. J Biol Chem 273: 18220-18229

Dolfini D, Zambelli F, Pedrazzoli M, Mantovani R, Pavesi G (2016) A high definition look at the NF-Y regulome reveals genome-wide associations with selected transcription factors. Nucleic Acids Res

Edwards D, Murray JA, Smith AG (1998) Multiple genes encoding the conserved CCAAT-box transcription factor complex are expressed in Arabidopsis. Plant physiology 117: 1015-1022

Fan C, Wang X, Hu R, Wang Y, Xiao C, Jiang Y, Zhang X, Zheng C, Fu YF The pattern of Phosphate transporter 1 genes evolutionary divergence in Glycine max L.

BMC Plant Biol 13: 48
Fan C, Wang X, Hu R, Wang Y, Xiao C, Jiang Y, Zhang X, Zheng C, Fu YF (2013) The pattern of Phosphate transporter 1 genes evolutionary divergence in Glycine max L. BMC plant biology 13: 48
Feng ZJ, He GH, Zheng WJ, Lu PP, Chen M, Gong YM, Ma YZ, Xu ZS (2015) Foxtail Millet NF-Y Families: Genome-Wide Survey and Evolution Analyses Identified Two Functional Genes Important in Abiotic Stresses. Front Plant Sci 6: 1142
Fornari M, Calvenzani V, Masiero S, Tonelli C, Petroni K (2013) The Arabidopsis NF-YA3 and NF-YA8 genes are functionally redundant and are required in early embryogenesis. PloS one 8: e82043
Gusmaroli G, Tonelli C, Mantovani R (2002) Regulation of novel members of the Arabidopsis thaliana CCAAT-binding nuclear factor Y subunits. Gene 283: 41-48
Hackenberg D, Keetman U, Grimm B (2012) Homologous NF-YC2 subunit from Arabidopsis and tobacco is activated by photooxidative stress and induces flowering. Int J Mol Sci 13: 3458-3477
Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W, Zimmermann P (2008) Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. Adv Bioinformatics 2008: 420747
Huang HS, Chen CJ, Chang WC (1999) The CCAAT-box binding factor NF-Y is required for the expression of phospholipid hydroperoxide glutathione peroxidase in human epidermoid carcinoma A431 cells. FEBS Lett 455: 111-116
Hwang YH, Kim SK, Lee KC, Chung YS, Lee JH, Kim JK (2016) Functional conservation of rice OsNF-YB/YC and Arabidopsis AtNF-YB/YC proteins in the regulation of flowering time. Plant Cell Rep 35: 857-865
Jami SK, Clark GB, Ayele BT, Ashe P, Kirti PB Genome-wide comparative analysis of annexin superfamily in plants. PLoS One 7: e47801
Jami SK, Clark GB, Ayele BT, Ashe P, Kirti PB (2012) Genome-wide comparative analysis of annexin superfamily in plants. PloS one 7: e47801
Kim KD, Shin JH, Van K, Kim DH, Lee SH (2009) Dynamic rearrangements determine genome organization and useful traits in soybean. Plant physiology 151: 1066-1076
Kim SK, Park HY, Jang YH, Lee KC, Chung YS, Lee JH, Kim JK (2016) OsNF-YC2 and OsNF-YC4 proteins inhibit flowering under long-day conditions in rice. Planta 243: 563-576
Kumimoto RW, Adam L, Hymus GJ, Repetti PP, Reuber TL, Marion CM, Hempel FD, Ratcliffe OJ (2008) The Nuclear Factor Y subunits NF-YB2 and NF-YB3 play additive roles in the promotion of flowering by inductive long-day photoperiods in Arabidopsis. Planta 228: 709-723
Kumimoto RW, Zhang Y, Siefers N, Holt BF, 3rd (2010) NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in Arabidopsis thaliana. The Plant journal : for cell and molecular biology 63: 379-391
Laloum T, Baudin M, Frances L, Lepage A, Billault-Penneteau B, Cerri MR, Ariel F, Jardinaud MF, Gamas P, de Carvalho-Niebel F, Niebel A (2014) Two CCAAT-box-binding transcription factors redundantly regulate early steps of the legume-rhizobia endosymbiosis. Plant J 79: 757-768
Laporte P, Lepage A, Fournier J, Catrice O, Moreau S, Jardinaud MF, Mun JH,

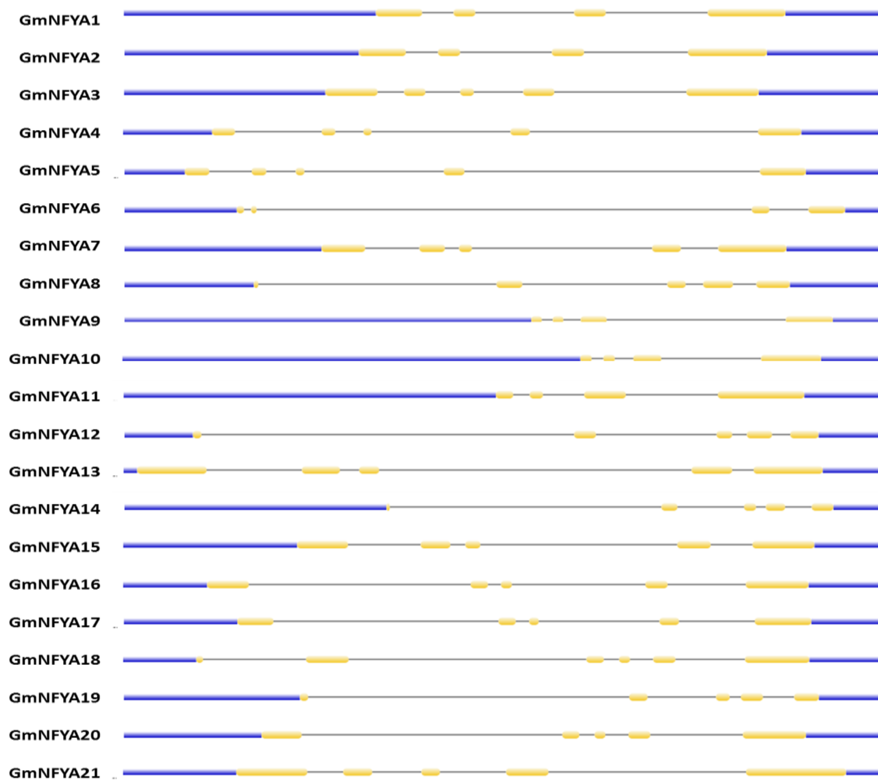
Larrainzar E, Cook DR, Gamas P, Niebel A (2014) The CCAAT box-binding transcription factor NF-YA1 controls rhizobial infection. J Exp Bot 65: 481-494
Lee DK, Kim HI, Jang G, Chung PJ, Jeong JS, Kim YS, Bang SW, Jung H, Choi YD, Kim JK (2015) The NF-YA transcription factor OsNF-YA7 confers drought stress tolerance of rice in an abscisic acid independent manner. Plant Sci 241: 199-210
Lee TH, Tang H, Wang X, Paterson AH (2013) PGDD: a database of gene and genome duplication in plants. Nucleic Acids Res 41: D1152-1158
Li S, Li K, Ju Z, Cao D, Fu D, Zhu H, Zhu B, Luo Y (2016) Genome-wide analysis of tomato NF-Y factors and their role in fruit ripening. BMC Genomics 17: 36
Liu JX, Howell SH (2010) bZIP28 and NF-Y transcription factors are activated by ER stress and assemble into a transcriptional complex to regulate stress response genes in Arabidopsis. Plant Cell 22: 782-796
Ma X, Li C, Wang M (2015a) Wheat NF-YA10 functions independently in salinity and drought stress. Bioengineered 6: 245-247
Ma X, Zhu X, Li C, Song Y, Zhang W, Xia G, Wang M (2015b) Overexpression of wheat NF-YA10 gene regulates the salinity stress response in Arabidopsis thaliana. Plant Physiol Biochem 86: 34-43
Maity SN, de Crombrugghe B (1992) Biochemical analysis of the B subunit of the heteromeric CCAAT-binding factor. A DNA-binding domain and a subunit interaction domain are specified by two separate segments. J Biol Chem 267: 8286-8292
Maity SN, de Crombrugghe B (1998) Role of the CCAAT-binding protein CBF/NF-Y in transcription. Trends Biochem Sci 23: 174-178
Mantovani R (1999) The molecular biology of the CCAAT-binding factor NF-Y. Gene 239: 15-27
Matuoka K, Chen KY (2002) Transcriptional regulation of cellular ageing by the CCAAT box-binding factor CBF/NF-Y. Ageing Res Rev 1: 639-651
Matuoka K, Yu Chen K (1999) Nuclear factor Y (NF-Y) and cellular senescence. Exp Cell Res 253: 365-371
McNabb DS, Xing Y, Guarente L (1995) Cloning of yeast HAP5: a novel subunit of a heterotrimeric complex required for CCAAT binding. Genes Dev 9: 47-58
Mei X, Liu C, Yu T, Liu X, Xu D, Wang J, Wang G, Cai Y (2015) Identification and characterization of paternal-preferentially expressed gene NF-YC8 in maize endosperm. Mol Genet Genomics 290: 1819-1831
Mu J, Tan H, Hong S, Liang Y, Zuo J (2013) Arabidopsis transcription factor genes NF-YA1, 5, 6, and 9 play redundant roles in male gametogenesis, embryogenesis, and seed development. Mol Plant 6: 188-201
Nardini M, Gnesutta N, Donati G, Gatta R, Forni C, Fossati A, Vonrhein C, Moras D, Romier C, Bolognesi M, Mantovani R (2013) Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. Cell 152: 132-143
Nelson DE, Repetti PP, Adams TR, Creelman RA, Wu J, Warner DC, Anstrom DC, Bensen RJ, Castiglioni PP, Donnarummo MG, Hinchey BS, Kumimoto RW, Maszle DR, Canales RD, Krolikowski KA, Dotson SB, Gutterson N, Ratcliffe OJ, Heard JE (2007) Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. Proc Natl Acad Sci U S A 104: 16450-16455

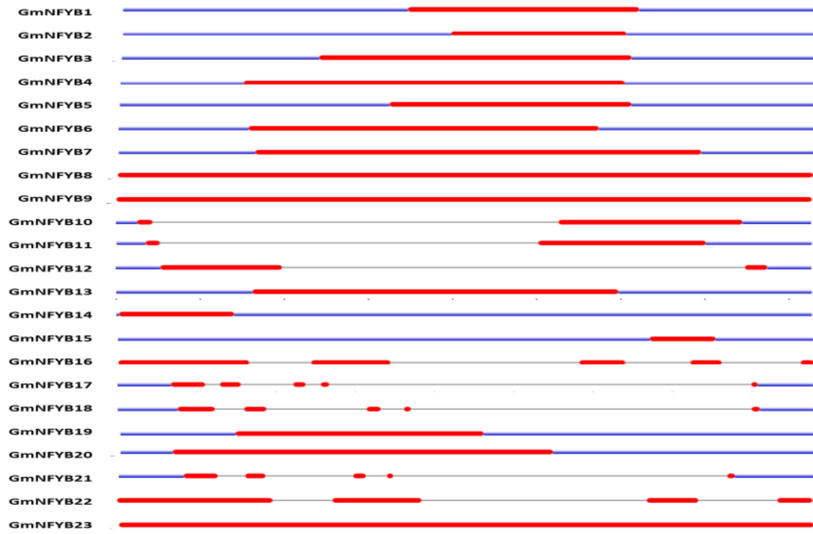
Ni Z, Hu Z, Jiang Q, Zhang H (2013) GmNFYA3, a target gene of miR169, is a positive regulator of plant tolerance to drought stress. Plant Mol Biol 82: 113-129
Oldfield AJ, Yang P, Conway AE, Cinghu S, Freudenberg JM, Yellaboina S, Jothi R (2014) Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. Mol Cell 55: 708-722
Palmeros-Suarez PA, Massange-Sanchez JA, Martinez-Gallardo NA, Montero-Vargas JM, Gomez-Leyva JF, Delano-Frier JP (2015) The overexpression of an Amaranthus hypochondriacus NF-YC gene modifies growth and confers water deficit stress resistance in Arabidopsis. Plant Sci 240: 25-40
Potkar R, Recla J, Busov V (2013) ptr-MIR169 is a posttranscriptional repressor of PtrHAP2 during vegetative bud dormancy period of aspen (Populus tremuloides) trees. Biochem Biophys Res Commun 431: 512-518
Quach TN, Nguyen HT, Valliyodan B, Joshi T, Xu D, Nguyen HT (2015) Genome-wide expression analysis of soybean NF-Y genes reveals potential function in development and drought response. Mol Genet Genomics 290: 1095-1115
Riechmann JL, Ratcliffe OJ (2000) A genomic perspective on plant transcription factors. Curr Opin Plant Biol 3: 423-434
Roulin A, Auer PL, Libault M, Schlueter J, Farmer A, May G, Stacey G, Doerge RW, Jackson SA The fate of duplicated genes in a polyploid plant genome. Plant J
Roulin A, Auer PL, Libault M, Schlueter J, Farmer A, May G, Stacey G, Doerge RW, Jackson SA (2013) The fate of duplicated genes in a polyploid plant genome. The Plant journal : for cell and molecular biology 73: 143-153
Siefers N, Dang KK, Kumimoto RW, Bynum WEt, Tayrose G, Holt BF, 3rd (2009) Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. Plant physiology 149: 625-641
Sinclair TR PL, King CA, Sneller CH, Chen P, Vadez V. (2007) Drought tolerance and yield increase of soybean resulting from improved symbiotic N2 fixation. Field Crops Res 101: 68-71
Sinha S, Maity SN, Lu J, de Crombrugghe B (1995) Recombinant rat CBF-C, the third subunit of CBF/NFY, allows formation of a protein-DNA complex with CBF-A and CBF-B and with yeast HAP2 and HAP3. Proc Natl Acad Sci U S A 92: 1624-1628
Siriwardana CL, Kumimoto RW, Jones DS, Holt BF, 3rd (2014) Gene Family Analysis of the Transcription Factors Reveals Opposing Abscisic Acid Responses During Seed Germination. Plant Mol Biol Report 32: 971-986
Sorin C, Declerck M, Christ A, Blein T, Ma L, Lelandais-Briere C, Njo MF, Beeckman T, Crespi M, Hartmann C (2014) A miR169 isoform regulates specific NF-YA targets and root architecture in Arabidopsis. New Phytol 202: 1197-1211
Soyano T, Kouchi H, Hirota A, Hayashi M (2013) Nodule inception directly targets NF-Y subunit genes to regulate essential processes of root nodule development in Lotus japonicus. PLoS Genet 9: e1003352
Stephenson TJ, McIntyre CL, Collet C, Xue GP (2007) Genome-wide identification and expression analysis of the NF-Y family of transcription factors in Triticum aestivum. Plant Mol Biol 65: 77-92
Stephenson TJ, McIntyre CL, Collet C, Xue GP (2010) TaNF-YC11, one of the light-upregulated NF-YC members in Triticum aestivum, is co-regulated with photosynthesis-related genes. Funct Integr Genomics 10: 265-276

Stephenson TJ, McIntyre CL, Collet C, Xue GP (2011) TaNF-YB3 is involved in the regulation of photosynthesis genes in Triticum aestivum. Funct Integr Genomics 11: 327-340
Sun X, Ling S, Lu Z, Ouyang YD, Liu S, Yao J (2014) OsNF-YB1, a rice endosperm-specific gene, is essential for cell proliferation in endosperm development. Gene 551: 214-221
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30: 2725-2729
Thirumurugan T, Ito Y, Kubo T, Serizawa A, Kurata N (2008) Identification, characterization and interaction of HAP family genes in rice. Mol Genet Genomics 279: 279-289
Toufighi K, Brady SM, Austin R, Ly E, Provart NJ (2005) The Botany Array Resource: e-Northerns, Expression Angling, and promoter analyses. Plant J 43: 153-163
Xu L, Lin Z, Tao Q, Liang M, Zhao G, Yin X, Fu R (2014) Multiple NUCLEAR FACTOR Y transcription factors respond to abiotic stress in Brassica napus L. PloS one 9: e111354
Yadav D, Shavrukov Y, Bazanova N, Chirkova L, Borisjuk N, Kovalchuk N, Ismagul A, Parent B, Langridge P, Hrmova M, Lopato S (2015) Constitutive overexpression of the TaNF-YB4 gene in transgenic wheat significantly improves grain yield. Journal of experimental botany 66: 6635-6650
Yamamoto A, Kagaya Y, Toyoshima R, Kagaya M, Takeda S, Hattori T (2009) Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. The Plant journal : for cell and molecular biology 58: 843-856
Yin G, Xu H, Xiao S, Qin Y, Li Y, Yan Y, Hu Y The large soybean (Glycine max) WRKY TF family expanded by segmental duplication events and subsequent divergent selection among subgroups. BMC Plant Biol 13: 148
Yin G, Xu H, Xiao S, Qin Y, Li Y, Yan Y, Hu Y (2013) The large soybean (Glycine max) WRKY TF family expanded by segmental duplication events and subsequent divergent selection among subgroups. BMC plant biology 13: 148
Zanetti ME, Blanco FA, Beker MP, Battaglia M, Aguilar OM (2010) A C subunit of the plant nuclear factor NF-Y required for rhizobial infection and nodule development affects partner selection in the common bean-Rhizobium etli symbiosis. Plant Cell 22: 4142-4157
Zhang F, Han M, Lv Q, Bao F, He Y (2015a) Identification and expression profile analysis of NUCLEAR FACTOR-Y families in Physcomitrella patens. Front Plant Sci 6: 642
Zhang L, Vision TJ, Gaut BS (2002) Patterns of nucleotide substitution among simultaneously duplicated gene pairs in Arabidopsis thaliana. Mol Biol Evol 19: 1464-1473
Zhang T, Zhang D, Liu Y, Luo C, Zhou Y, Zhang L (2015b) Overexpression of a NF-YB3 transcription factor from Picea wilsonii confers tolerance to salinity and drought stress in transformed Arabidopsis thaliana. Plant Physiol Biochem 94: 153-164

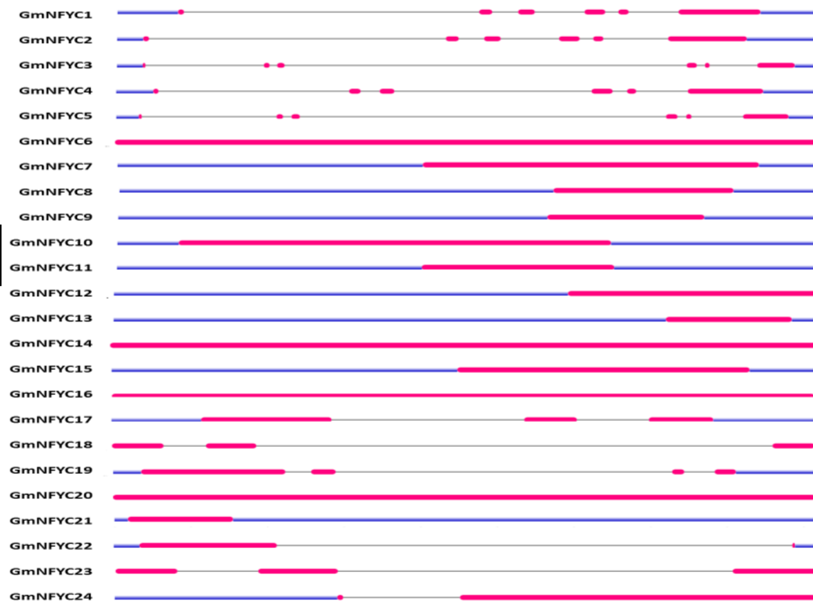
Legends and figures

GmNF-YA



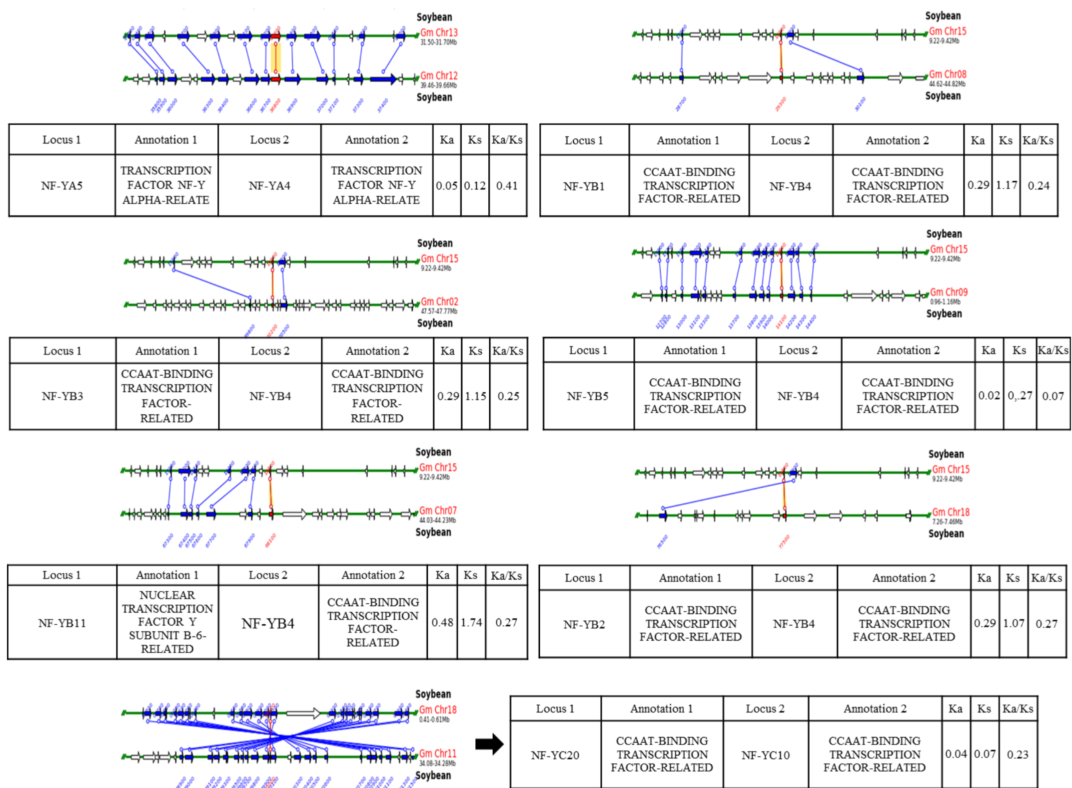


GmNF-YB



GmNF-YC

Figure 1 . *GmNF-Y* genes structure and onservation of NF-Y domains. The total height of each cell indicates the conserved sequence at each position. The height of each letter is proportional to the corresponding relative frequency. The amino acids are colored according to their chemical properties.



Locus 1	Annotation 1	Locus 2	Annotation 2	Ka	Ks	Ka/Ks
NF-YA5	TRANSCRIPTION FACTOR NF-Y ALPHA-RELATE	NF-YA4	TRANSCRIPTION FACTOR NF-Y ALPHA-RELATE	0.05	0.12	0.41

Locus 1	Annotation 1	Locus 2	Annotation 2	Ka	Ks	Ka/Ks
NF-YB1	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	NF-YB4	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	0.29	1.17	0.24

Locus 1	Annotation 1	Locus 2	Annotation 2	Ka	Ks	Ka/Ks
NF-YB3	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	NF-YB4	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	0.29	1.15	0.25

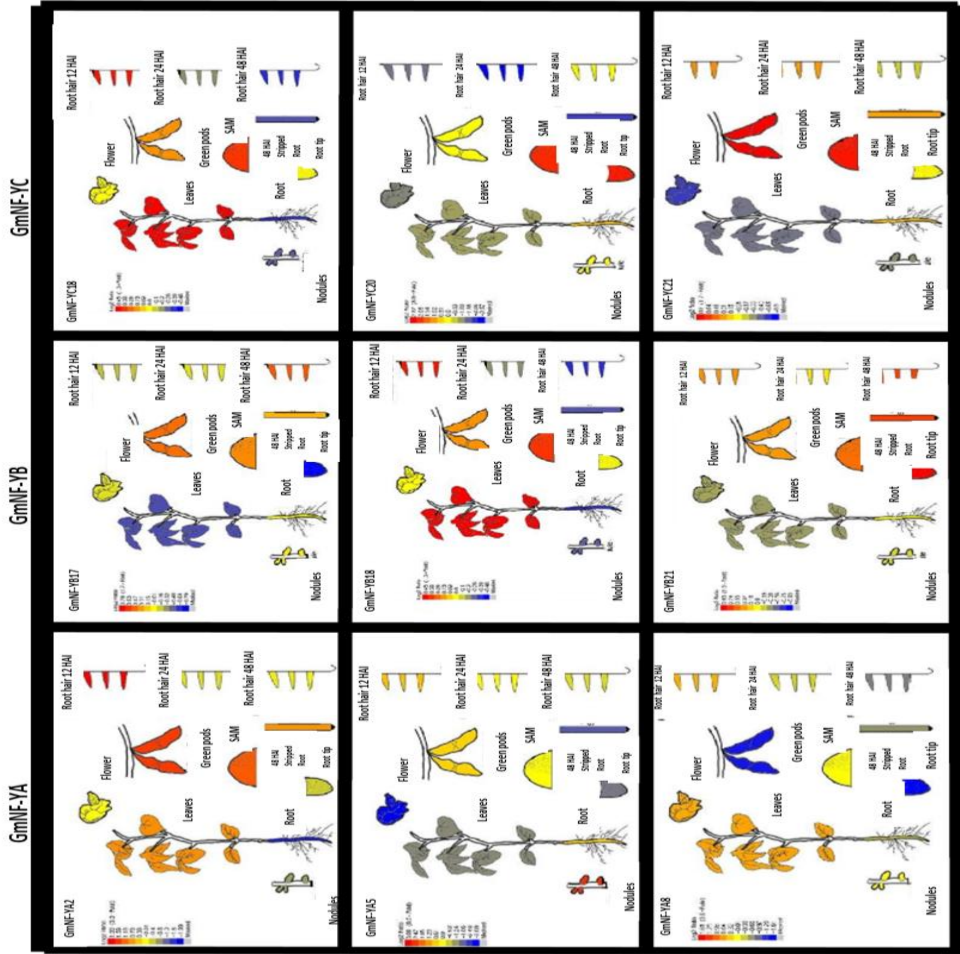
Locus 1	Annotation 1	Locus 2	Annotation 2	Ka	Ks	Ka/Ks
NF-YB5	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	NF-YB4	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	0.02	0,.27	0.07

Locus 1	Annotation 1	Locus 2	Annotation 2	Ka	Ks	Ka/Ks
NF-YB11	NUCLEAR TRANSCRIPTION FACTOR Y SUBUNIT B-6-RELATED	NF-YB4	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	0.48	1.74	0.27

Locus 1	Annotation 1	Locus 2	Annotation 2	Ka	Ks	Ka/Ks
NF-YB2	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	NF-YB4	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	0.29	1.07	0.27

Locus 1	Annotation 1	Locus 2	Annotation 2	Ka	Ks	Ka/Ks
NF-YC20	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	NF-YC10	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	0.04	0.07	0.23

Figure 2. Syntenic regions between *GmNF-Y* genes analyzed using the PGDD database. The green horizontal line represents the chromosome. The blue and red vertical lines and arrows represent the duplicated orthologous; the red line represents the sequence that was used as the search query. The white arrow indicates genes without correspondent paralogous.



A

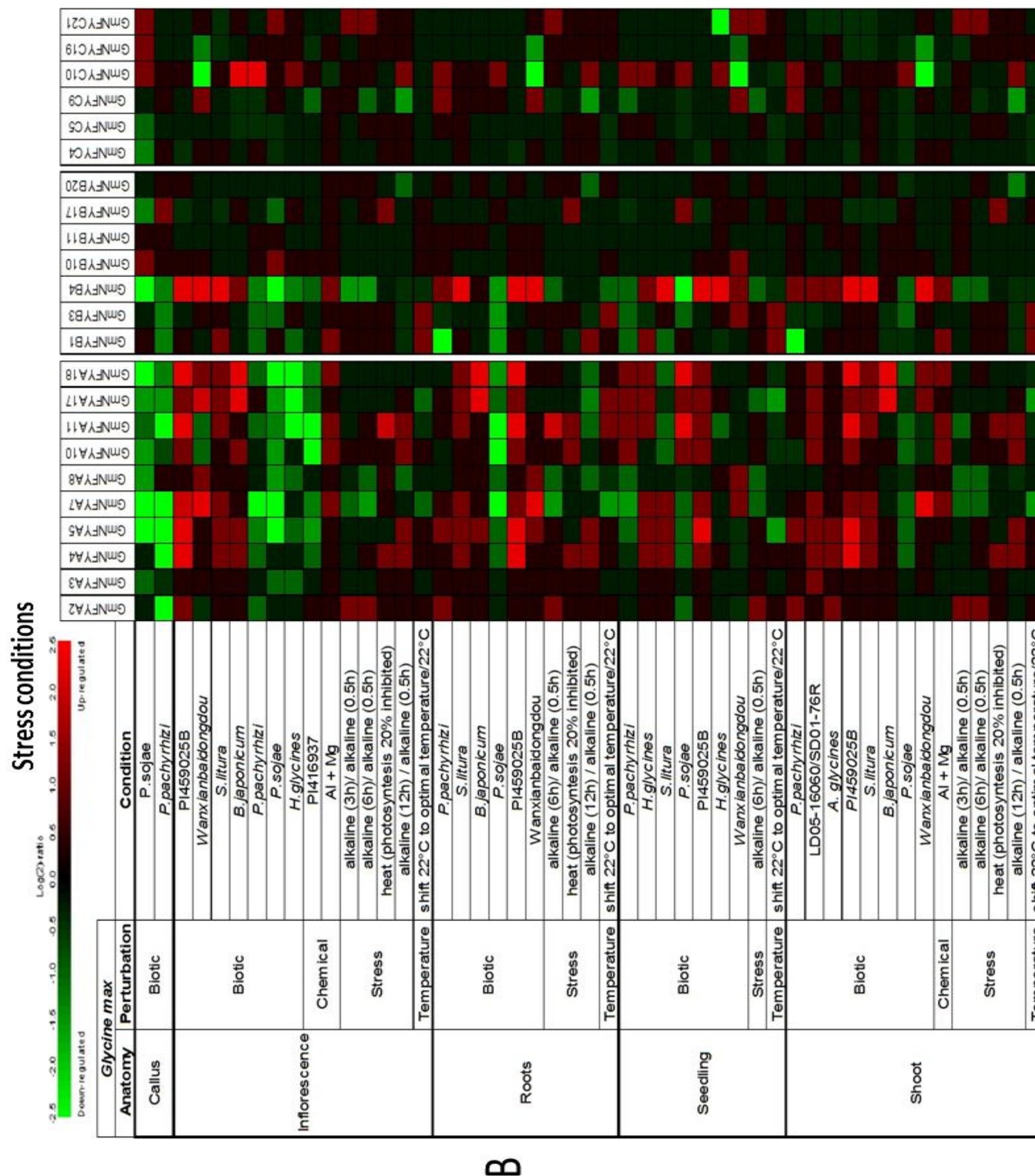


Figure 3. A) Expression profile of *GmNF-YA, GmNF-YB* and *GmNF-YC* genes in leaves, flowers, green pods, SAM (Shoot Apical Meristem), roots, root tips, nodules, and roots inoculated with *Bradyrhizobium japonicum*. Red color indicates up- and blue indicates downregulated genes (Log 2 ratio). HAI: Hours After Inoculation. B) Expression profile of *GmNF-Y* genes under different stress conditions. Red color indicates up- and green indicates downregulated genes (Log 2 ratio).

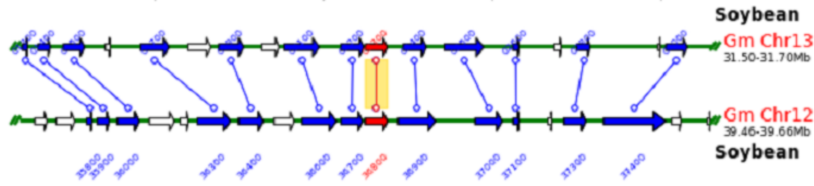
Acronym	Gene ID	WG Mapping Analysis (PLAZA database)
GmNFYA1	Glyma.14G010000	Block duplicated
GmNFYA2	Glyma.02G303800	Block duplicated
GmNFYA3	Glyma.08G335900	Block duplicated
GmNFYA4	Glyma.13G202300	Block duplicated
GmNFYA5	Glyma.12G236800	Block duplicated
GmNFYA6	Glyma.15G173300	Block duplicated
GmNFYA7	Glyma.07G036200	Block duplicated
GmNFYA8	Glyma.08G124200	Block duplicated
GmNFYA9	Glyma.17G051400	Block duplicated
GmNFYA10	Glyma.13G107900	Block duplicated
GmNFYA11	Glyma.09G068400	Block duplicated
GmNFYA12	Glyma.05G166100	Block duplicated
GmNFYA13	Glyma.15G027400	Block duplicated
GmNFYA14	Glyma.15G129900	Block duplicated
GmNFYA15	Glyma.16G005500	Block duplicated
GmNFYA16	Glyma.10G082800	Block duplicated
GmNFYA17	Glyma.02G195000	Block duplicated
GmNFYA18	Glyma.19G200800	Block duplicated
GmNFYA19	Glyma.09G023800	Block duplicated
GmNFYA20	Glyma.03G203000	Block duplicated
GmNFYA21	Glyma.18G071000	Block duplicated
GmNFYB1	Glyma.08G329300	Block duplicated
GmNFYB2	Glyma.18G077500	Block duplicated
GmNFYB3	Glyma.02G300200	Block duplicated
GmNFYB4	Glyma.15G118800	Block duplicated
GmNFYB5	Glyma.09G014100	Block duplicated
GmNFYB6	Glyma.08G141000	Block duplicated
GmNFYB7	Glyma.05G183200	Block duplicated
GmNFYB8	Glyma.05G193300	Block duplicated
GmNFYB9	Glyma.08G001300	Block duplicated
GmNFYB10	Glyma.17G005600	Block duplicated
GmNFYB11	Glyma.07G268100	Block duplicated
GmNFYB12	Glyma.09G046200	Block duplicated

GmNFYB13	Glyma.15G153900	Block duplicated
GmNFYB14	Glyma.10G153500	Block duplicated
GmNFYB15	Glyma.20G234900	Block duplicated
GmNFYB16	Glyma.19G178500	Block duplicated
GmNFYB17	Glyma.03G177700	Block duplicated
GmNFYB18	Glyma.20G198500	Block duplicated
GmNFYB19	Glyma.07G249000	Block duplicated
GmNFYB20	Glyma.02G154000	Block duplicated
GmNFYB21	Glyma.10G192000	Block duplicated
GmNFYB22	Glyma.10G048900	Block duplicated
GmNFYB23	Glyma.17G025300	Block duplicated
GmNFYC1	Glyma.02G277000	Block duplicated
GmNFYC2	Glyma.14G038800	Block duplicated
GmNFYC3	Glyma.13G189400	Block duplicated
GmNFYC4	Glyma.06G311400	Block duplicated
GmNFYC5	Glyma.15G227300	Block duplicated
GmNFYC6	Glyma.20G232400	Block duplicated
GmNFYC7	Glyma.10G155900	Block duplicated
GmNFYC8	Glyma.15G261300	Block duplicated
GmNFYC9	Glyma.08G165700	Block duplicated
GmNFYC10	Glyma.18G007100	Block duplicated
GmNFYC11	Glyma.19G236400	Block duplicated
GmNFYC12	Glyma.03G239400	Block duplicated
GmNFYC13	Glyma.13G207700	Tandem duplicated
GmNFYC14	Glyma.08G148200	No duplicated
GmNFYC15	Glyma.13G207500	Tandem duplicated
GmNFYC16	Glyma.13G284000	Block duplicated
GmNFYC17	Glyma.02G089600	No duplicated
GmNFYC18	Glyma.12G069200	Tandem and Block duplicated
GmNFYC19	Glyma.11G148000	Block duplicated
GmNFYC20	Glyma.11G250000	Block duplicated
GmNFYC21	Glyma.06G169600	Block duplicated
GmNFYC22	Glyma.04G196200	Block duplicated
GmNFYC23	Glyma.06G038200	No duplicated
GmNFYC24	Glyma.13G207600	Tandem duplicated

ESM 1. Detailed data on gene acronym, locus, and duplication type of *GmNF-Y*

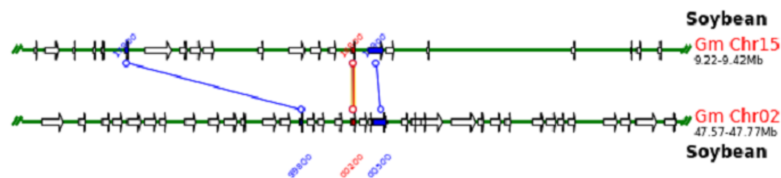
genes.

GmNF-YA4 and GmNF-YA5



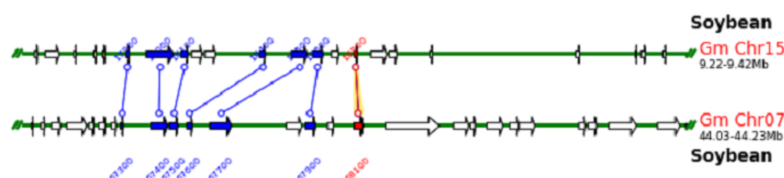
Locus 1	Annotation 1	Locus 2	Annotation 2	Ka	Ks	Ka/Ks
Glyma.12G235800	Dehydrin	Glyma.13G201300	Dehydrin	0.09	0.23	0.39
Glyma.12G235900	Protein tyrosine kinase (Pkinase Tyr) /Di-glucose binding within endoplasmic reticulum (Malectin)	Glyma.13G201400	Protein tyrosine kinase (Pkinase Tyr) /Di-glucose binding within endoplasmic reticulum (Malectin)	0.02	0.19	0.10
Glyma.12G236000	SUMO-ACTIVATING ENZYME SUBUNIT 2	Glyma.13G201500	SUMO-ACTIVATING ENZYME SUBUNIT 2	0.01	0.11	0.09
Glyma.12G236300	EXPRESSED PROTEIN	Glyma.13G201700	EXPRESSED PROTEIN	0.04	0.11	0.36
Glyma.12G236400	snRNP-specific protein-like factor and related proteins // G-protein beta subunit-like protein	Glyma.13G201900	snRNP-specific protein-like factor and related proteins // G-protein beta subunit-like protein (contains WD40 repeats)	0.02	0.12	0.16
Glyma.12G236600	SMC proteins	Glyma.13G202100	SMC proteins	0.01	0.08	0.12
Glyma.12G236700	CAROTENOID 9,10(9',10')-CLEAVAGE DIOXYGENASE 1	Glyma.13G202200	CAROTENOID 9,10(9',10')-CLEAVAGE DIOXYGENASE 1	0.03	0.12	0.25
Glyma.12G236800	TRANSCRIPTION FACTOR NF-Y ALPHA-RELATE	Glyma.13G202300	TRANSCRIPTION FACTOR NF-Y ALPHA-RELATE	0.05	0.12	0.41
Glyma.12G236900	ACID PHOSPHATASE-RELATED	Glyma.13G202400	ACID PHOSPHATASE-RELATED	0.02	0.08	0.25
Glyma.12G237000	X-BOX TRANSCRIPTION FACTOR-RELATED	Glyma.13G202500	X-BOX TRANSCRIPTION FACTOR-RELATED	0.01	0.13	0.07
Glyma.12G237100	No annotation	Glyma.13G202600	No annotation	0.05	0.15	0.33
Glyma.12G237300	ZINC FINGER CCCH DOMAIN-CONTAINING PROTEIN 56	Glyma.13G202800	ZINC FINGER CCCH DOMAIN-CONTAINING PROTEIN 56	0.03	0.19	0.15
Glyma.12G237400	MATE EFFLUX FAMILY PROTEIN FRD3 (MULTIDRUG RESISTANCE PROTEIN)	Glyma.13G203000	MATE EFFLUX FAMILY PROTEIN FRD3 (MULTIDRUG RESISTANCE PROTEIN)	0.20	0.53	0.37

GmNF-YB4 and GmNF-YB3



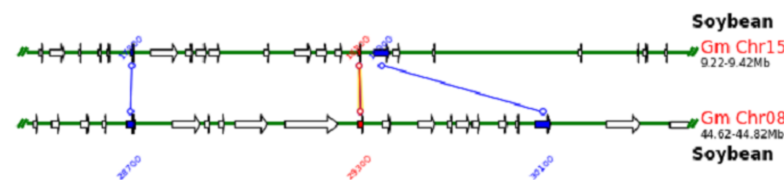
Locus 1	Annotation 1	Locus 2	Annotation 2	Ka	Ks	Ka/Ks
Glyma.02G299800	No Annotation	Glyma.15G117900	No Annotation	0.60	0.00	
Glyma.02G300200	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	Glyma.15G118800	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	0.29	1.15	0.25
Glyma.02G300500	ADP-ribose diphosphatase / ADPR-PPase // NAD(+) diphosphatase / NADP pyrophosphatase	Glyma.15G118900	ADP-ribose diphosphatase / ADPR-PPase // NAD(+) diphosphatase / NADP pyrophosphatase	0.32	1.41	0.22

GmNF-YB4 and GmNF-YB11



Locus 1	Annotation 1	Locus 2	Annotation 2	Ka	Ks	Ka/Ks
Glyma.07G267300	No annotation	Glyma.15G117900	No annotation	0.15	0.68	0.22
Glyma.07G267400	MICROTUBULE-ASSOCIATED PROTEIN 70-5	Glyma.15G118000	MICROTUBULE-ASSOCIATED PROTEIN 70-5	0.13	0.39	0.33
Glyma.07G267500	PPR repeat (PPR)	Glyma.15G118100	Glyma.07G267500	0.26	0.77	0.33
Glyma.07G267600	COPPER TRANSPORT PROTEIN ATOX1-RELATED	Glyma.15G118400	METAL ION BINDING PROTEIN-RELATED	0.34	1.63	0.20
Glyma.07G267700	COPPER TRANSPORT PROTEIN ATOX1-RELATED	Glyma.15G118500	COPPER TRANSPORT PROTEIN ATOX1-RELATED	0.07	0.52	0.13
Glyma.07G267900	SGMNF1-RELATED PROTEIN KINASE REGULATORY SUBUNIT BETA-1	Glyma.15G118600	SGMNF1-RELATED PROTEIN KINASE REGULATORY SUBUNIT BETA-1	0.22	0.70	0.31
Glyma.07G268100	NUCLEAR TRANSCRIPTION FACTOR Y SUBUNIT B-6-RELATED	Glyma.15G118800	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	0.48	1.74	0.27

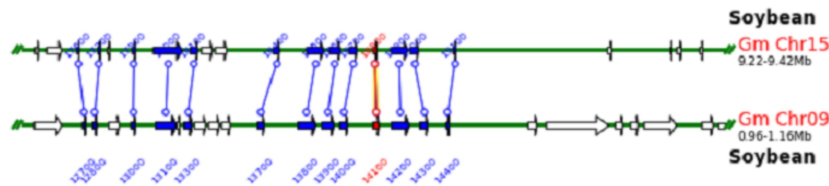
GmNF-YB4 and GmNF-YB1



Locus 1	Annotation 1	Locus 2	Annotation 2	Ka	Ks	Ka/Ks
Glyma.08G328700	No annotation	Glyma.15G117900	No annotation	0.54	0.00	0
Glyma.08G329300	CCAAT-BINDING TRANSCRIPTION	Glyma.15G118800	CCAAT-BINDING TRANSCRIPTION	0.29	1.17	0.24

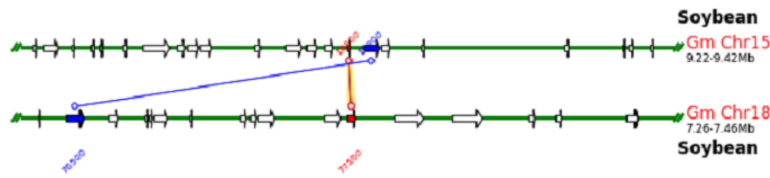
	FACTOR-RELATED		FACTOR-RELATED			
Glyma.08G330100	ADP-ribose diphosphatase / ADPR-PPase // NAD(+) diphosphatase / NADP pyrophosphatase	Glyma.15G118900	ADP-ribose diphosphatase / ADPR-PPase // NAD(+) diphosphatase / NADP pyrophosphatase	0.38	1.19	0.31

GmNF-YB4 and GmNF-YB5



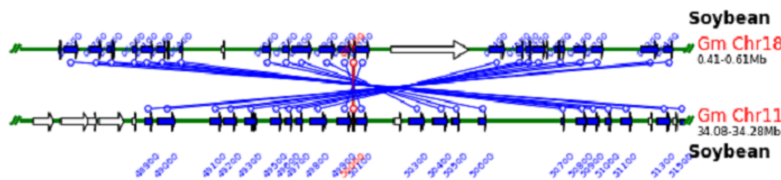
Locus 1	Annotation 1	Locus 2	Annotation 2	Ka	Ks	Ka/K
Glyma.09G012700	No annotation	Glyma.15G117600	No annotation	0.05	0.17	0.29
Glyma.09G012800	CottoGmNFibreexpressedprotein	Glyma.15G117700	No annotation	0.20	0.25	0.8
Glyma.09G013000	No annotation	Glyma.15G117900	No annotation	0.04	0.19	0.21
Glyma.09G013100	MICROTUBULE-ASSOCIATED PROTEIN 70-5	Glyma.15G118000	MICROTUBULE-ASSOCIATED PROTEIN 70-5	0.02	0.09	0.22
Glyma.09G013300	PPR repeat (PPR)	Glyma.15G118100	PPR repeat (PPR	0.04	0.13	0.30
Glyma.09G013700	METAL ION BINDING PROTEIN-RELATED	Glyma.15G118400	METAL ION BINDING PROTEIN-RELATED	0.07	0.35	0.2
Glyma.09G013800	Glyma.09G013800	Glyma.15G118500	COPPER TRANSPORT PROTEIN ATOX1-RELATED	0.02	0.13	0.15
Glyma.09G013900	SGMNF1-RELATED PROTEIN KINASE REGULATORY SUBUNIT BETA-1	Glyma.15G118600	SGMNF1-RELATED PROTEIN KINASE REGULATORY SUBUNIT BETA-1	0.02	0.13	0.15
Glyma.09G014000	ACYL CARRIER PROTEIN 3, MITOCHONDRIAL	Glyma.15G118700	ACYL CARRIER PROTEIN 3, MITOCHONDRIAL	0.05	0.14	0.35
Glyma.09G014100	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	Glyma.15G118800	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	0.02	0.27	0.07
Glyma.09G014200	ADP-ribose diphosphatase / ADPR-PPase // NAD(+) diphosphatase / NADP pyrophosphatase	Glyma.15G118900	ADP-ribose diphosphatase / ADPR-PPase // NAD(+) diphosphatase / NADP pyrophosphatase	0.04	0.09	0.44
Glyma.09G014300	DNA-BINDING PROTEIN-LIKE PROTEIN	Glyma.15G119000	DNA-BINDING PROTEIN-LIKE PROTEIN	0.04	0.12	0.33
Glyma.09G014400	BIFUNCTIONAL INHIBITOR/LIPID-TRANSFER PROTEIN/SEED STORAGE 2S ALBUMIN SUPERFAMILY PROTEIN	Glyma.15G119100	BIFUNCTIONAL INHIBITOR/LIPID-TRANSFER PROTEIN/SEED STORAGE 2S ALBUMIN SUPERFAMILY PROTEIN	0.00	0.14	0

GmNF-YB4 and GmNF-YB2



Locus 1	Annotation 1	Locus 2	Annotation 2	Ka	Ks	Ka/Ks
Glyma.15G118800	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	Glyma.18G077500	CCAAT-BINDING TRANSCRIPTION FACTOR-RELATED	0.29	1.07	0.27
Glyma.15G118900	ADP-ribose diphosphatase / ADPR-PPase // NAD(+) diphosphatase / NADP pyrophosphatase	Glyma.18G076500	ADP-ribose diphosphatase / ADPR-PPase // NAD(+) diphosphatase / NADP pyrophosphatase	0.27	1.28	0.21

GmNF-YC10 and GmNF-YC20

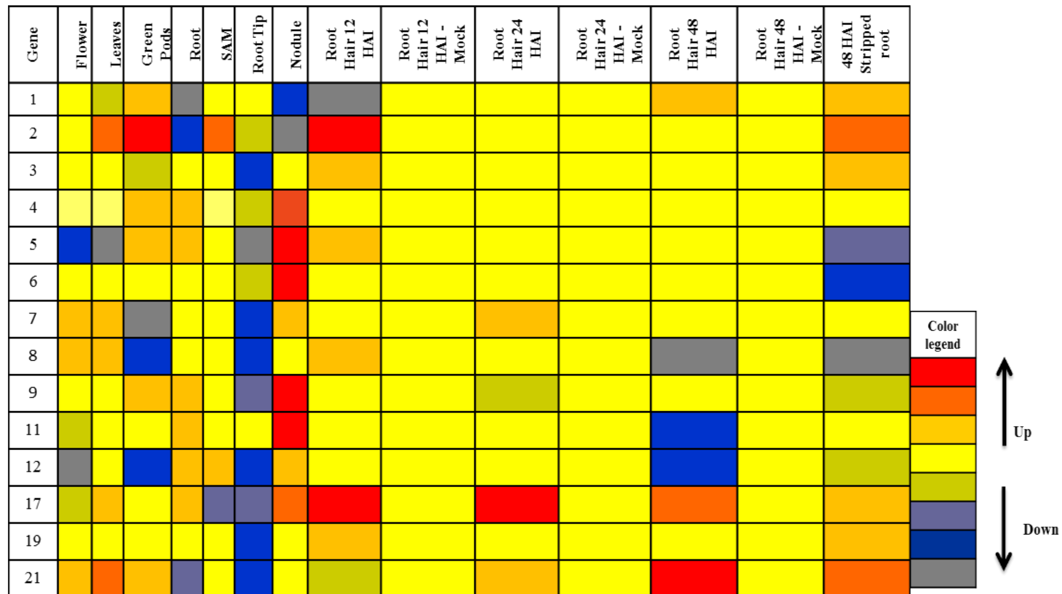


Locus 1	Annotation 1	Locus 2	Annotation 2	Ka	Ks	Ka/Ks
Glyma.11G248900	OTU DOMAIN CONTAINING PROTEIN	Glyma.18G008400	OTU DOMAIN CONTAINING PROTEIN	0.19	0.35	0.54
Glyma.11G249000	U-BOX DOMAIN-CONTAINING PROTEIN 45-RELATED	Glyma.18G008300	U-BOX DOMAIN-CONTAINING PROTEIN 45-RELATED	0.02	0.06	0.33
Glyma.11G249100	No Annotation	Glyma.18G008200	No Annotation	0.00	0.19	0
Glyma.11G249200	NUCLEOPORIN P54	Glyma.18G008100	Nuclear pore complex, p54 component (sc Nup57)	0.02	0.12	0.16
Glyma.11G249300	F-box domain (F-box) // Kelch motif (Kelch_1)	Glyma.18G007900	F-box domain (F-box) // Kelch motif (Kelch_1)	0.02	0.10	0.2
Glyma.11G249500	MINICHROMOSOME MAINTENANCE (MCM2/3/5) FAMILY PROTEIN	Glyma.18G007700	MINICHROMOSOME MAINTENANCE (MCM2/3/5) FAMILY PROTEIN	0.02	0.17	0.11
Glyma.11G249600	MICROSOMAL SIGNAL PEPTIDASE 12 KDA SUBUNIT	Glyma.18G007600	MICROSOMAL SIGNAL PEPTIDASE 12 KDA SUBUNIT	0.05	0.11	0.45
Glyma.11G249700	60S RIBOSOMAL PROTEIN L17	Glyma.18G007500	60S RIBOSOMAL PROTEIN L17	0.00	0.23	0
Glyma.11G249800	ELMO/CED-12 DOMAIN-CONTAINING PROTEIN-RELATED	Glyma.18G007400	ELMO/CED-12 DOMAIN-CONTAINING PROTEIN-RELATED	0.01	0.11	0.09
Glyma.11G249900	65-KDA MICROTUBULE-ASSOCIATED PROTEIN 8	Glyma.18G007200	65-KDA MICROTUBULE-ASSOCIATED PROTEIN 8	0.02	0.15	0.13

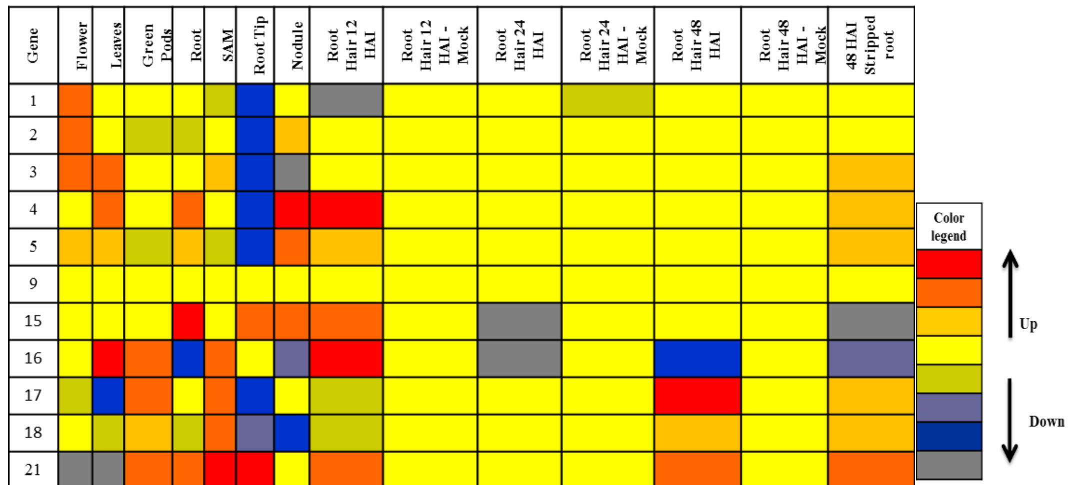
ESM 2. Functional annotation, *Ka* and *Ks* values, and *Ka/Ks* ratios of the

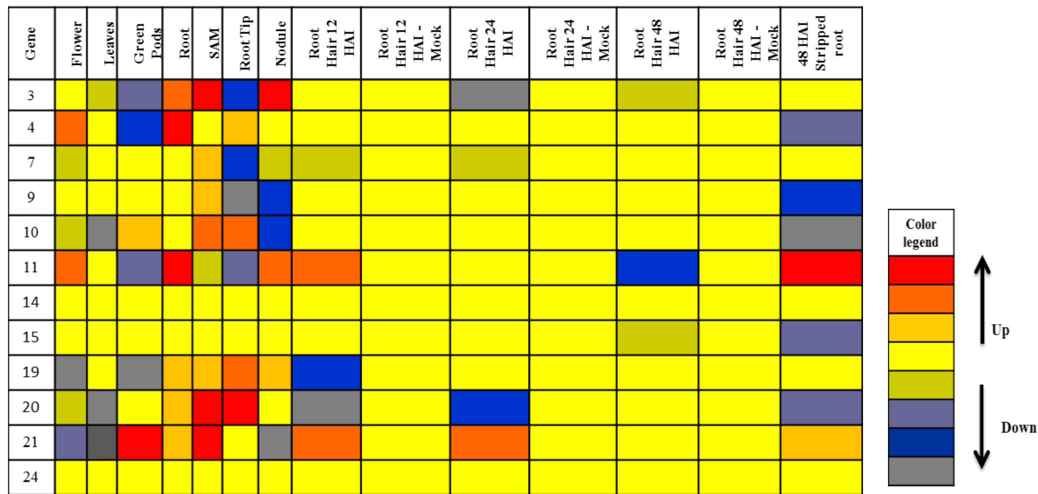
paralogous *GmNF-Y* genes analyzed in the PGDD database.

GmNF-YA



GmNF-YB





Supplementary Figure 1-3: Expression profile of *GmNF-YA, GmNF-YB* and *GmNF-YC* genes in leaves, flowers, green pods, SAM (Shoot Apical Meristem), roots, root tips, nodules, and roots inoculated with *Bradyrhizobium japonicum*. In order to generate more understandable figures, the expression data were adapted from BAR images. Red color indicates up- and blue indicates downregulated genes. A direct comparison of each specific gene in each organs/tissues is applicable. However, comparison among different genes is not adequate.

CAPÍTULO 4

Artigo em revisão no Periódico *Molecular Phylogenetics and Evolution*

Revising the PLAC8 Gene Family:
From a Central Role in Differentiation, Proliferation, and Apoptosis in Mammals to a Multifunctional Role in Plants

Cabreira-Cagliari, C.[1]*; Dias, N. C. F. [2]*; Bohn, B.; Fagundes, D. G. S. [2]; Margis-Pinheiro, M[1]; Bodanese-Zanettini, M.H.[1]; Cagliari, A[2].

[1] Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil[2] Universidade Estadual do Rio Grande do Sul (UERGS), Santa Cruz do Sul, RS, Brazil

* These authors contributed equally.

Corresponding author: Alexandro Cagliari, e-mail: alexandro-cagliari@uergs.edu.br, phone: +55-51-3715-6926, address: Universidade Estadual do Rio Grande do Sul - UERGS, CEP 96816-501, Santa Cruz do Sul, RS, Brazil.

Abstract

PLAC8 is a cysteine-rich protein described as a central mediator of tumor evolution in mammals, being a promising candidate for diagnostic and therapeutic targeting. The PLAC8 gene also acts in the contact hypersensitivity response and serves a role in psoriatic skin. In plants, PLAC8 motif-containing proteins are involved in determination of organ size, growth, response to infection, Ca^{2+} influx, Cd resistance, and zinc detoxification. In general, the PLAC8 motif-containing proteins present the conserved CCXXXXCPC or CLXXXXCPC region. However, there is no devised nomenclature for the PLAC8 motif-containing proteins. Here, through the analysis of 445 sequences, we show that PLAC8 motif-containing proteins compose a unique gene family, and we propose a unified nomenclature. This is the first report indicating the existence of different groups of PLAC8 proteins, which we have called types I, II, and III. The type I genes are found in mammals, fungi, plants, and algae, while the types II and III are exclusive to plants. Our study describes for the first time the PLAC8 type III proteins. Whether

these sequences maintain their known functional role or possess distinct functions of types I and II genes, remains unclear.

Key words: PLAC8 genes; phylogenetic reconstruction; Programmed Cell Death; Cell Differentiation; Cysteine-rich proteins.

Introduction

Programmed cell death (PCD) is a genetically regulated process of cellular suicide that plays a fundamental role in a variety of developmental and physiological processes, removing infected or damaged cells in all multicellular organisms. A significant increase in the knowledge about PCD has occurred in the last years. In the plant model *Arabidopsis thaliana*, several genes involved with PCD have been identified. Recently, a network called the "deathsome", composed by several genes related to PCD control, has been proposed[1]. LSD1 (*Lesion simulating disease 1*), a central gene in this network, negatively regulates PCD under stress conditions[2-4] by physically interacting with positive regulators of Hypersensitive Response (HR), such as bZIP10 (a basic region leucine zipper (bZIP) transcription factor)[5], and type I metacaspase (MC1)[6]. In addition, two genes encoding PLAC8 (Placenta-specific 8) motif proteins, At1g52200 and At4g23470, were also described as components of the *A. thaliana* deathsome[1].

The first PLAC8 protein was identified through a microarray analysis on placental (hence the name PLAC: Placenta-specific) and embryonic RNA in mouse[7]. Mouse PLAC8 is a cysteine-rich protein of 112 amino acids, encoded by a gene located at chromosome 5. PLAC8 expression is not restricted to the uterus and placenta of mammals; it is highly expressed in phagocytes, macrophages, and neutrophils, and is essential to optimal killing of bacteria by these cells[8]. In myeloid cells, PLAC8 expression plays a role in differentiation, proliferation, and apoptosis[9-11], and provides a mechanistic link between primary oncogenic mutations and the induction of autophagy[12]. The overexpression of PLAC8 increased growth rate, resistance to apoptotic stimuli, and loss of the cell cycle checkpoint, and promoted tumorigenic conversion[10]. PLAC8 is also required is a critical upstream regulator of brown fat differentiation, and thermoregulation [13].

In pancreatic ductal adenocarcinoma, PLAC8 is a central mediator of tumor progression, being a promising candidate gene for diagnostic and therapeutic targeting[14]. When overexpressed, PLAC8 can suppress p53 expression[10]. In addition, PLAC8 expression is affected by p53 mutation and Ras activation. Downregulation of PLAC8 by RNA interference reduced tumor formation by 100% in murine colon cancer[15]. Therefore, PLAC8 represents a critical node in gene networks underlying the malignant phenotype[15].

Increased expression of PLAC8 has been reported in early stages of pancreatic cancer[16], while decreased PLAC8 expression is associated with oncogenesis in the liver[17]. PLAC8 acts in apoptosis protection in fibroblasts[10] and apoptosis induction in human lymphocytes[18]; inhibits cell differentiation in primary acute myeloid leukaemia cells[11], and inducts the epithelial-to-mesenchym transition in cultured colon cancer cells[19]. Altogether, these results suggest that the role of PLAC8 seems to be highly cellular and is physiologically context-dependent.

Since the first report in humans more than ten years ago[7], the increased numbers of proteins containing the cysteine-rich PLAC8 domain have provided an ongoing challenge in their clear identification and logical classification across species. Until now, there has been no devised nomenclature for naming these proteins. In this context, comparative functional genomics associated with a high-throughput phylogenetic analysis covering several genomes can contribute to understanding the relationship among the PLAC8 motif-containing proteins.

From this perspective, we first discuss the PLAC8 motif-containing proteins' functions in mammals and plants. Then, we suggest the first unified nomenclature based on the initial characterized PLAC8 gene. Finally, we propose a classification of the PLAC8 family in three different groups (types I, II, and III), two of them restricted to plants.

Material and Methods

PLAC8 genes identification

In order to overcome the confused nomenclature and contribute to the currently incomplete characterization of the PLAC8-containing proteins, we retrieved all of the PLAC8-like sequences using the cysteine-rich domain that defines these proteins (Pfam PF04749) according the SMART database (http:// smart.embl-heidelberg.de/). The baits At4g23470 and At1g52200 were used as queries in BLAST searches conducted against the NCBI

(http://www.ncbi.nlm.nih.gov), Metazome (https://metazome.jgi.doe.gov/), FungiDB (http://fungidb.org/fungidb/), Phytozome (http://www.phytozome.org/), and PLAZA (http://bioinformatics.psb.ugent.be/plaza/versions/plaza_v3_dicots/) databases.

Evolutionary processes

The complete sequences from representative species were aligned using the Muscle algorithm as implemented in MEGA v.5.05 (Tamura et al. 2011). A Bayesian inference was generated using BEAST v. 1.4.7[33] and a run of 10^7 chains was conducted[34] and the trees were sampled every 1000 generations. The Yule tree prior, the JTT substitution model and the uncorrelated log-normal relaxed clock were used in the BEAST analysis. The TRACER v.1.4 (http://beast.bio.ed.ac.uk/Tracer) was used to check the convergence of the Monte Carlo Markov Chains (MCMCs) and for adequate effective sample sizes (EES >200) after the first 10% of the generations had been deleted as burn-in. The final joint sample was used to estimate the maximum clade credibility tree with the TreeAnnotator program, which is part of the BEAST package. The statistical support for the clades was determined by accessing the Bayesian posterior probability (PP). The trees were visualized using FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Gene structure and PLAC8 proteins characterization

The gene structure was analyzed using the Exon-Intron Graphic Maker (http://wormweb.org/exonintron) or manually drawn. The cysteine-rich domain that defines these proteins (Pfam PF04749) was selected according the SMART database (http:// smart.embl-heidelberg.de/). Their amino acid sequences were analyzed MEGA v.5.05 (Tamura et al. 2011) and manually aligned to access domain conservation using MEME software (http://meme.sdsc.edu/meme/).

Results and discussion

Distinct names of the PLAC8 motif-containing proteins in mammals and plants

The search for PLAC8 genes showed that several distinct names have been used to designate the PLAC8 motif-containing proteins. Onzin and Cornifelin (CNFN) acronym have been used to designate the PLAC8 human protein[20, 21]. Fruit weight-2.2 (FW2.2) and Cell number regulator (CNR) have been used to described PLAC8-containing proteins related to organ size and cell number control[20, 21, 22, 23, 24, 25, 26]**.** Cys-rich PCR (Plant Cadmium Resistance) acronym had been used to denominate proteins involved with resistance to chemical elements[27,28]. The Mid1-complementing activity (MCA) name has been employed to nominated proteins related to Ca^{2+} influx at the plasma membrane[29,30,31,32].

Onzin and Cornifelin (CNFN)

The name 'cornifelin' was proposed to specify a gene highly expressed in psoriatic skin, whose overexpression alters the protein composition of the cornified cell envelope (a structure that is formed beneath the plasma membrane)[20]. Regarding the name 'onzin', it was used to nominate a leukocyte inhibitory factor-regulated gene in the uterus, which presented an uncharacterized cysteine-rich domain that did not conform to consensus zinc- or RING-finger domains[21].

Since the first report[21], several studies have investigated the role of onzin. Onzin was described as a new c-Myc target gene highly expressed in 32D cells and markedly repressed by c-Myc[10]. Onzin also interacts with PLSCR1 (Phospholipid Scramblase 1), which is proposed to mediate the bidirectional movement of the plasma membrane during proliferation and apoptosis[9]. Overexpression of onzin in fibroblasts resulted in growth, loss of cell cycle control, resistance to apoptosis, and tumorigenesis[10]. The opposite result was found when the reduction of endogenous onzin levels was obtained by shRNA.

Onzin contributes to intracellular killing within the neutrophil[8]. Leukaemic cell differentiation-inducing agents significantly inhibited the expression of onzin in acute myeloid leukaemic cell lines and primary cells as well as cells of some onzin-expressing solid tumor cells[11]. In addition, onzin also plays a critical role in the contact hypersensitivity response through the regulation of pro-inflammatory factors[21].

Fruit weight-2.2 (FW2.2) and Cell number regulator (CNR)

The *fw2.2* acronym designated a PLAC8 motif-containing protein included in a quantitative trait locus (QTL), the expression of which modulates fruit size[22]. *fw2.2* was demonstrated to act as a negative regulator of cell division during the early stages of fruit development following pollination[23]. FW2.2 interacts with the regulatory subunit (beta) of casein kinase II (CKII)[24]. The CKII protein has been linked to a wide variety of cellular functions, including cell division control. In soybean, expression of *FWL1* (*Glycine max FW2.2-like 1*) is induced after *Bradyrhizobium japonicum* inoculation, suggesting that it is involved in the cellular remodeling programm required for response to rhizobial infection[25]. In addition, *GmFWL1* controls (directly or indirectly) nuclear size and chromatin condensation.

The CNR genes were described as orthologs of *fw2.2* in maize (*Zea mays*)[26]. *CNR1* acts to reduced overall plant size when ectopically overexpressed. The opposite phenotype is observed when *CNR1* is cosuppressed or silenced. The differences in plant and organ size occurred by changes in cell number, rather than cell size. In addition, *CNR2* expression is negatively correlated with tissue growth activity and hybrid seedling vigor[26]. All these results are consistent with the observations of fw2.2 genes.

Cys-rich PCR (Plant Cadmium Resistance)

The PLAC8 motif-containing proteins PCR were described as sequences that confers a strong Cd (a widespread soil pollutant) resistance at the plasma membrane level[27]. These authors showed that AtPCR1 presents two hydrophobic segments, which form helix structures that may function as transmembrane domains. These domains (CCXXXXCPC) are situated in the N-terminal region of the protein and deletion or mutations in CC and CPC residues to AA and AAA strongly reduce the capacity of *AtPcr1* to confer Cd(II) resistance. Interestingly, the complete Cys-rich region in the first putative transmembrane domain of *AtPcr1* is important for Cd(II) resistance at the plasma membrane level. In addition, *AtPCR2* is involved in detoxification in the presence of high concentrations of zinc and in the transfer of zinc from the root to the shoot[28].

The Mid1-complementing activity (MCA)

The PLAC8 motif-containing protein MCA1 corresponds to an integral plasma membrane protein MCA1 (Mid1-complementing activity) that connects Ca^{2+} influx with

mechanical stimulation in *A. thaliana*[29]. MCA1 rescues the ability to take up Ca^{2+} ions in response to mating factor; *mca* mutants are defective in Ca^{2+} uptake[30]. MCA1 has no significant similarity to that of any protein characterized as an ion channel component. Analysis of MCA1 domains showed at least two potential transmembrane segments and the PLAC8 domain at the carboxyl-terminal half. In addition, the MCA2 protein has a distinct role in Ca^{2+} uptake in roots, and an overlapping role with MCA1 in plant growth[31].

In rice, MCA1 is involved in the regulation of plasma membrane Ca^{2+} influx and reactive oxygen species generation induced by hypo-osmotic stress[32]. *OsMCA1*-suppressed plants showed retarded growth and shortened rachises[32]. In addition, a SNP in *OsMCA1* responds for a plant architecture defect by deactivation of bioactive gibberellin[31].

The PLAC8 family: A consolidated nomenclature to designate PLAC8 proteins

Including all above-described names for designate PLAC8 genes, a total of 445 proteins containing the PLAC8 domain were identified (Supplementary Table 1 and Figure 1A). The common characteristic observed in all proteins is the cys-rich PLAC8 domain. In the context of different names to refer PLAC8 proteins, an evident confusion and incomplete characterization of these proteins is presently common. A new nomenclature containing PLAC8 acronym and the number of identified gene (according BLAST output) was set. This unified can clarify the present arbitrary gene annotation of the PLAC8 motif-containing proteins.

Considering the number of genes identified, lineage-specific fluctuations in gene family size were observed among taxa (Figure 1, Supplementary Table 1). In humans, three proteins containing the PLAC8 motif were identified. The first one corresponds to the cornifelin protein[20]. The second one matches to the PLAC8 itself[7], which is also named onzin[21]. Finally, the third one corresponds to a protein sequence that was not previously described, and we called it PLAC8-like 1 (Supplementary Table 1, Figure 2). In *Aspergillus nidulans*, two PLAC8 genes were found (Supplementary Table 1). In comparison with mammals and fungi, Viridiplantae presented a greater or equal number of genes, with soybean being the species with the highest number of members (Figure 1A).

Based on domain structure analysis (SMART database), we can distinguish three groups of PLAC8 domains, allowing the classification of PLAC8 proteins in types I, II, and III (Figure 1B-E). A prominent discrepancy in numbers of copies between groups was observed, since the

representative genes of type I were always larger than those of types II and III (Figure 1A). In addition, single-copy genes were found exclusively in types II and III.

The lengths of the three types of PLAC8 domains vary from 95 to 110 amino acids, all them starting with a tryptophan residue (W) (Figure 1 B-E and Supplementary Table 1). Based on this classification, we notice that the sequences belonging to the type I group correspond to the most well-characterized PLAC8 proteins until now. Their involvement in PCD and differentiation in mammals, in addition to different biological roles already described in plants, highlight their functional importance. Regarding the type II group, representative genes were investigated exclusively in *Arabidopsis*[29] and rice[31-32], and different functions have been proposed in comparison to type I. Concerning the genes classified as type III, they are all functionally uncharacterized and are now described for the first time.

The evolutionary history of the PLAC8 motif

The relationship of the previously named onzin, cornifelin, FWL, PCR, CNR, and MCA genes and the sequences now identified was emphasized by the evolutionary analysis (Supplementary Material 1, tab 1 and 2, and Figure 2). The unrooted phylogenetic tree performed according bayesian inference indicated that these genes compose a unique gene family whose unique domain present in the sequences is the PLAC8. All these results summarize the importance to group together all the families acronym used until now (onzin, cornifelin, FWL, PCR, CNR, and MCA) in unify the name to designate all PLAC8 genes.

During the evolutionary process, the well-conserved cysteine-rich region of the PLAC8 domain (CCXXXXCPC or CLXXXXCPC) was maintained in all PLAC8 proteins. It starts at residue 15 in types I and II, and at residue 16 in type III (Figure 1 B-E, black lines). These motifs are present in the N-terminal region of PLAC8 proteins, allowing anti-apoptotic activity[10] and mediating the interaction of onzin with Akt and Mdm2[10], and with PLSCR1[9] in mammals. In plants, deletion or mutations in the CC and CPC residues reduce the capacity of *AtPcr1* to confer Cd(II) resistance[27]. These results, in addition to the high conservation of these cysteine-rich regions, highlight the importance of their maintenance throughout the evolution, allowing the functionality of PLAC8 proteins.

The existence of three different types of PLAC8 genes was corroborated by a well-supported phylogenetic tree topology, confirming the proposed classification (Figure 2). The spread of the PLAC8 domain through the evolutionary process presents an interesting evolutionary scenario. While the type I PLAC8 proteins (337 genes) were widely distributed (algae, fungi, plants, and mammals), the type II (48 genes) and type III (71 genes) were present only in plants, emerging from lycophyte and bryophyte, respectively (Figure 1). The existence of types II and III groups exclusively in plants entails a question: were these groups not maintained in algae, humans, and fungi through the evolutionary process (as the copies were lost), or did the emergence of these different types of PLAC8 proteins occur exclusively in plants?

Human PLAC1 and PLAC9 are not phylogenetically related to PLAC8 gene

PLAC1 and PLAC9 are common cited as genes related to PLAC8, since they share the same acronym name for genes. PLAC1 is a gene strongly expressed in all trophoblast derived cells in the placenta[35]. Similar to PLAC8, studies indicated that PLAC1 is also expressed by a wide variety of human cancers and cancer-derived PLAC1 has the potential to promote tumor growth and function[36]. Another gene that carries the PLAC acronym is PLAC9, which is weakly expressed though highly enriched in placenta[7].

To clarify if PLAC1 and PLAC9 human genes are members of PLAC8 family, these sequences were investigated and included in the phylogenetic analysis. We observed that the PLAC8 motif was absent in their proteins. Furthermore, the phylogenetic analysis evidenced that PLAC1 and PLAC9 human genes have no relationship with PLAC8 genes (Figure 2). These genes localized as outgroup in the phylogenetic tree, highlighting that they share different evolutionary history. Is important to note the names PLAC2–PLAC7 have been assigned in GenBank to other genes derived from placent[7].

Conclusion and Perspectives

In light of the data discussed here, the PLAC8 family constitutes a group of proteins involved in several important biological processes in plants and mostly apoptosis networks in

mammals. The existence of PLAC8 genes in several clades highlights the importance of their maintenance through the evolutionary process (Figure 3).

In the present study, the evolutionary history of the PLAC8 proteins was reconstructed, providing the first evidence that the PLAC8 motif-containing proteins compose a unique gene family (Figure 3). We propose a unified nomenclature for the PLAC8 family, overcoming the current nomenclature confusion. Our developed unified nomenclature will be helpful to achieve quick prediction considering the future investigation of new PLAC8 genes, since from the nomenclature point of view, the newly cloned gene(s) will always be characterized/named based on sequence similarity with the presently characterized PLAC8 proteins.

Here we provide the first report comprising the existence of different groups of PLAC8 proteins (types I, II, and III) based on domain composition. The proteins of type III group were described for the first time. Whether the type III have the same functional role of type I and II genes, or whether this type is functionally distinct from them, remains unclear.

The authors declare that they have no conflicts of interest.

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Capacitação de Pessoal de Nível Superior (CAPES), Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul (PPGBM/UFRGS), Universidade Estadual do Rio Grande do Sul (Uergs), GENOSOJA/CNPq and BIOTECSUR II/MCT.

References

1. Coll NS, Epple P, Dangl JL. Programmed cell death in the plant immune system. Cell Death Differ [Internet]. 2011;18(8):1247–56. Available from: http://dx.doi.org/10.1038/cdd.2011.37

2. Dietrich RA, Richberg MH, Schmidt R, Dean C, Dangl JL. A novel zinc finger protein is encoded by the arabidopsis LSD1 gene and functions as a negative regulator of plant cell death. Cell [Internet]. 1997;88(5):685–94. Available from: <Go to ISI>://WOS:A1997WM41300014

3. Epple P, Mack AA, Morris VR, Dangl JL. 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 [Internet]. 2003;100(11):6831–6. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12732715

4. Cabreira C, Cagliari A, Bücker-Neto L, Wiebke-Strohm B, De Freitas LB, Marcelino-Guimarães FC, et al. The Lesion Simulating Disease (LSD) gene family as a variable in soybean response to Phakopsora pachyrhizi infection and dehydration. Funct Integr Genomics. 2013;13(3):323–38.

5. Kaminaka H, Näke C, Epple P, Dittgen J, Schütze K, Chaban C, et al. bZIP10-LSD1 antagonism modulates basal defense and cell death in Arabidopsis following infection. EMBO J. 2006;25(18):4400–11.

6. Coll NS, Vercammen D, Smidler A, Clover C, Breusegem F Van, Dangl JL, et al. Arabidopsis type I metacaspases control cell death. Science. 2010;330(6009):1393–7.

7. Galaviz-Hernandez C, Stagg C, De Ridder G, Tanaka TS, Ko MSH, Schlessinger D, et al. Plac8 and Plac9, novel placental-enriched genes identified through microarray analysis. Gene. 2003;309(2):81–9.

8. Ledford JG, Kovarova M, Koller BH. Impaired host defense in mice lacking ONZIN. J Immunol. 2007;178(8):5132–43.

9. Li Y, Rogulski K, Zhou Q, Sims PJ, Prochownik E V. The negative c-Myc target onzin affects proliferation and apoptosis via its obligate interaction with phospholipid scramblase 1. Mol Cell Biol. 2006;26(9):3401–13.

10. Rogulski K, Li Y, Rothermund K, Pu L, Watkins S, Yi F, et al. Onzin, a c-Myc-

repressed target, promotes survival and transformation by modulating the Akt-Mdm2-p53 pathway. Oncogene. 2005;24(51):7524–41.

11. Wu S-F, Huang Y, Hou J-K, Yuan T-T, Zhou C-X, Zhang J, et al. The downregulation of onzin expression by PKCepsilon-ERK2 signaling and its potential role in AML cell differentiation. Leuk Off J Leuk Soc Am Leuk Res Fund, UK. 2010;24(3):544–51.

12. Kinsey C, Balakrishnan V, O'Dell MR, Huang JL, Newman L, Whitney-Miller CL, et al. Plac8 links oncogenic mutations to regulation of autophagy and is critical to pancreatic cancer progression. Cell Rep. 2014;7(4):1143–55.

13. Jimenez-Preitner M, Berney X, Uldry M, Vitali A, Cinti S, Ledford JG, et al. Plac8 is an inducer of C/EBP?? required for brown fat differentiation, thermoregulation, and control of body weight. Cell Metab. 2011;14(5):658–70.

14. Kaistha BP, Lorenz H, Schmidt H, Sipos B, Pawlak M, Gierke B, et al. PLAC8 localizes to the inner plasma membrane of pancreatic cancer cells and regulates cell growth and disease progression through critical cell-cycle regulatory pathways. Cancer Res. 2016;76(1):96–107.

15. McMurray HR, Sampson ER, Compitello G, Kinsey C, Newman L, Smith B, et al. Synergistic response to oncogenic mutations defines gene class critical to cancer phenotype. Nature [Internet]. 2008;453(7198):1112–6. Available from: http://dx.doi.org/10.1038/nature06973\nhttp://www.nature.com/nature/journal/v453/n7198/pdf/nature06973.pdf

16. Buchholz M, Braun M, Heidenblut A, Kestler H a, Klöppel G, Schmiegel W, et al. Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions. Oncogene [Internet]. 2005;24(44):6626–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16103885

17. Grate LR. Many accurate small-discriminatory feature subsets exist in microarray transcript data: biomarker discovery. BMC Bioinformatics [Internet]. 2005;6:97. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15826317

18. Mourtada-Maarabouni M, Watson D, Munir M, Farzaneh F, Williams GT. Apoptosis suppression by candidate oncogene PLAC8 is reversed in other cell types. Curr Cancer Drug Targets [Internet]. 2013;13:80–91. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22920440

19. Li C, Ma H, Wang Y, Cao Z, Graves-Deal R, Powell AE, et al. Excess PLAC8

promotes an unconventional ERK2-dependent EMT in colon cancer. J Clin Invest. 2014;124(5):2172–87.

20. Michibata H, Chiba H, Wakimoto K, Seishima M, Kawasaki S, Okubo K, et al. Identification and characterization of a novel component of the cornified envelope, cornifelin. Biochem Biophys Res Commun. 2004;318(4):803–13.

21. Sherwin JR, Sharkey AM, Smith SK. Identification of LIF regulated genes in the mouse uterus. Direct submission to NCBI. 2000 AF263458.

22. Ledford JG, Kovarova M, Jania L a, Nguyen M, Koller BH. ONZIN deficiency attenuates contact hypersensitivity responses in mice. Immunol Cell Biol. 2012;90(7):733–42.

23. Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, et al. Fw2.2: a Quantitative Trait Locus Key To the Evolution of Tomato Fruit Size. Science (80-) [Internet]. 2000;289(5476):85–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10884229

24. Cong B, Liu J, Tanksley SD. Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. Proc Natl Acad Sci U S A [Internet]. 2002;99(21):13606–11. Available from: http://www.pnas.org/content/99/21/13606.full

25. Cong B, Tanksley SD. FW2.2 and cell cycle control in developing tomato fruit: A possible example of gene co-option in the evolution of a novel organ. Plant Mol Biol. 2006;62(6):867–80.

26. Libault M, Stacey G. Evolution of FW2.2-like (FWL) and PLAC8 genes in eukaryotes. Plant Signal Behav [Internet]. 2010;5(10):1226–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20855956

27. Guo M, Rupe M a, Dieter JA, Zou J, Spielbauer D, Duncan KE, et al. Cell Number Regulator1 affects plant and organ size in maize: implications for crop yield enhancement and heterosis. Plant Cell. 2010;22(4):1057–73.

28. Song W-Y, Martinoia E, Lee J, Kim D, Kim D-Y, Vogt E, et al. A novel family of cys-rich membrane proteins mediates cadmium resistance in Arabidopsis. Plant Physiol [Internet]. 2004;135(2):1027–39. Available from: http://www.plantphysiol.org/cgi/doi/10.1104/pp.103.037739\npapers3://publication/doi/10.1104/pp.103.037739

29. Song W-Y, Choi KS, Kim DY, Geisler M, Park J, Vincenzetti V, et al. Arabidopsis PCR2 is a zinc exporter involved in both zinc extrusion and long-distance zinc transport. Plant Cell [Internet]. 2010;22(7):2237–52. Available from:

http://www.plantcell.org/content/22/7/2237/F7.expansion

30. Nakagawa Y, Katagiri T, Shinozaki K, Qi Z, Tatsumi H, Furuichi T, et al. Arabidopsis plasma membrane protein crucial for Ca2+ influx and touch sensing in roots. Proc Natl Acad Sci U S A [Internet]. 2007;104(9):3639–44. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1802001&tool=pmcentrez&rendertype=abstract

31. Yamanaka T, Nakagawa Y, Mori K, Nakano M, Imamura T, Kataoka H, et al. MCA1 and MCA2 that mediate Ca2+ uptake have distinct and overlapping roles in Arabidopsis. Plant Physiol [Internet]. 2010;152(3):1284–96. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2832256&tool=pmcentrez&rendertype=abstract

31. Liu Z, Cheng Q, Sun Y, Dai H, Song G, Guo Z, et al. A SNP in OsMCA1 responding for a plant architecture defect by deactivation of bioactive GA in rice. Plant Mol Biol. 2015;87(1–2):17–30.

32. Kurusu T, Nishikawa D, Yamazaki Y, Gotoh M, Nakano M, Yamanaka T, et al. Plasma membrane protein OsMCA1 is involved in regulation of hypo-osmotic shock-induced Ca2+influx and modulates generation of reactive oxygen species in cultured rice cells. BMC Plant Biol. 2012; 12:11. Available from: http://www.biomedcentral.com/1471-2229/12/11

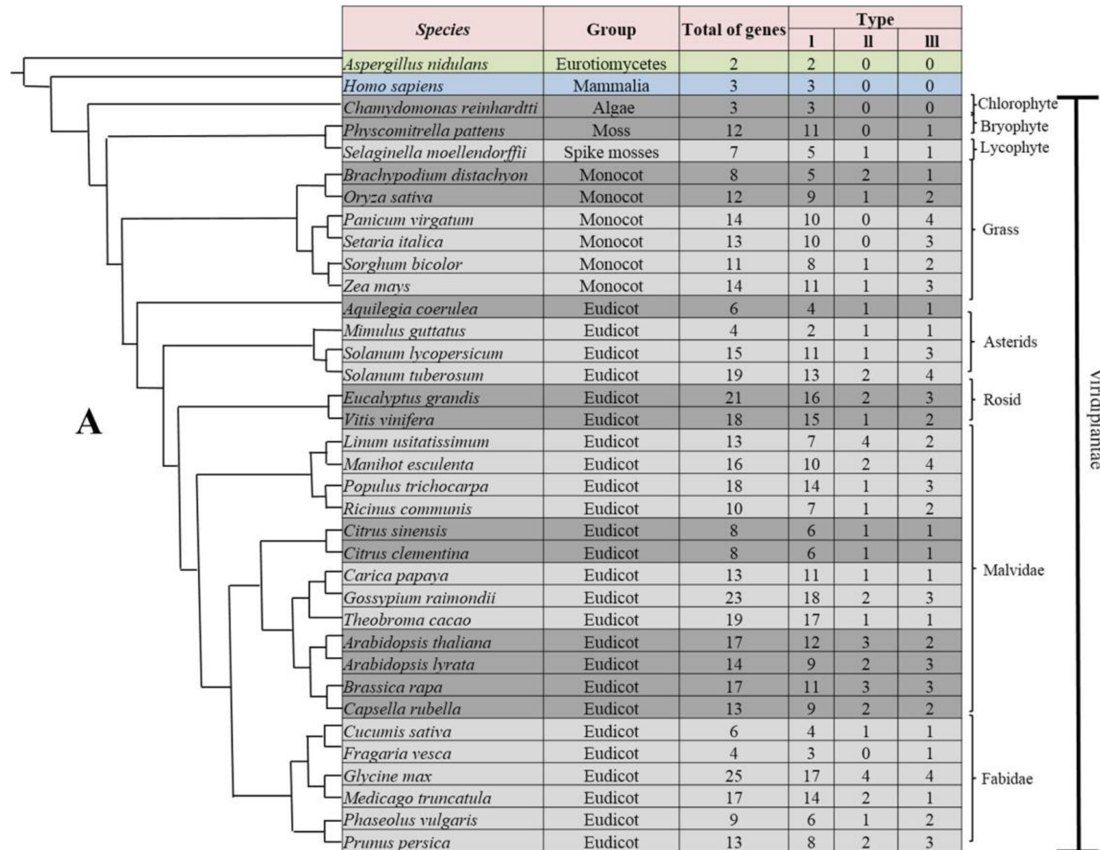
33. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol [Internet]. 2007;7(1):214. Available from: http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-7-214

34. Cabreira C, Cagliari A, Bücker-Neto L, Margis-Pinheiro M, de Freitas LB, Bodanese-Zanettini MH. The phylogeny and evolutionary history of the Lesion Simulating Disease (LSD) gene family in Viridiplantae. Mol Genet Genomics. 2015 Dec;290(6):2107-19.

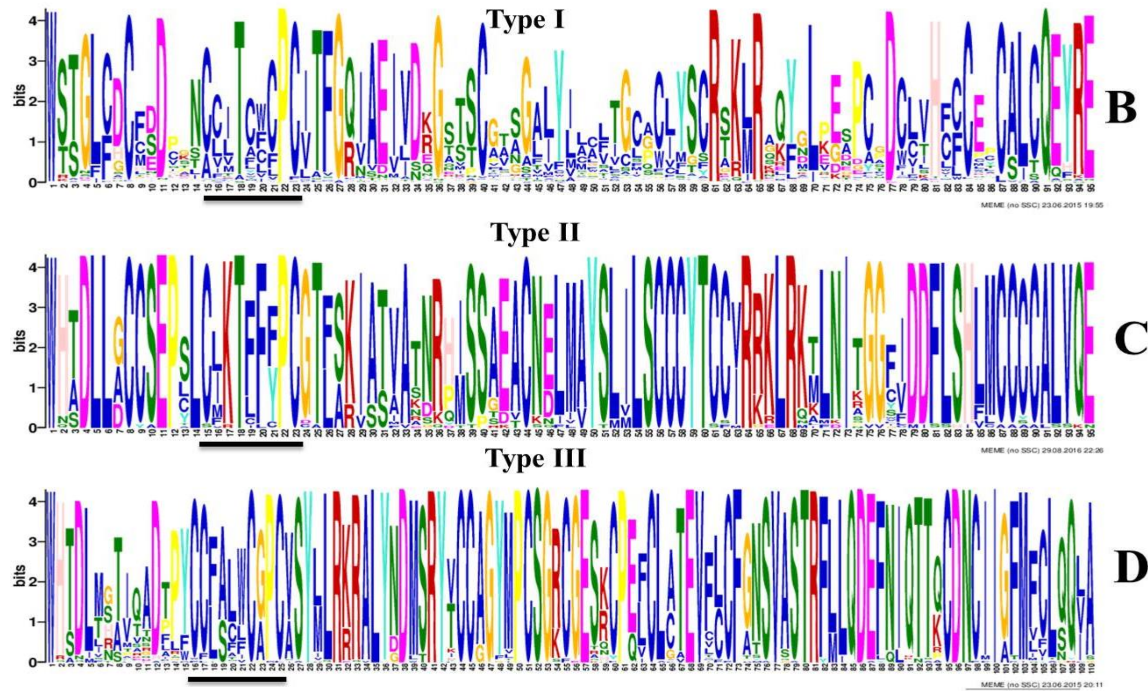
35. Cocchia M, Huber R, Pantano S, Chen E Y, Ma P, Forabosco A, et al. PLAC1, an Xq 26 gene with placenta-specific expression. Genomics (2000) 68:305–12. doi:10.1006/geno.2000.6302

36. Fant M, Farina A, Nagaraja R, Schlessinger D. PLAC1 (Placenta-specific 1): A novel, X-linked gene with roles in reproductive and cancer biology. Prenat Diagn. 2010 Jun;30(6):497-502. doi: 10.1002/pd.2506.

Figure and legends



Species	Group	Total of genes	Type I	Type II	Type III
Aspergillus nidulans	Eurotiomycetes	2	2	0	0
Homo sapiens	Mammalia	3	3	0	0
Chamydomonas reinhardtti	Algae	3	3	0	0
Physcomitrella pattens	Moss	12	11	0	1
Selaginella moellendorffii	Spike mosses	7	5	1	1
Brachypodium distachyon	Monocot	8	5	2	1
Oryza sativa	Monocot	12	9	1	2
Panicum virgatum	Monocot	14	10	0	4
Setaria italica	Monocot	13	10	0	3
Sorghum bicolor	Monocot	11	8	1	2
Zea mays	Monocot	14	11	1	3
Aquilegia coerulea	Eudicot	6	4	1	1
Mimulus guttatus	Eudicot	4	2	1	1
Solanum lycopersicum	Eudicot	15	11	1	3
Solanum tuberosum	Eudicot	19	13	2	4
Eucalyptus grandis	Eudicot	21	16	2	3
Vitis vinifera	Eudicot	18	15	1	2
Linum usitatissimum	Eudicot	13	7	4	2
Manihot esculenta	Eudicot	16	10	2	4
Populus trichocarpa	Eudicot	18	14	1	3
Ricinus communis	Eudicot	10	7	1	2
Citrus sinensis	Eudicot	8	6	1	1
Citrus clementina	Eudicot	8	6	1	1
Carica papaya	Eudicot	13	11	1	1
Gossypium raimondii	Eudicot	23	18	2	3
Theobroma cacao	Eudicot	19	17	1	1
Arabidopsis thaliana	Eudicot	17	12	3	2
Arabidopsis lyrata	Eudicot	14	9	2	3
Brassica rapa	Eudicot	17	11	3	3
Capsella rubella	Eudicot	13	9	2	2
Cucumis sativa	Eudicot	6	4	1	1
Fragaria vesca	Eudicot	4	3	0	1
Glycine max	Eudicot	25	17	4	4
Medicago truncatula	Eudicot	17	14	2	1
Phaseolus vulgaris	Eudicot	9	6	1	2
Prunus persica	Eudicot	13	8	2	3



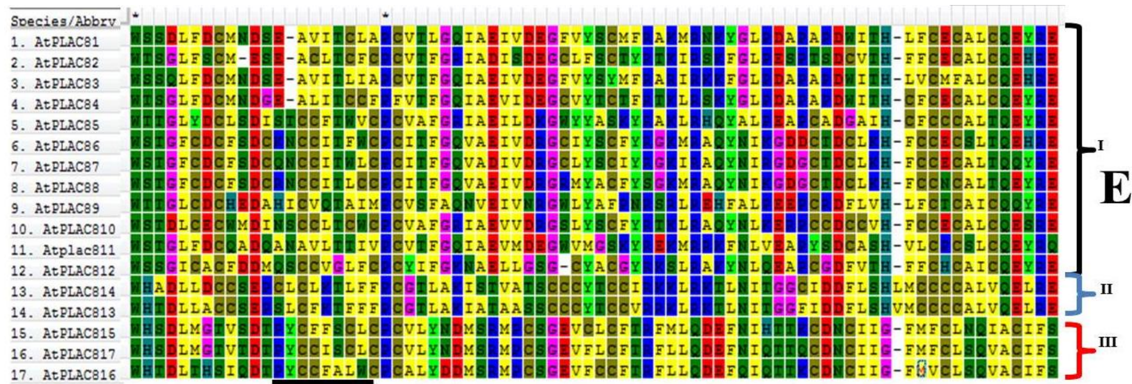


Figure 1. A) The total number of *PLAC8* genes annotated per species and the number of types I, II, and III proteins. Adapted from the Phytozome (http://www.phytozome.org/). B-D) PLAC8 domain analysis. The height of each letter is proportional to the corresponding relative frequency. The amino acids are coloured according to their chemical properties: Blue for most hydrophobic residues (A, C, F, I, L, V, and M); Green for polar, non-charged, non-aliphatic residues (N, Q, S, and T); Red for positively-charged residues (K and R); Orange for glycine (G); Rose for histidine (H); Yellow for proline (P); and Turquoise for tyrosine (Y). The black bar indicates the conserved cysteine-rich region. B) 337 sequences of the type I. C) 48 sequences of the type II. D) 71 sequences of the type III. E) Detailed illustration of PLAC8 domain of types I, II, and III proteins from *Arabidopsis thaliana*. The black bar indicates the conserved cysteine-rich region.

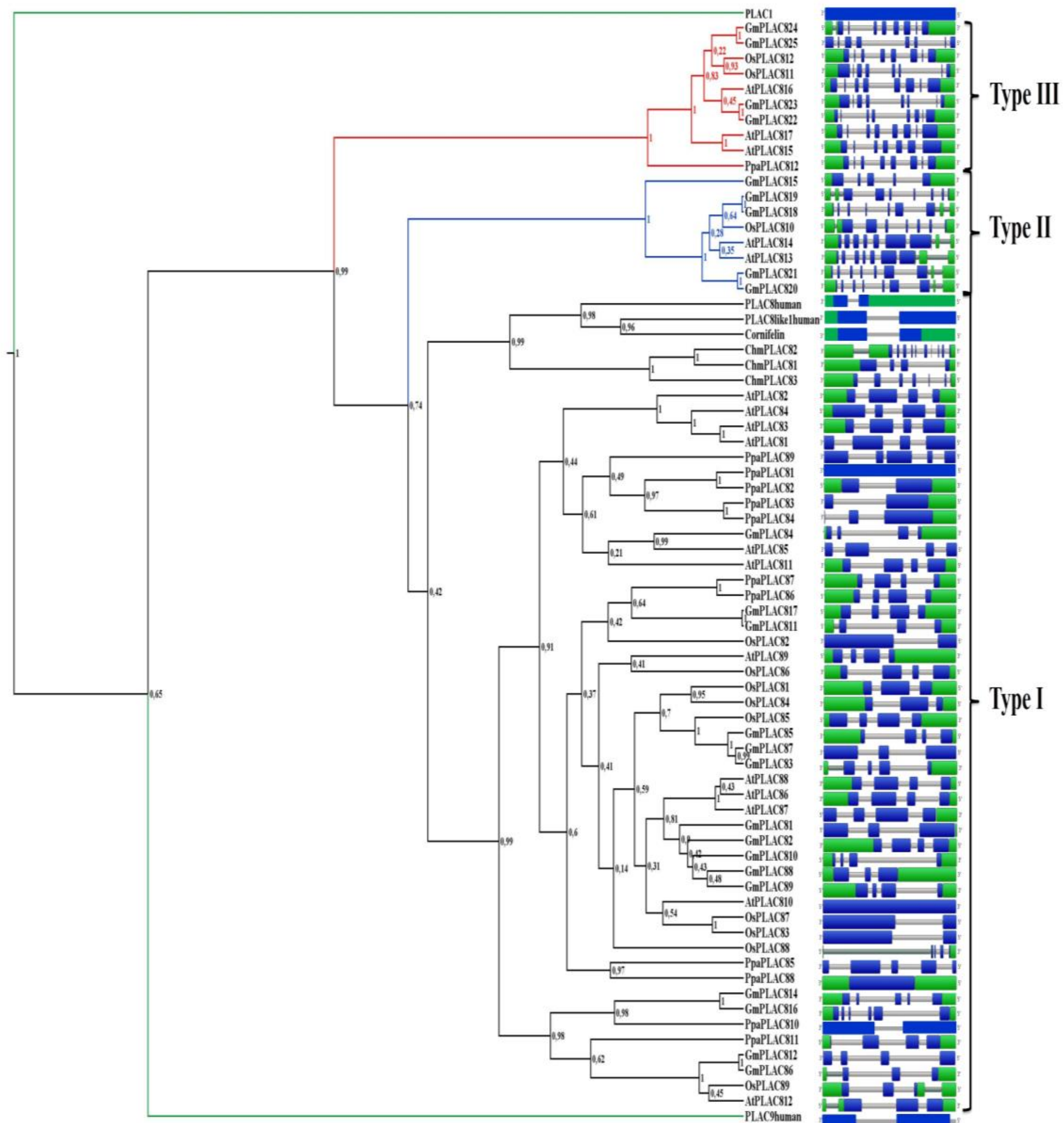


Figure 2. Unrooted tree (Bayesian inference) showing the evolutionary relationships among PLAC8 genes from representative species. The posterior probability is shown above the branches. The taxa terminologies are abbreviated using the first letter of the genus and the first two letters of the species (see Supplementary Material 1). Black shading indicates type I proteins. Blue shading indicates type II proteins. Red shading indicates type III proteins. The right side provides a detailed illustration of the relative intron/exon length and number in each gene. PLAC1 and PLAC9 genes were used as the outgroup. The intron/exon figures were generated using the PLAZA database (Viridiplantae) or manually drawn (human genes). The green boxes represent the 5´and 3´ UTR. The blue boxes show exons, and the lines connecting them represent the introns.

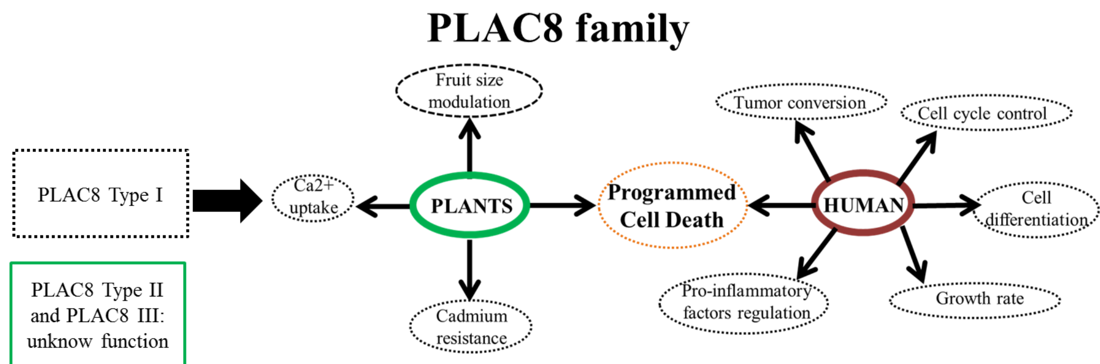


Figure 3. Overview of PLAC8 family function and evolution. Dotted circles represents functions observed to Type I gene, while dotted plus dashed circles shows function found in Type II genes. Blue arrow show genes exclusively found in plants, while green arrow indicates PLAC8 genes observed in plants and humans.

SPECIES	TYPE	ACRONYM	GENE / LOCUS
Manihot esculenta	I	MePLAC81	>cassava4.1_026034m CDS
		MePLAC82	>cassava4.1_030847m CDS
		MePLAC83	>cassava4.1_020997m CDS
		MePLAC84	>cassava4.1_016859m CDS
		MePLAC85	>cassava4.1_023829m CDS
		MePLAC86	>cassava4.1_018107m CDS
		MePLAC87	>cassava4.1_026618m CDS
		MePLAC88	>cassava4.1_024830m CDS
		MePLAC89	>cassava4.1_009432m CDS
		MePLAC810	>cassava4.1_018068m
	II	MePLAC812	ME06512G01420
		MePLAC812	ME07520G02900
	III	MePLAC813	>cassava4.1_014438m CDS
		MePLAC814	>cassava4.1_013090m CDS
		MePLAC815	>cassava4.1_014516m CDS
		MePLAC816	>cassava4.1_016062m CDS
Ricinus communis	I	RcPLAC81	>30174.m008709 CDS
		RcPLAC82	>30115.m001188 CDS
		RcPLAC83	>29692.m000524 CDS
		RcPLAC84	>29801.m003194 CDS
		RcPLAC85	>29692.m000523 CDS
		RcPLAC86	>28883.m000737 CDS

		RcPLAC87	>29806.m000933
	II	RcPLAC88	29729.m002275
	III	RcPLAC89	>30190.m011251 CDS
		RcPLAC810	>30174.m008635 CDS
Linum usitatissimum	I	LuPLAC81	>Lus10021696 CDS
		LuPLAC82	>Lus10034298 CDS
		LuPLAC83	>Lus10019478 CDS
		LuPLAC84	>Lus10043325 CDS
		LuPLAC85	>Lus10004063 CDS
		LuPLAC86	>Lus10008890 CDS
		LuPLAC87	>Lus10009036 CDS
	II	LuPLAC88	>Lus10014256
		LuPLAC89	>Lus10025955
		LuPLAC810	>Lus10041875
		LuPLAC811	>Lus10028421
	III	LuPLAC812	>Lus10017509 CDS
		LuPLAC813	>Lus10018826 CDS
Medicago truncatula	I	MtPLAC81	>Medtr8g104870.1 CDS
		MtPLAC82	>Medtr4g086330.1 CDS
		MtPLAC83	>Medtr4g086320.1 CDS
		MtPLAC84	>Medtr8g104890.1 CDS
		MtPLAC85	>Medtr1g075550.1
		MtPLAC86	>Medtr2g101660.1
		MtPLAC87	>Medtr5g008240.4
		MtPLAC88	>Medtr6g084940.1
		MtPLAC89	>Medtr8g104830.1
		MtPLAC810	Medtr7g044840
		MtPLAC811	Medtr8g075150
		MtPLAC812	Medtr5g022670
		MtPLAC813	Medtr5g017650
		MtPLAC814	Medtr7g100870
	II	MtPLAC815	MT8G075150
		MtPLAC816	MT5G022670
	III	MtPLAC817	Medtr8g098735.1 CDS
Phaseolus vulgaris	I	PvPLAC81	>Phvul.005G169200.1 CDS
		PvPLAC82	>Phvul.002G319200.2 CDS
		PvPLAC83	>Phvul.002G319100.1 CDS
		PvPLAC84	>Phvul.002G318900.1 CDS
		PvPLAC85	>Phvul.002G319000.1 CDS
		PvPLAC86	>Phvul.007G124700.1 CDS
	II	PvPLAC87	>Phvul.002G106200.3

	III	PvPLAC88	>Phvul.002G158600.1 CDS
		PvPLAC89	>Phvul.002G158700.1 CDS
Glycine max	I	GmPLAC81	>Glyma05g34883.1 CDS
		GmPLAC82	>Glyma08g04830.1 CDS
		GmPLAC83	>Glyma07g33441.1 CDS
		GmPLAC84	>Glyma01g20980.1 CDS
		GmPLAC85	>Glyma02g15020.1 CDS
		GmPLAC86	>Glyma01g44060.1
		GmPLAC87	>Glyma02g29850.2
		GmPLAC88	>Glyma05g34870.2
		GmPLAC89	>Glyma05g34876.1
		GmPLAC810	>Glyma07g30410.1
		GmPLAC811	>Glyma11g01530.1
		GmPLAC812	>Glyma13g43327.1
		GmPLAC813	>Glyma15g01990.3
		GmPLAC814	Glyma13g32140.1
		GmPLAC815	Glyma07g30800.1
		GmPLAC816	Glyma02g43090.1
		GmPLAC817	Glyma09g31910.1
	II	GmPLAC818	Glyma01g36860.1
		GmPLAC819	Glyma11g08430.1
		GmPLAC820	Glyma02g05130.1
		GmPLAC821	Glyma16g23240.1
	III	GmPLAC822	>Glyma08g01990.1 CDS
		GmPLAC823	>Glyma05g37590.1 CDS
		GmPLAC824	>Glyma01g43010.1 CDS
		GmPLAC825	>Glyma11g02480.2 CDS
Prunus persica	I	PpPLAC81	>ppa021734m CDS
		PpPLAC82	>ppa011430m CDS
		PpPLAC83	>ppa019128m CDS
		PpPLAC84	>ppa023567m CDS
		PpPLAC85	>ppa018965m CDS
		PpPLAC86	>ppa017696m CDS
		PpPLAC87	>ppa017482m CDS
		PpPLAC88	>ppa020240m CDS
	II	PpPLAC89	Prupe.1G307600.1
		PpPLAC810	Prupe.5G219500.1
	III	PpPLAC811	>ppa010454m CDS
		PpPLAC812	>ppa015403m CDS
		PpPLAC813	>ppa022606m CDS
Fragaria vesca	I	FvPLAC81	>mrna26238.1-v1.0-hybrid CDS

		FvPLAC82	>mrna00157.1-v1.0-hybrid CDS
		FvPLAC83	>mrna28586.1-v1.0-hybrid CDS
	III	FvPLAC84	mrna31806.1-v1.0-hybrid CDS
Arabidopsis thaliana	I	AtPLAC81	>AT3G18460.1 CDS
		AtPLAC82	>AT3G18470.1 CDS
		AtPLAC83	>AT3G18450.1 CDS
		AtPLAC84	>AT1G49030.1 CDS
		AtPLAC85	>AT1G58320.1 CDS
		AtPLAC86	>AT1G14870.1 CDS
		AtPLAC87	>AT5G35525.1 CDS
		AtPLAC88	>AT1G14880.1 CDS
		AtPLAC89	>AT1G68630.1 CDS
		AtPLAC810	>AT1G68610.1 CDS
		AtPLAC811	>AT1G52200.1
		AtPLAC812	AT2G40935.1
	II	AtPLAC813	AT4G35920
		AtPLAC814	AT2G17780
	III	AtPLAC815	>AT5G41390.1 CDS
		AtPLAC816	>AT1G63830.2 CDS
		AtPLAC817	>AT4G23470.1
Arabidopsis lyrata	I	AlPLAC81	>923732 CDS
		AlPLAC82	>479349 CDS
		AlPLAC83	>479347 CDS
		AlPLAC84	>337030 CDS
		AlPLAC85	>911287 CDS
		AlPLAC86	>493614 CDS
		AlPLAC87	>475448 CDS
		AlPLAC88	>926146 CDS
		AlPLAC89	>483158 CDS
	II	AlPLAC810	AL7G04830
		AlPLAC811	AL3G38580
		AlPLAC812	492525 CDS
	III	AlPLAC813	924569 CDS
		AlPLAC814	330382 CDS
Capsella rubella	I	CrPLAC81	>Carubv10010385m CDS
		CrPLAC82	>Carubv10011498m CDS
		CrPLAC83	>Carubv10018179m CDS
		CrPLAC84	>Carubv10010549m CDS
		CrPLAC85	>Carubv10016079m CDS
		CrPLAC86	>Carubv10021324m CDS
		CrPLAC87	>Carubv10022063m CDS

		CrPLAC88	>Carubv10021976m CDS
		CrPLAC89	>Carubv10024118m CDS
	II	CrPLAC810	CRU_007G04020
		CrPLAC811	CRU_003G30700
	III	CrPLAC812	>Carubv10005554m CDS
		CrPLAC813	>Carubv10020877m CDS
Brassica rapa	I	BrPLAC81	>Bra018953 CDS
		BrPLAC82	>Bra022322 CDS
		BrPLAC83	>Bra018774 CDS
		BrPLAC84	>Bra022323 CDS
		BrPLAC85	>Bra026796 CDS
		BrPLAC86	>Bra026181 CDS
		BrPLAC87	>Bra026794 CDS
		BrPLAC88	>Bra004328 CDS
		BrPLAC89	>Bra004325 CDS
		BrPLAC810	>Bra016960
		BrPLAC811	>Bra027875
	II	BrPLAC812	BR03G55000
		BrPLAC813	BR01G02530
		BrPLAC814	BR09G09150
	III	BrPLAC815	>Bra019301 CDS
		BrPLAC816	>Bra013693 CDS
		BrPLAC817	>Bra027663 CDS
Carica papaya	I	CpPLAC81	>evm.model.supercontig_296.1 CDS
		CpPLAC82	>evm.model.supercontig_13.19 CDS
		CpPLAC83	>evm.model.supercontig_62.18 CDS
		CpPLAC84	>evm.model.supercontig_2.225 CDS
		CpPLAC85	>evm.model.supercontig_2.228 CDS
		CpPLAC86	>evm.model.supercontig_2.226 CDS
		CpPLAC87	>evm.TU.supercontig_1.195
		CpPLAC88	>evm.TU.supercontig_2.212
		CpPLAC89	>evm.TU.supercontig_2.213
		CpPLAC810	>evm.TU.supercontig_2.215
		CpPLAC811	>evm.TU.supercontig_92.67
	II	CpPLAC812	CP00006G02880
	III	CpPLAC813	>evm.model.supercontig_180.15 CDS
Gossypium raimondii	I	GrPLAC81	>Gorai.002G174700.1 CDS
		GrPLAC82	>Gorai.002G174600.1 CDS

		GrPLAC83	>Gorai.005G159200.1 CDS
		GrPLAC84	>Gorai.002G174500.1 CDS
		GrPLAC85	>Gorai.006G252400.1 CDS
		GrPLAC86	>Gorai.012G150200.1 CDS
		GrPLAC87	>Gorai.002G175600.1 CDS
		GrPLAC88	>Gorai.013G043200.1 CDS
		GrPLAC89	>Gorai.006G062300.1 CDS
		GrPLAC810	>Gorai.002G175100.1 CDS
		GrPLAC811	>Gorai.002G175300.1 CDS
		GrPLAC812	>Gorai.002G175400.1
		GrPLAC813	>Gorai.003G023000.1
		GrPLAC814	>Gorai.006G195300.1
		GrPLAC815	>Gorai.007G006900.1
		GrPLAC816	>Gorai.008G184700.1
		GrPLAC817	>Gorai.008G184800.1
		GrPLAC818	>Gorai.012G012300.1
	II	GrPLAC819	GR03G00740
		GrPLAC820	GR07G19200
	III	GrPLAC821	>Gorai.008G248500.1 CDS
		GrPLAC821	>Gorai.004G139700.1 CDS
		GrPLAC822	>Gorai.003G156800.1 CDS
Theobroma cacao	I	TcPLAC81	>Thecc1EG012297t1 CDS
		TcPLAC82	>Thecc1EG012293t1 CDS
		TcPLAC83	>Thecc1EG012292t1 CDS
		TcPLAC84	>Thecc1EG012291t1 CDS
		TcPLAC85	>Thecc1EG011842t1 CDS
		TcPLAC86	>Thecc1EG011011t1 CDS
		TcPLAC87	>Thecc1EG000600t1 CDS
		TcPLAC88	>Thecc1EG010997t1 CDS
		TcPLAC89	>Thecc1EG011009t1 CDS
		TcPLAC810	>Thecc1EG020938t1 CDS
		TcPLAC811	>Thecc1EG007048t1 CDS
		TcPLAC812	>Thecc1EG011843t1 CDS
		TcPLAC813	>Thecc1EG010995t1 CDS
		TcPLAC814	>Thecc1EG011002t1 CDS
		TcPLAC815	>Thecc1EG011102t1 CDS
		TcPLAC816	>Thecc1EG012285t1 CDS
		TcPLAC817	>Thecc1EG011840t1 CDS
	II	TcPLAC819	TC0001G01310
	III	TcPLAC818	>Thecc1EG016303t1 CDS
Citrus sinensis	I	CsiPLAC81	>orange1.1g041663m CDS

		CsiPLAC82	>orange1.1g031629m CDS
		CsiPLAC83	>orange1.1g031916m CDS
		CsiPLAC84	>orange1.1g042250m CDS
		CsiPLAC85	>orange1.1g030039m CDS
		CsiPLAC86	>orange1.1g029667m CDS
	II	CsiPLAC88	CS00017G01580
	III	CsiPLAC87	>orange1.1g025094m CDS
Critus clementina	I	CcPLAC81	>Ciclev10017041m CDS
		CcPLAC82	>Ciclev10017071m CDS
		CcPLAC83	>Ciclev10016920m CDS
		CcPLAC84	>Ciclev10016918m CDS
		CcPLAC85	>Ciclev10033938m CDS
		CcPLAC86	>Ciclev10016864m CDS
	II	CcPLAC88	>Ciclev10025703m
	III	CcPLAC87	>Ciclev10021784m CDS
Eucalyptus grandis	I	EgPLAC81	>Eucgr.B00710.1 CDS
		EgPLAC82	>Eucgr.H02012.1 CDS
		EgPLAC83	>Eucgr.G02148.1 CDS
		EgPLAC84	>Eucgr.H01656.1 CDS
		EgPLAC85	>Eucgr.G02143.1 CDS
		EgPLAC86	>Eucgr.G02147.1 CDS
		EgPLAC87	>Eucgr.G02154.1 CDS
		EgPLAC88	>Eucgr.G02142.1 CDS
		EgPLAC89	>Eucgr.I01157.1 CDS
		EgPLAC810	>Eucgr.G02119.1 CDS
		EgPLAC811	>Eucgr.B00709.1 CDS
		EgPLAC812	>Eucgr.G02100.1 CDS
		EgPLAC813	>Eucgr.B00708.1 CDS
		EgPLAC814	>Eucgr.K02209.1 CDS
		EgPLAC815	>Eucgr.B02945.1 CDS
		EgPLAC816	>Eucgr.G02155
	II	EgPLAC827	EG0010G21440
		EgPLAC821	EG0009G19540
	III	EgPLAC817	>Eucgr.D01828.1 CDS
		EgPLAC818	>Eucgr.D02309.1 CDS
		EgPLAC819	>Eucgr.E00640.1 CDS
Vitis vinifera	I	VvPLAC81	>GSVIVT01017057001 CDS
		VvPLAC82	>GSVIVT01008259001 CDS
		VvPLAC83	>GSVIVT01031066001 CDS
		VvPLAC84	>GSVIVT01011690001 CDS
		VvPLAC85	>GSVIVT01011692001 CDS

		VvPLAC86	>GSVIVT01031754001 CDS
		VvPLAC87	>GSVIVT01031755001 CDS
		VvPLAC88	>GSVIVT01011687001 CDS
		VvPLAC89	>GSVIVT01031753001 CDS
		VvPLAC810	>GSVIVT01011693001 CDS
		VvPLAC811	>GSVIVT01011689001 CDS
		VvPLAC812	>GSVIVT01011694001 CDS
		VvPLAC813	>GSVIVT01011928001 CDS
		VvPLAC814	>GSVIVG01011926001
		VvPLAC815	>GSVIVG01025863001
	II	VvPLAC816	VV00G16540
	III	VvPLAC817	>GSVIVT01026997001 CDS
		VvPLAC818	>GSVIVT01019394001 CDS
Solanum tuberosum	I	StPLAC81	>PGSC0003DMT400042614 CDS
		StPLAC82	>PGSC0003DMT400042615 CDS
		StPLAC83	>PGSC0003DMT400014353 CDS
		StPLAC84	>PGSC0003DMT400014354 CDS
		StPLAC85	>PGSC0003DMT400081919 CDS
		StPLAC86	>PGSC0003DMT400081921 CDS
		StPLAC87	>PGSC0003DMT400075916 CDS
		StPLAC88	>PGSC0003DMT400014899 CDS
		StPLAC89	>PGSC0003DMT400014898 CDS
		StPLAC810	>PGSC0003DMT400033759 CDS
		StPLAC811	>PGSC0003DMT400014897 CDS
		StPLAC812	>PGSC0003DMT400014896 CDS
		StPLAC813	>PGSC0003DMT400056038 CDS
	II	StPLAC814	ST03G009030
		StPLAC815	ST02G026830
	III	StPLAC816	>PGSC0003DMT400012782 CDS
		StPLAC817	>PGSC0003DMT400012781 CDS
		StPLAC818	>PGSC0003DMT400049953 CDS
		StPLAC819	>PGSC0003DMT400000176 CDS
Solanum lycopersicum	I	SlPLAC81	>Solyc06g066590.2.1 CDS

		SlPLAC82	>Solyc03g120600.2.1 CDS
		SlPLAC83	>Solyc03g119660.1.1 CDS
		SlPLAC84	>Solyc01g005470.2.1 CDS
		SlPLAC85	>Solyc04g007900.2.1 CDS
		SlPLAC86	>Solyc02g090730.2.1 CDS
		SlPLAC87	>Solyc08g013920.2.1 CDS
		SlPLAC88	>Solyc08g013910.2.1 CDS
		SlPLAC89	>Solyc12g037950.1.1 CDS
		SlPLAC810	>Solyc12g013570.1.1 CDS
		SlPLAC811	>Solyc10g081410.1.1
	II	SlPLAC812	SL02G083540
	III	SlPLAC813	>Solyc10g012080.2.1 CDS
		SlPLAC814	>Solyc01g094870.1.1 CDS
		SlPLAC815	>Solyc08g081360.2.1 CDS
Aquilegia caerulea	I	AcPLAC81	>Aquca_026_00254.1 CDS
		AcPLAC82	>Aquca_015_00087.1 CDS
		AcPLAC83	>Aquca_017_00774.1 CDS
		AcPLAC84	>Aquca_002_01130.1 CDS
	II	AcPLAC85	>Aquca_075_00029.1
	III	AcPLAC86	>Aquca_013_00246.1 CDS
Sorghum bicolor	I	SbPLAC81	>Sb01g002370.1 CDS
		SbPLAC82	>Sb06g019050.1 CDS
		SbPLAC83	>Sb01g002360.1 CDS
		SbPLAC84	>Sb04g003560.1 CDS
		SbPLAC85	>Sb04g034060.1 CDS
		SbPLAC86	>Sb01g002350.1 CDS
		SbPLAC87	>Sb04g024020.1 CDS
		SbPLAC88	>Sb01g002340.1 CDS
	II	SbPLAC89	Sobic.001G499200.1
	III	SbPLAC810	>Sb01g048960.1 CDS
		SbPLAC811	>Sb01g030600.1 CDS
Zea mays	I	ZmPLAC81	>GRMZM2G015941_T01 CDS
		ZmPLAC82	>GRMZM2G053387_T01 CDS
		ZmPLAC83	>GRMZM2G023081_T01 CDS
		ZmPLAC84	>GRMZM2G119755_T01 CDS
		ZmPLAC85	>GRMZM2G151230_T01 CDS
		ZmPLAC86	>>GRMZM2G367431_P01
		ZmPLAC87	>GRMZM5G892035_P01
		ZmPLAC88	GRMZM2G325477_T01
		ZmPLAC89	GRMZM2G168257_T02
		ZmPLAC810	GRMZM2G334628_T05

		ZmPLAC811	GRMZM2G325477_T02
	II	ZmPLAC812	ZM01G04050
	III	ZmPLAC813	>GRMZM2G060564_T01 CDS
		ZmPLAC814	>GRMZM2G065696_T01 CDS
Setaria italica	I	SrPLAC81	>Si019325m CDS
		StrPLAC82	>Si008410m CDS
		SrPLAC83	>Si018556m CDS
		SrPLAC84	>Si038888m CDS
		SrPLAC85	>Si038822m CDS
		SrPLAC86	>Si008245m CDS
		SrPLAC87	>Si039847m CDS
		SrPLAC88	>Si039668m CDS
		SrPLAC89	Seita.9G296400.1
		SrPLAC810	Seita.9G296600.2
	III	SrPLAC811	>Si036850m CDS
		SrPLAC812	>Si037708m CDS
		SrPLAC813	>Si037597m CDS
Panicum virgatum	I	PviPLAC81	>Pavirv00052537m CDS
		PviPLAC82	>Pavirv00008509m CDS
		PviPLAC83	>Pavirv00066814m CDS
		PviPLAC84	>Pavirv00022453m CDS
		PviPLAC85	>Pavirv00022476m CDS
		PviPLAC86	>Pavirv00061728m CDS
		PviPLAC87	>Pavirv00010814m CDS
		PviPLAC88	>Pavirv00049055m CDS
		PviPLAC89	>Pavirv00051291m CDS
		PviPLAC810	>Pavirv00050063m CDS
	III	PviPLAC811	>Pavirv00001210m CDS
		PviPLAC812	>Pavirv00001211m CDS
		PviPLAC813	>Pavirv00027236m CDS
		PviPLAC814	>Pavirv00027879m CDS
Oryza sativa	I	OsPLAC81	>LOC_Os02g36940.1 CDS
		OsPLAC82	>LOC_Os03g61470.1 CDS
		OsPLAC83	>LOC_Os03g61440.1 CDS
		OsPLAC84	>LOC_Os02g36950.1 CDS
		OsPLAC85	>LOC_Os02g52550.1 CDS
		OsPLAC86	>LOC_OS10G02300.1
		OsPLAC87	>LOC_OS03G61500.1
		OsPLAC88	>LOC_OS02G18540.1
		OsPLAC89	LOC_Os01g16210.1
	II	OsPLAC810	LOC_Os03g06120.1

	III	OsPLAC811	>LOC_Os03g03180.1 CDS
		OsPLAC812	>LOC_Os10g39100.1 CDS
Brachypodium distachyon	I	BdPLAC81	>Bradi5g12460.1 CDS
		BdPLAC82	>Bradi3g46930.1 CDS
		BdPLAC83	>Bradi3g58080.1 CDS
		BdPLAC84	>Bradi1g30630.1 CDS
		BdPLAC85	>Bradi1g03090.1 CDS
	II	BdPLAC86	>Bradi1g43866.1
		BdPLAC87	>Bradi2g09980.1
	III	BdPLAC88	>Bradi3g32060.1 CDS
Selaginella moellendorffii	I	SmPLAC81	>122730 CDS
		SmPLAC82	>69283 CDS
		SmPLAC83	>38123 CDS
		SmPLAC84	>69284 CDS
		SmPLAC85	439474
	II	SmPLAC87	>934284
	III	SmPLAC86	>164732 CDS
Physcomitrella patens		PpaPLAC81	>Pp1s189_77V6.1 CDS
		PpaPLAC82	>Pp1s24_101V6.1 CDS
		PpaPLAC83	>Pp1s28_13V6.1 CDS
		PpaPLAC84	>Pp1s68_109V6.1 CDS
		PpaPLAC85	>Pp1s1_243V6.1
		PpaPLAC86	>Pp1s35_289V6.1
		PpaPLAC87	>**Pp1s188_69V6.1**
		PpaPLAC88	>**Pp1s322_40V6.1**
		PpaPLAC89	>**Pp1s372_61V6.1**
		PpaPLAC810	Pp3c21_3760V3.2
		PpaPLAC811	Pp3c17_4810V3.2
	III	PpaPLAC812	>Pp1s28_120V6.1 CDS
Populus trichocarpa	I	PtPLAC81	Potri.008G132900.1 CDS
		PtPLAC82	Potri.010G108800.1 CDS
		PtPLAC83	>Potri.010G127700.1 CDS
		PtPLAC84	Potri.012G060900.1 CDS
		PtPLAC85	Potri.012G092200.1 CDS
		PtPLAC86	Potri.010G109100.1 CDS
		PtPLAC87	Potri.008G132800.1 CDS
		PtPLAC88	Potri.010G109000.1 CDS
		PtPLAC89	Potri.007G042800.1 CDS
		PtPLAC810	Potri.010G127800.1 CDS
		PtPLAC811	Potri.010G108900.1 CDS
		PtPLAC812	Potri.010G109100.3 CDS

		PtPLAC813	Potri.006G024900.2 CDS
		PtPLAC814	Potri.015G088800.1
	II	PtPLAC815	PT05G11000
	III	PtPLAC816	>Potri.003G130700.1 CDS
		PtPLAC817	>Potri.001G101100.1 CDS
		PtPLAC818	>Potri.014G087600.1 CDS
Cucumis sativus	I	CsPLAC81	Cucsa.074420.1 CDS
		CsPLAC82	Cucsa.009110.1 CDS
		CsPLAC83	Cucsa.337240.1 CDS
		CsPLAC84	Cucsa.045730.1 CDS
	II	CsPLAC85	>Cucsa.322780.6
	III	CsPLAC86	>Cucsa.106690.1 CDS
Mimulus guttatus	I	MgPLAC81	Migut.I00935.1 CDS
		MgPLAC82	Migut.I00936.1 CDS
	II	MgPLAC83	>Migut.H02182.1
	III	MgPLAC84	>mgv1a012408m CDS
Homo sapiens	I	Cornifelin	>ENST00000222032 CDS
		PLAC8 human (ONZIN)	ENST00000311507
		PLAC8-like 1 human	ENST00000311450
Aspergillus nidulans	I	AnPLAC81	>CADANIAT00000240
		AnPLAC82	>CADANIAT00001906
Chlamydomonas reinhardtii	I	ChrPLAC81	Cre17.g738050
		ChrPLAC82	Cre17.g738000
		ChrPLAC83	Cre02.g145900

Species	Acronym PLAC8	Acronym at FWL	Acronym at PCR	Acronym at CNR	Acronym at MCA
Glycine max	GmPLAC82	GmFWL3			
	GmPLAC83	GmFWL2			
	GmPLAC84	GmFWL7			
Arabidopsis thaliana	AtPLAC81	AtFWL9	PCR4		
	AtPLAC82	AtFWL6	PCR7		
	AtPLAC83	AtFWL4	PCR5		
	AtPLAC84	AtFWL8	PCR6		
	AtPLAC85	AtFWL1	PCR9		
	AtPLAC86	AtFWL3	PCR2		
	AtPLAC87	AtFWL2	PCR3		
	AtPLAC88	AtFWL7	PCR1		

	AtPLAC810	AtFWL5			
	AtPLAC813				MCA1
	AtPLAC814				MCA2
Zea mays	ZmPLAC81			ZmCNR1	
	ZmPLAC82			ZmCNR3	
	ZmPLAC83			ZmCNR9	
	ZmPLAC84			ZmCNR7	
	ZmPLAC85			ZmCNR2	
	ZmPLAC87			ZmCNR05	
	ZmPLAC88			ZmCNR04	
	ZmPLAC89			ZmCNR06	
	ZmPLAC810			ZmCNR08	
	ZmPLAC811			ZmCNR10	
Oryza sativa	OsPLAC81	OsFWL1		OsCNR02	
	OsPLAC82	OsFWL8		OsCNR09	
	OsPLAC83	OsFWL5		OsCNR10	
	OsPLAC84	OsFWL2		OsCNR04	
	OsPLAC85	OsFWL3			
	OsPLAC86		OsPcr1	OsCNR11_12	
	OsPLAC87			OsCNR08	
	OsPLAC88			OsCNR14	
	OsPLAC89			OsCNR15	
	OsPLAC810				OsMCA1

Supplementary Material 1, tab 1 and 2: Detailed data on species, type, acronym, and gene/locus of PLAC8 genes.

CAPÍTULO 5

Manuscrito submetido ao periódico *Functional and Integrative Genomics*

GILP Family: A Stress-Responsive Group of Plant Proteins Containing a LITAF Motif

Cabreira-Cagliari, C.[1]; Fagundes, D. G. S.[2], Dias, N. C. F.[2]; Bohn, B. [2]; Margis-Pinheiro, M[1]; Bodanese-Zanettini, M.H.[1] and Cagliari, A[2].

[1] Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil[2] Universidade Estadual do Rio Grande do Sul (UERGS), Santa Cruz do Sul, RS, Brazil

Corresponding author: Alexandro Cagliari, e-mail: alexandro-cagliari@uergs.edu.br, phone: +55-51-3715-6926, address: Universidade Estadual do Rio Grande do Sul - UERGS, CEP 96816-501, Santa Cruz do Sul, RS, Brazil.

Abstract

Lipopolysaccharide-induced tumor necrosis factor-α (LITAF) is a membrane protein that is highly dependent on correct location to exert transcription factor activity and protein quality control. In humans, *LITAF*, *PIG7* (p53-inducible gene 7), and *SIMPLE* (small integral membrane protein of the lysosome/late endosome) refer to the same gene, which acts as a tumor suppressor. In animals, several studies have shown that the transcription factor activity and nuclear translocation of LITAF protein are critical for the induction of several immune cells via classical pathways. In plants, LITAF protein corresponds to the plasma membrane protein AtGILP (*Arabidopsis thaliana* GSH-induced LITAF domain protein). The conservation of LITAF proteins across species and their putative role is still unclear. In this study, we investigate the LITAF containing-proteins, which we call GILP proteins, in Viridiplantae. We identified a total of 59 genes in 46 species, whose gene copies range from one to three. Phylogenetic analysis showed that multiple copies were originated via block duplication posteriorly to monocot and eudicot separation. Analysis of the LITAF domain of GILP proteins allowed the identification of a putative domain signature in Viridiplantae, containing a CXXCX41HXCPXC motif. The subcellular location for the majority

of GILP proteins was predicted to be in the plasma membrane, based on a transmembrane domain positioned within the LITAF domain. *In silico* analysis showed that the GILP genes are neither tissue-specific nor ubiquitously expressed, being responsive to stresses conditions. Finally, investigation of the GILP protein network resulted in the identification of genes whose families are known to be involved with biotic and/or abiotic stress responses. Together, the expression modulation of GILP genes, associated with their plasma membrane location suggests that they could act in the signaling of biotic/abiotic stress response in plants.

Introduction

The basis of innate immunity in plants occurs through a single overarching inception, the recognition of danger signals (Andolfo and Ercolano, 2015). Plants perceive pathogen-associated molecular patterns (PAMPs) through transmembrane pattern recognition receptors (PRRs) and initiate PAMP-triggered immunity (PTI). PTI acts as the first defense line against pathogens (He et al., 2011). The second line of defense is known as effector-triggered immunity (ETI), which is one of the most effective innate immune strategies in plants, normally leading to a hypersensitive response (HR) (He et al., 2011). At the infection site, through the initiation of HR, a genetically regulated process of cellular suicide known as programmed cell death (PCD) is activated to remove mutated, infected or damaged cells.

Recently, knowledge about PCD in plants has increased, especially in *Arabidopsis thaliana*. Coll et al. (2011) identified several genes involved with PCD control, which constitute what the authors named the "deathsome". The *LSD1* (*Lesion simulating disease 1*) gene, which negatively regulates PCD under stress conditions (Dietrich et al., 1997; Epple et al., 2003; Cabreira et al., 2013), is a central gene in the deathsome. LSD1 physically interacts with bZIP10 (a basic region leucine zipper transcription factor) (Kaminaka et al., 2006), and type I metacaspase (MC1) (Coll et al., 2010), retaining these proteins in the cytoplasm and inhibiting their action as positive regulators of HR.

A protein encoded by the *At5g13190* gene, and containing the LITAF [lipopolysaccharide (LPS)-induced tumor necrosis factor alpha (TNF-a) factor] domain was also described as *At*LSD1 interaction-protein (He et al., 2011). The LITAF domain is composed of 68 amino acids, with an *N*-terminal CxxC knuckle, a hydrophobic region, and a *C*-terminal (H)xCxxC knuckle (Ponting et al., 2001).

The human gene coding the LITAF motif-containing protein has been described in three independent studies. Firstly, the gene was called *PIG7* (p53-inducible gene 7), which was positively regulated by the tumor suppressor protein p53 (Polyak et al., 1997). Second, this same gene was named as an LPS-regulated gene, which encodes a protein that translocates into the nucleus, following cellular activation by LPS and upregulation of TNF-α transcription, hence their name *LITAF* (Myokai et al., 1999). Finally, this gene was described as an unglycosylated small integral membrane protein of the lysosome/late endosome (*SIMPLE*) (Moriwaki et al., 2001). Therefore, *PIG7, LITAF* and *SIMPLE* refer to the same gene, whose *N*-terminal protein region contains two PPXY domains responsible for binding to WWOX, NEDD4 and TSG101 proteins (Ludes-Meyers et al., 2004; Shirk et al., 2005). In this manuscript, we will refer to LITAF motif-containing protein in human as LITAF protein.

Several studies in humans have shown that the transcription factor activity and nuclear translocation of LITAF protein are critical for the induction of several immune cells via classical pathways (Zou et al., 2015). LITAF protein has become a novel target for developing better therapeutics for the treatment of systemic and local forms of inflammation (Merril et al., 2011). Based on the fact that LITAF expression can be induced by p53, LITAF has been identified as a potential tumor suppressor gene (Polyak et al., 1997). LITAF expression is significantly lower in tumor tissues when compared with isogenic normal tissues (Abba et al., 2004; Fernandez-Cobo et al., 2006; Wang et al., 2009). In prostate cancer, shRNA-LITAF cells significantly enhanced cancer malignancy (Zhou et al., 2001). In patients with acute leukemia, LITAF sensitizes leukemic cells to chemotherapeutic agents (Wang et al., 2009). Therefore, LITAF protein may serve as a switch in the balance between classical inflammation and alternative activation in cancer, acting as a transcription factor or as a recruiting factor targeting partner proteins to the lysosome for degradation (Zou et al., 2015).

A single report has investigated a plasma membrane LITAF protein called GILP (GSH-induced LITAF domain protein) in *A. thaliana* (He et al., 2011). The authors showed that AtGILP uses both its *N*- and *C*-terminal domains to interact with the *N*-terminal domain of AtLSD1. Moreover, *AtGILP* was upregulated in response to avirulent pathogen and fumonisin B_1 (FB_1), which are known to trigger PCD. The overexpression of *AtGILP* suppressed pathogen-induced PCD. Taken together, these results suggest that AtGILP may acts as a membrane anchor, bringing other regulators of PCD to the plasma membrane (He et al., 2011).

Since the first report in humans around twenty years ago (Polyak et al., 1997), the existence and putative roles of LITAF proteins across species remains unexplored. In this context, comparative functional genomics associated with a high-throughput phylogenetic analysis covering several genomes can contribute to understanding the relationship between the LITAF proteins. In this study, we investigate the LITAF-containing proteins, which we call GILP proteins, in Viridiplantae. We propose the nomenclature of the plant *GILP* genes and their phylogenetic reconstruction and characterization are presented. *In silico* expression analysis and protein-protein interaction networks are also shown. These insights could be useful in further functional studies of individual GILP proteins in plants.

Material and Methods

Identification of GILP Genes

To identify gene sequences encoding LITAF proteins, the complete nucleotide and protein sequences of At5g13190 (*AtGILP* gene) were used as queries to perform BLAST searches against 46 species with sequenced genomes. tBLASTx and BLASTn were conducted in NCBI (National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) and Phytozome (http://www.phytozome.org/) databases. Genes were individually analyzed for the presence of the previously characterized LITAF domain (He et al., 2011) using SMART (http:// smart.embl-heidelberg.de/) database. We adopted the acronym GILP, previously used by He et al. (2011) to refer to the LITAF protein in *A. thaliana* (*AtGILP* gene). Nomenclature of identified genes was based on BLAST output. Gene structure composition was analyzed using the Exon-Intron Graphic Maker (http://wormweb.org/exonintron).

Evolutionary Processes

The complete amino acid sequences encoding the LITAF domains from representative species were aligned using the Muscle algorithm, implemented in MEGA v.5.05 (Tamura et al., 2011). A Bayesian inference was generated using BEAST v.1.4.7 (Drummond and Rambaut 2007). A run of 10^7 chains was performed, and the trees were sampled every 1000 generations. The Yule tree, the JTT substitution model and the uncorrelated log-normal relaxed clock were used in the BEAST analysis. The TRACER v.1.4 (http://beast.bio.ed.ac.uk/Tracer) was used to

check the convergence of the Monte Carlo Markov Chains (MCMCs) and for adequate effective sample sizes (EES > 200) after the first 10% of the generations had been deleted as burn-in. The final joint sample was used to estimate the maximum clade credibility tree with the TreeAnnotator program, which is part of the BEAST package. The statistical support for the clades was determined by accessing the Bayesian posterior probability. The trees were visualized using FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

We also investigated the mechanisms involved with the evolution of multiples copies of *GILP* genes from species available in the PLAZA v3.0 database (http://bioinformatics.psb.ugent.be/plaza/). The physical co-localization of *GILP* genes was analyzed in the Genome Duplication Database (PGDD) (http://chibba. agtec.uga.edu/duplication/), considering a 100 kb syntenic region.

GILP Protein Characterization

LITAF motif regions were selected based on SMART prediction. Their amino acid sequences were analyzed in MEGA v.5.05 (Tamura et al., 2011) and manually aligned to access domain conservation using MEME software (http://meme.sdsc.edu/meme/). To investigate subcellular localization, TargetP (http://www.cbs.dtu.dk/services/TargetP/) and Cello (http://cello.life.nctu.edu.tw/) prediction tools were used. TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0/) and Phobius (http://phobius.sbc.su.se/) were used to detect transmembrane domains.

In silico Functional Investigation of GILP Genes

To gain insights about gene expression, *GILP* genes were examined using the RNA-Seq database BAR (The Botany Array Resource, http://bar.utoronto.ca/). Plant organs and tissues of the representative species *Arabidopsis*, *Glycine max* and *Oryza sativa* in normal and stress conditions were examined (Toufighi et al., 2005). To investigate the proteins that potentially interact with GILP proteins, the functional protein association network of *Oryza sativa, Arabidopsis* and *Glycine max* were investigated. These analyses were conducted in STRING (string-db.org).

Results

A Small Family of GILP Genes in Viridiplantae

Using the *AtGILP* gene (the unique LITAF protein investigated in plants until now) as bait to BLAST search, 59 genes were identified in Viridiplantae (Supplementary material 1). *GILP* genes were absent in basal organisms, such as the chrolophyte species *Volvox Carteri* and *Chlamydomonas reinhardtii*. At least one gene copy was identified from bryophyte species. Duplication of GILP sequences was found from monocot to eudicot species. The hexaploid bread wheat *Triticum aestivum* presented the highest number of copies identified (three genes). Monocots, wheat and *Ananas comosus* were the unique species with more than one gene. In eudicots a maximum of two *GILP* genes were found in *Amaranthus hypochondriacus, Gossypium raimondii, Kalanchoe marnieriana, Brassica rapa, Malus domestica, Glycine max, Populus trichocarpa, Salix purpurea* and *Manihot esculenta.*

In regard to the exon-intron structure, the number of exons ranged from two to three (Figure 1). Considering the representative genes analyzed, the majority presented three exons, even in the basal organisms. Some genes exhibited a wide intronic region, as observed in *ZmGILP1* and *SbGILP1*. In the duplicated genes, very similar structures were observed, as the *PtGILP1* and *PtGILP2* genes.

The Evolutionary History of *GILP* Genes

The unrooted phylogenetic tree showed a clear separation between bryophyte, lycophyte, monocot and eudicot species (Figure 1). A high posterior probability value supported this separation. *GILP* genes from monocot and eudicot species were separated in two well-supported subclades, indicating that these genes are more related inside the specific group.

The investigation concerning what kind of duplication (tandem or block duplication processes) contributed to the acquisition of multiples copies of *GILP* genes in *Glycine max, Populus trichocarpa, Brassica rapa* and *Gossypium raimondii* suggested that only block duplication occurred in this gene family (Figure 2A). The syntenic regions are located in a huge duplication block (*Glycine max, Populus trichocarpa, Brassica rapa*) or in a large duplication block (*Gossypium raimondii*). In regards to the other species with duplicated genes, no results were found in PLAZA and/or PGDD database. Finally, the analysis of *Ka/Ks* rate between

duplicated genes showed values < 1, indicating that a subfunctionalization process could be acting in these sequences (Figure 2B).

Many GILP Proteins Have A Transmembrane Domain Present

We analyzed all putative GILP proteins to investigate the LITAF motif and their rate of conservation at each amino acid position. According to the SMART tool prediction, the motif region is composed of 66 amino acids (Figure 3). Comparison of all GILP proteins indicated a total conservation of residues in some positions, as detailed in Figure 3. Positions 6 to 9 (CxxC) and 52 to 57 (HxCPxC) showed a conserved cysteine, histidine and proline region, which corresponds to the two knuckles. In addition, the initial (PAPF) and final (DFEK) amino acid positions were also well conserved.

The subcellular localization of all GILP proteins was also investigated. The TargetP tool uses the presence of any of the *N*-terminal presequences to predict targeting to chloroplast (cTP), mitochondrial (mTP) or secretory pathways (SP). None of these target sequences were found in GILP proteins (Supplementary table 2). Subsequently, subcellular localization prediction was performed using the CELLO database, which identifies sequences targeting proteins to chloroplasts, the cytoplasm, the cytoskeleton, the endoplasmic reticulum, the extracellular/secretory space, the golgi, lysosomes, mitochondria, the nucleus, peroxisomes, the plasma membrane, and vacuoles. The results showed that the majority of GILP proteins (40 out of 59) were predicted to be localized to the plasma membrane (Supplementary table 2).
Transmembrane regions were identified in the majority of GILP proteins, corroborating the prediction that GILP proteins are putatively located in the plasma membrane (Supplementary table 2). The region containing the transmembrane domain in the majority of GILP proteins is located inside the LITAF domain (Figure 3), around the 69-88 amino acid position for short proteins and around the 89-107 position in long proteins (as illustrated in Figure 4).

The *GILP* Genes are Neither Tissue-Specific Nor Ubiquitously Expressed

In order to generate a wide characterization of the *GILP* family, the relative expression of *GILP* genes was investigated in representative species using the RNA-Seq BAR database. Considering monocot species, rice *OsGILP1* was highly expressed in the inflorescence (P5), root,

and stigma and downregulated in seeds (S5) (Figure 5A). In addition, under abiotic stress, *OsGILP1* was downregulated in salt, upregulated in cold stress and in coleoptile in aerobic conditions.

GILP genes were investigated in eudicot species, soybean and *Arabidopsis*. In soybean, the expression showed slight variation among tissues/organs analyzed, except for nodules inoculated with *Bradyrhizobium japonicum* (*GmGILP1* and *GmGILP2*) and in roots (*GmGILP2*) (Figure 5B). *At*GILP1 showed a variable expression in tissues/organs (Figure 6). *At*GILP1 expression was upregulated in leaves and during the final stages of flower development. In the initial stages of flower development, seed development, and shoot apex, *AtGILP1* was down regulated.

Under abiotic stress conditions, *AtGILP1* was upregulated in all conditions analyzed (hypoxia, selenate, auxin (IAA) and cytokinin treatment), except in guard and mesophyll cells sprayed with abscisic acid (ABA), where the expression was downregulated (Figure 6). Under biotic stress, *AtGILP1* was upregulated after *Golovinomyces orontii* and *Hyaloperonospora arabidopsidis* infection.

We also investigated the network of genes that are coexpressed or interact with GILP proteins (OsGILP1, AtGILP1, GmGILP1 and GmGILP2) according to database predictions. For this purpose, we considered the results based on experimental evidence or coexpression by microarray, and text-mining. We observed that few genes were present in GILP protein networks (Figure 8). OsGILP1 network is composed of an amine oxidase and genes from the *LSD* family (Figure 8A). Proteins involved with membrane processes were found in the network of AtGILP1, including the transmembrane proteins At1g50740, SYP122 and SYP121 (Figure 8B). Proteins related to PCD processes were also found in the AtGILP1 network, including At4g34150 and At2g23810. In GmGILP1 and GmGILP2 networks, ring and zinc finger proteins were observed, as well as several proteins from the Lysine Specific Demethylase family (Figure 8C). GmGILP1 and GmGILP2 networks share the majority of proteins, except the Lysine Specific Demethylase 1 Glyma02g18210 (GmGILP2) and the Lysine Specific Demethylase 1 Glyma02g59100 (GmGILP1).

Genes belonging to the *LSD* (*Lesion simulating disease*) gene family are present in OsGILP1, AtGILP1, GmGILP1 and GmGILP2 networks. Based on experimental predictions, OsLSD3 and OsLSD4 were found in the OsGILP1 network, while GmLSD1, GmLSD2 and GmLSD4 were found in GmGILP1 and GmGILP2 networks. Considering the AtGILP1 network,

an AtLSD1 putative interaction-protein was also identified based on text-mining and coexpression data.

Discussion

Gene duplications originating from genome polyploidization or from region-specific duplication is a remarkable feature in plant genome evolution. Gene duplication is an important mechanism for the appearance of new genes, as well as genetic novelty in organisms (Lawton-Rauh A 2003; Magadum et al., 2013). The plasticity of a genome or species in response to alterations in environments would be impaired without gene duplication (Magadum et al., 2013). Investigation of the mechanisms that generate duplicate copies and the subsequent dynamics and fate of the duplicated genes is vital to clarify many aspects of evolutionary forces shaping intra-specific and inter-specific genome content, evolutionary relationships, and interactions (Magadum et al., 2013). In this study, we identified for the first time, a small gene family in plants, which contains the LITAF domain and we have named the *GILP* family. LITAF domain-containing proteins have been extensively investigated as p53-inducible genes in humans (Polyak et al., 1997). However, LITAF-containing proteins in plants have not been described until now.

Several questions considering the molecular evolutionary dynamics of gene duplication are unclear, including the relative roles in the maintenance of duplicated genes, and whether different gene classes (such as transcriptional regulators versus structural housekeeping genes) have distinct evolutionary rates (Lawton-Rauh A 2003). While some families present a big number of genes, such as the WRKY gene family (3035 WRKY genes in 43 species) (Mohanta et al., 2016), some families are small, e.g. the LSD gene family (113 LSD genes in 32 species) (Cabreira et al., 2015). Regarding the GILP proteins, we have identified a small gene family composed of 59 genes in 46 species, and the gene copies range from one to three in each species. The monocot *Triticum aestivum* presented the highest number of GILP genes (three copies), which is not surprising, since this species originated from two recent hybridization events between three diploid progenitors, donors of the A, B, and D subgenomes (Marcussen et al., 2014).

Duplicated genes have the property of initially being functionally redundant, allowing compensation for deleterious variation in a sister copy (Bozorgmehr, 2012). Plants are not the only organisms in which LITAF-containing proteins were duplicated. In dipteran insects, an expansion

likely due to an ancient gene duplication event occurred, in contrast to mammals and other invertebrates that contain a single *LITAF* gene (Smith et al., 2012). Since gene duplication did not occur with the *LITAF* gene in humans, functional redundancy and compensation by a sister copy is absent. In the case of deleterious mutations, variations in the LITAF gene are responsible for important disease phenotypes. In Charcot-Marie-Tooth (CMT), LITAF mutations correspond to 1-2% of the reported cases. These mutations occur mostly in the LITAF *C*-terminus region (SLD), specifically around the hydrophobic domain that is flanked by the two CX2C motifs (Lacerda et al., 2014). Some human LITAF mutants mislocalize to the mitochondria instead of late endosomes/lysosomes, while others show partial mislocalization with a portion of the protein in the late endosome/lysosome and the rest localized to mitochondria (Lacerda et al., 2014; Sinkiewicz-Darol et al., 2015). LITAF mutations could affect the regulation of endosomal trafficking, which not only controls protein degradation but also regulates intracellular signaling (Zhu et al., 2013; Chin et al., 2013). Additionally, the accumulation of non-functional LITAF protein, resulting from mutations, has been reported in cancer (Matsumura et al., 2004).

Due to the existence of duplicated *GILP* genes in plants, the mutation of a single gene might be compensated by their paralogs. The *in silico* investigation of *Gm*GILP1 and *Gm*GILP2 duplicated genes showed a similar expression pattern, suggesting a putative expression redundancy (Figure 8). However, the resultant mutant phenotype and/or functional compensation from a duplicated copy are completely unknown, as GILP mutants have never been phenotypically analyzed.

Here we present the first phylogenetic reconstruction of LITAF-containing proteins (GILP), including sequences from eudicot, monocot, lycophyta and bryophyta (Figure 2A). Extensive BLAST searches showed that *GILP* genes are absent in algae organisms, suggesting that the rise of this gene family occurred in bryophyte and expanded from monocot. The unrooted tree formed two clades. The first clade contains the lycophyta and bryophyte organisms. The second one is subdivided in two groups, composed of eudicot or monocot *GILP* genes. These results indicate that the gene duplication process, which was the origin of the multiplication in each species, occurred posterior to monocot and eudicot separation (Figure 1). The *GILP* gene structure containing three exons potentially represents the ancestral condition for the *GILP* family. It is present in lycophyta and bryophyta and was maintained during the evolution in most of the identified genes. Taken together, the phylogenetic reconstruction shows the emergence of the

GILP family in embryophyte, followed by an expansion into multiple copies, in contrast to humans and other vertebrates.

The subcellular localization was predicted to be in the plasma membrane for the majority of GILP proteins (Figure 3, double black bar and Figure 4). The predicted transmembrane region was necessary for *At*GILP localization in the plasma membrane (He et al., 2011). Under biotic and abiotic stresses, the plasma membrane represents a boundary for plants to sense changes in their environment. PRRs act as part of multi-protein complexes at the plasma membrane, signaling the recognition and response to stresses in a highly sensitive and specific manner (Monaghan and Zipfel 2012). In this context, plasma membrane proteins, such as *At*GILP1 (which interacts with *At*LSD1, the negative regulator of PCD), function in the regulation of PCD triggered by stresses (He at al., 2011). Our expression analysis shows that *Os*GILP1, *At*GILP1, *Gm*GILP1 and *Gm*GILP2 are modulated in response to stresses and/or stress signaling conditions, such as drought, salt and cold (*Os*GILP1) and IAA, cytokinin, hypoxia and *Hyaloperonospora arabidopsidis* fungus infection (*At*GILP1) (Figures 5 and 7). The expression modulation of GILP genes associated with the cellular localization at the plasma membrane (Supplementary material 2) suggests that they could be acting in stress response signaling. It has already been proposed that LITAF proteins may act as membrane anchors, recruiting other proteins to the membrane via protein–protein interaction (He at al., 2011).

GILP-interaction network investigation allowed the identification of several genes, whose families are known to be involved with biotic and/or abiotic stress responses (Figure 8). Amino oxidase, which was identified in the OsGILP1 network (Figure 8A), is involved in symbiotic interactions and plant defense responses (Cona et al., 2006). Lysine specific histone demethylase 1 (found in the GmGILP1 and GmGILP2 networks) is included in a family known to act in chromatin modification, which regulates plant immunity against biotic stress (Ding and Wang 2015). In regard to the AtGILP1 network, TET8 for example, is a transmembrane protein involved in development, pathogenesis and immune responses (Wang et al., 2012). In addition, *SYP122* and *SYP121* are members of the syntaxin family and are involved in the fusion of transport vesicles to target membranes. In soybean, roots engineered to overexpress *SYP38* (*syntaxin 31*) induced the expression of a *LSD* gene (Pant et al., 2015), suggesting that they could act together. Therefore, the identification of several proteins related to PCD converges with the modulation of gene expression during stress conditions. These results reinforce the putative involvement of the *GILP* gene family in PCD.

The *LSD* genes are present in all GILP networks investigated (Figure 8). LSD proteins are zinc finger proteins related to negative regulation of PCD (Dietrich et al., 1997; Epple et al., 2003; Coll et al., 2010; Cabreira et al., 2013). He et al. (2011) showed the physical interaction of AtGILP1 and AtLSD1 by both the *N*-terminal and the *C*-terminal domains of AtGILP with the *N*-terminal domain of AtLSD1. The existence of LSD sequences in all analyzed networks, suggests that during PCD, GILP members act in coordination with with LSD genes. It is known that *AtGILP* negatively regulates hypersensitive cell death, similarly to *AtLSD1* (He et al., 2011). Therefore, GILP interaction network analysis reinforces that the *GILP* genes could act at the plasma membrane in parallel with other regulators of PCD, especially LSD members.

Human LITAF acts as a LPS-induced transcription factor involved in activating TNF- *α* gene expression, since it directly binding to specific sequences in the TNF- *α* promoter region (Myokai et al., 1999; Stucchi et al., 2006). LITAF also binds to the *TNF-α* promoter in response to bacterial LPS stimulation to influence the expression of *TNF-α* (Tang et al., 2003), as well as additional LPS-induced cytokines (Tang et al., 2006). *Anopheles gambiae* LITAF-like 3 was also described as a transcription factor, which modulate SRPN6 transcript abundance, indicating that it acts as part of the mosquito anti-*Plasmodium* immune response (Smith et al., 2012). In plants, the possible binding of a GILP sequence to the promoter of a target gene was not investigated until now. Therefore, it is still unclear if *GILP* genes from plants act as transcription factors, as observed in other species.

Another open question is whether paralogs play the same function during PCD. For example, the expression modulation of the *Gm*GILP1 and *Gm*GILP2 paralogs during *Bradyrhizobium japonicum* infection has not indicated that these genes act as each other. In the *LSD* family, it is known that *AtLSD1* and *AtLOL1* have opposite roles during PCD, since *AtLSD1* regulates and *AtLOL1* induces, cell death (Epple et al., 2003). Future detailed studies should analyze the function of paralogous genes during PCD.

In the present work we described a new gene family, composed of one (minimum) or three (maximum) genes in each Viridiplantae species. Our results indicate that this family expanded by duplication through the evolution, leading to the acquisition of new gene copies. A domain signature of GILP proteins from plants was described and it is different to vertebrates. In accordance with previous studies, we showed that the *GILP* family represents a class of plasma membrane proteins. Finally, the modulation of *GILP* gene expression in stresses conditions and a network of protein-protein interactions composed of sequences related to PCD control were

described. Together, our results indicate that *GILP* genes from plants could act in the PCD network, perhaps bringing others important proteins to the transmembrane region. Future deep characterization of these newly identified GILP proteins could serve to clarify the role of this gene family in the PCD process.

Bibliography

Abba, M.C. Drake, J. A.; Hawkins, K. A.; Yuhui, H.; Hongxia, S.; Notcovich, C.; Gaddis, S.; Sahin, A.; Baggerly, K.; Aldaz, C. M. (2004) Transcriptomic changes in human breast cancer progression as determined by serial analysis of gene expression. Breast cancer research: BCR, 6(5), pp.R499-513.

Andolfo, G.; Ercolano, M.R. (2015) Plant Innate Immunity Multicomponent Model. Frontiers in plant science, 6(November), p.987.

Cabreira, C.; Cagliari, A.; Bücker-Neto, L.; Wiebke-Strohm, B.; de Freitas, L.B.; Marcelino-Guimarães, F. C.; Nepomuceno; A. L.; Margis-Pinheiro, M. M.; Bodanese-Zanettini, M. H. (2013) The Lesion Simulating Disease (LSD) gene family as a variable in soybean response to Phakopsora pachyrhizi infection and dehydration. Functional and Integrative Genomics, 13(3), pp.323–338.

Chin, L. S., Lee, S. M.; Li, L. (2013) Simple: A new regulator of endosomal trafficking and signaling in health and disease. Communicative and Integrative Biology, 6(3).

Coll, N. S.; Vercammen, D.; Smidler, A.; Clover, C.; Van Breusegem, F.; Dangl, J. L.; Epple, P. (2010). Arabidopsis type I metacaspases control cell death. Science (New York, N.Y.), 330(6009), pp.1393–1397.

Coll, N.S.; Epple, P.; Dangl, J. L. (2011) Programmed cell death in the plant immune system. Cell death and differentiation, 18(8), pp.1247–56.

Cona, A. Rea, G.; Angelini, R.; Federico, R.; Tavladoraki, P. (2006) Functions of amine oxidases in plant development and defence. Trends in Plant Science, 11(2), pp.80–88.

Dietrich, R.A. Richberg, M. H.; Schmidt, R.; Dean, C.; Dangl, J. L. (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(5), pp.685–694.

Ding, B.; Wang, G-L. (2015) Chromatin versus pathogens: the function of epigenetics in plant immunity. Front Plant Sci, 6: 675.

Drummond, A. J.; Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC evolutionary biology, 7(1), p.214.

Epple, P. Mack, A. A.; Morris, V. R.; 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(11), pp.6831–6836.

Fagundes, D.; Bohn, B.; Cabreira, C.; Leipelt, F.; Dias, N.; Bodanese-Zanettini, M. H.; Cagliari, A. (2015) Caspases in plants: metacaspase gene family in plant stress responses. Functional and Integrative Genomics, 15(6), pp.639–649.

Fernandez-Cobo, M.; Holland, J. F.; Pogo, B. G. (2006) Transcription profiles of non-immortalized breast cancer cell lines. BMC Cancer. 6:992006.

He, S.; Tan, G.; Liu, Q.; Huang, K.; Ren, J.; Zhang, X.; Yu, X.; Huang, P.; An, C. (2011) The lsd1-interacting protein gilp is a litaf domain protein that negatively regulates hypersensitive cell death in arabidopsis. PLoS ONE, 6(4).

Kaminaka, H.; Nake, C.; Epple, P.; Dittgen, J.; Schutze, K.; Chaban, C.; Holt, B. F III.; Merkle, T.; Schafer, E.; Harter, K.; Dangl, J. L. (2006) bZIP10-LSD1 antagonism modulates basal defense and cell death in Arabidopsis following infection. The EMBO journal, 25(18), pp.4400–4411.

Toufighi, K.; Brady,S. M.; Austin, R.; Ly, E.; Provart, N. J. (2005) The Botany Array Resource: e-Northerns, Expression Angling, and promoter analyses. The Plant Journal, 43, 153–163.

Lacerda, A. F.; Hartjes, E.; Brunetti, C. R. (2014) LITAF mutations associated with Charcot-Marie-Tooth disease 1C show mislocalization from the late endosome/lysosome to the mitochondria. PloS one, 9(7), p.e103454.

Lawton-Rauh, A. (2003). Evolutionary Dynamics of Duplicated Genes in Plants. Molecular Phylogenetics and Evolution, 29(3), pp.396–409.

Ludes-Meyers, J.H. Kil, H.; Bednarek, A. K.; Drake, J. Bedford, M. T.; Aldaz, C. M. (2004) WWOX binds the specific proline-rich ligand PPXY: identification of candidate interacting proteins. Oncogene, 23(29), pp.5049–55.

Magadum, S.; Banerjee, U.; Murugan, P.; Gangapur, D.; Ravikesavan, R. (2013) Gene duplication as a major force in evolution. Journal of Genetics, 92(1), pp.155–161.

Marcussen, T.; Sandve S. R.; Heier, L.; Spannagl, M.; Pfeifer, M.; International Wheat Genome Sequencing Consortium.; Jakobsen, K. S.; Wulff, B. B.; Steuernagel, B.; Mayer, K. F.; Olsen, O. A. (2014) Ancient hybridizations among the ancestral genomes of bread wheat. Science, 345(6194), p.1250092.

Matsumura, Y.; Matsumura, Y.; Nishigori, C.; Horio, T.; Miyachi, Y. (2004) PIG7/LITAF gene mutation and overexpression of its gene product in extramammary Paget's disease. Int J Cancer. 111:218–223.

Merrill, J. C.; Jou, J.; Constable, C.; Leeman, S. E.; Amar, S. (2011) Whole-body deletion of LPS-induced TNF- factor (LITAF) markedly improves experimental endotoxic shock and inflammatory arthritis. Proceedings of the National Academy of Sciences, 108(52), pp.21247–21252.

Mohanta, T. K.; Park, Y-H.; Bae, H. (2016) Novel Genomic and Evolutionary Insight of WRKY

Transcription Factors in Plant Lineage. Sci Rep 6, 37309.

Monaghan, J.; Zipfel, C. (2012) Plant pattern recognition receptor complexes at the plasma membrane. Curr Opin Plant Biol, 15(4), pp.349–357.

Moriwaki, Y.; Begum, N. A.; Kobayashi, M.; Matsumoto, M.; Toyoshima, K.; Seya, T. 2001. Mycobacterium bovis Bacillus Calmette-Guerin and its cell wall complex induce a novel lysosomal membrane protein, SIMPLE, that bridges the missing link between lipopolysaccharide and p53-inducible gene, LITAF(PIG7), and estrogen-inducible gene, EET-1. J Biol Chem, 276(25):23065-76.

Myokai, F. Takashiba, S.; Lebo, R.; Amar, S. (1999) A novel lipopolysaccharide-induced transcription factor regulating tumor necrosis factor alpha gene expression: molecular cloning, sequencing, characterization, and chromosomal assignment. Proc Natl Acad Sci U S A, 96(8), pp.4518–23.

Polyak, K. Xia, Y.; Zweier, J. L.; Kinzler, K. W.; Vogelstein, B. (1997) A model for p53-induced apoptosis. Nature, 389(6648), pp.300–305.

Ponting, C.P.; Mott, R.; Bork, P.; Copley, R. R. (2001) Novel protein domains and repeats in Drosophila melanogaster: insights into structure, function, and evolution. Genome Res, 11(12), pp.1996–2008.

Shirk, A.J. Anderson, S. K.; Hashemi, S. H.; Chance, P. F.; Bennett, C. L. (2005) SIMPLE interacts with NEDD4 and TSG101: Evidence for a role in lysosomal sorting and implications for Charcot-Marie-Tooth disease. Journal of Neuroscience Research, 82(1), pp.43–50.

Sinkiewicz-Darol, E.; Lacerda, A. F.; Kostera-Pruszczyk, A.; Potulska-Chromik, A.; Sokołowska, B.; Kabzińska, D.; Brunetti, C. R.; Hausmanowa-Petrusewicz, I.; Kochański, A. (2015) The LITAF/SIMPLE I92V sequence variant results in an earlier age of onset of CMT1A/HNPP diseases. Neurogenetics, 16: 27–32.

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

Wang, D.; Liu, J.; Tang, K.; Xu, Z.; Xiong, X.; Rao, Q.; Wang, M.; Wang, J. (2009) Expression of pig7 gene in acute leukemia and its potential to modulate the chemosensitivity of leukemic cells. Leukemia Research, 33(1), pp.28–38.

Wang, D.; Liu, J.; Tang, K.; Xu, Z.; Xiong, X.; Rao, Q.; Wang, M.; Wang, J. (2009) Expression of pig7 gene in acute leukemia and its potential to modulate the chemosensitivity of leukemic cells. Leuk Res. 33:28–38.

Wang, F.; Vandepoele, K.; Van Lijsebettens, M. 2012. Tetraspanin genes in plants. Plant Science, 190, pp.9–15.

Zhu, H.; Guariglia, S.; Yu, R. Y.; Li, W.; Brancho, D.; Peinado, H.; Lyden, D.; Salzer, J.; Bennett, C.; Chow, C. W. (2013) Mutation of SIMPLE in Charcot-Marie-Tooth 1C alters production of exosomes. Molecular biology of the cell, 24(11), pp.1619–37, S1-3.

Zhou, J.; Yang, Z.; Tsuji, T.; Gong, J.; Xie, J.; Chen, C.; Li, W.; Amar, S.; Luo, Z. (2011) LITAF and TNFSF15, two downstream targets of AMPK, exert inhibitory effects on tumor growth. Oncogene. 30:1892–1900.

Zou, J.; Guo, P.; Lv, N.; Huang, D. (2015) Lipopolysaccharide-induce tumor necrosis factor-α enhances inflammation and is associated with cancer. Molecular Medicine Reports 12(5).

Figures and legends

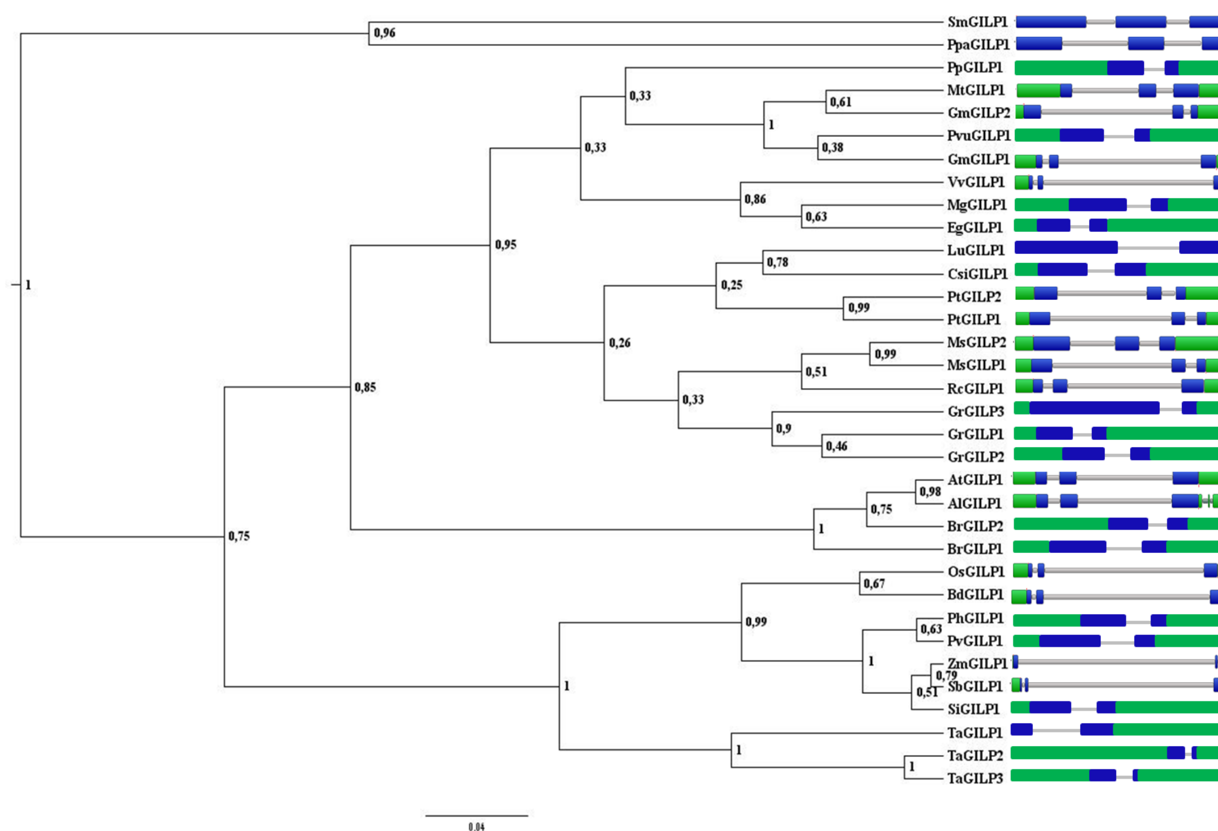
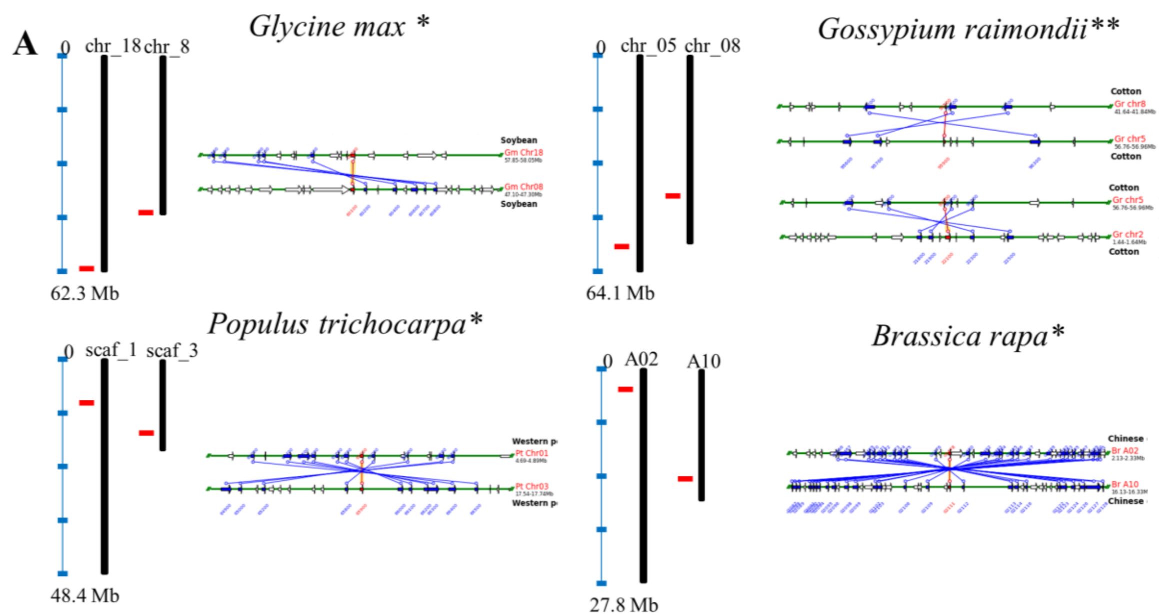


Figure 1. Unrooted tree (Bayesian inference) showing the evolutionary relationships among *GILP* genes from representatives species. The posterior probability is shown above the branches. The taxa terminologies are abbreviated using the first letter of the genus and the first two letters of the species (see Supplementary Material 1). The right side provides a detailed illustration of the relative intron/exon length and number in each gene. The intron/exon structure were generated using the PLAZA database or manually drawn. The green boxes represent the 5´and 3´UTR. The blue boxes shows exons and the lines connecting them represent the introns.



B

Sequences	Ka	Ks	Ka/Ks
GmGILP1xGmGILP2	0.04	0.09	0.44
PtGILP1xPtGILP2	0.03	0.22	0.13
GrGILP1xGrGILP2	0.09	0.57	0.15
GrGILP2xGrGILP3	0.12	0.72	0.16
BrGILP1xBrGILP2	0.06	0.26	0.23

Figure 2. A) *GILP* genes location analyses conducted in PLAZA database. *GILP* genes from *Glycine max, Populus trichocarpa, Gossypium raimondii* and *Brassica rapa* were showed. Syntenic regions between *GILP* genes analyzed using the PGDD database. The green horizontal line represents the chromosome. The blue and red vertical lines and arrows represent the duplicated orthologous genes; the red line represents the sequence that was used as the search query (indicated at the top of each syntenic region). * huge duplication block (*Glycine max, Populus trichocarpa, Brassica rapa*) ** large duplication block (*Gossypium raimondii*). The white arrow indicates genes without correspondent paralogous genes in this duplicated region. B) *Ka* and *Ks* values, and *Ka/Ks* ratios of the paralogous *GmNF-Y* genes analyzed in the PGDD database.

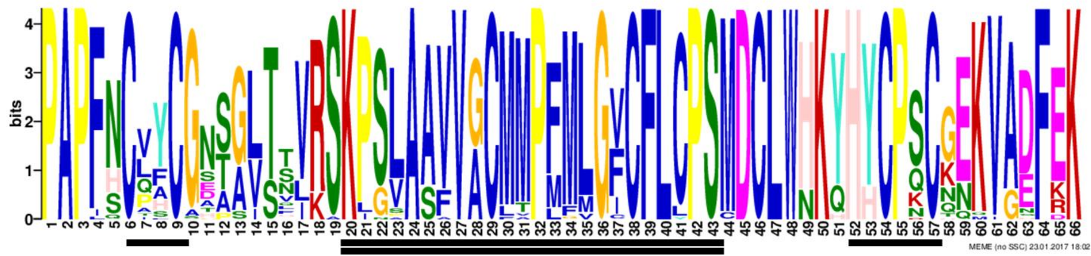


Figure 3. GILP proteins domain analysis. The total height of each cell indicates the conserved sequence at each position. The height of each letter is proportional to the corresponding relative frequency. The amino acids are colored according to their chemical properties: Blue for most hydrophobic residues (A, C, F, I, L, V, and M); Green for polar, non-charged, non-aliphatic residues (N, Q, S, and T); Red for positively charged residues (K and R); Orange for glycine (G); Rose for histidine (H); Yellow for proline (P); and Turquoise for tyrosine (Y). Single line represents the two knuckles, while double line showed the transmembrane region.

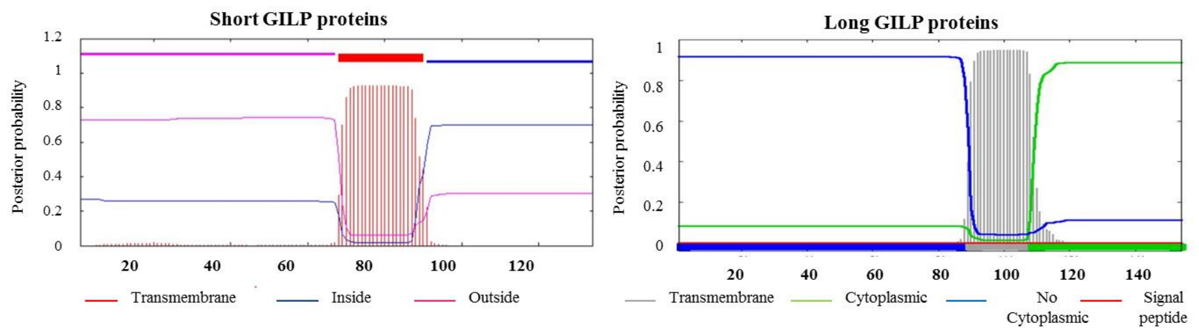


Figure 4: Analysis of the complete sequence of short and long GILP proteins using TMHMM tool. Amino acid region outside, inside and anchored in plasma membrane are showed by different colors.

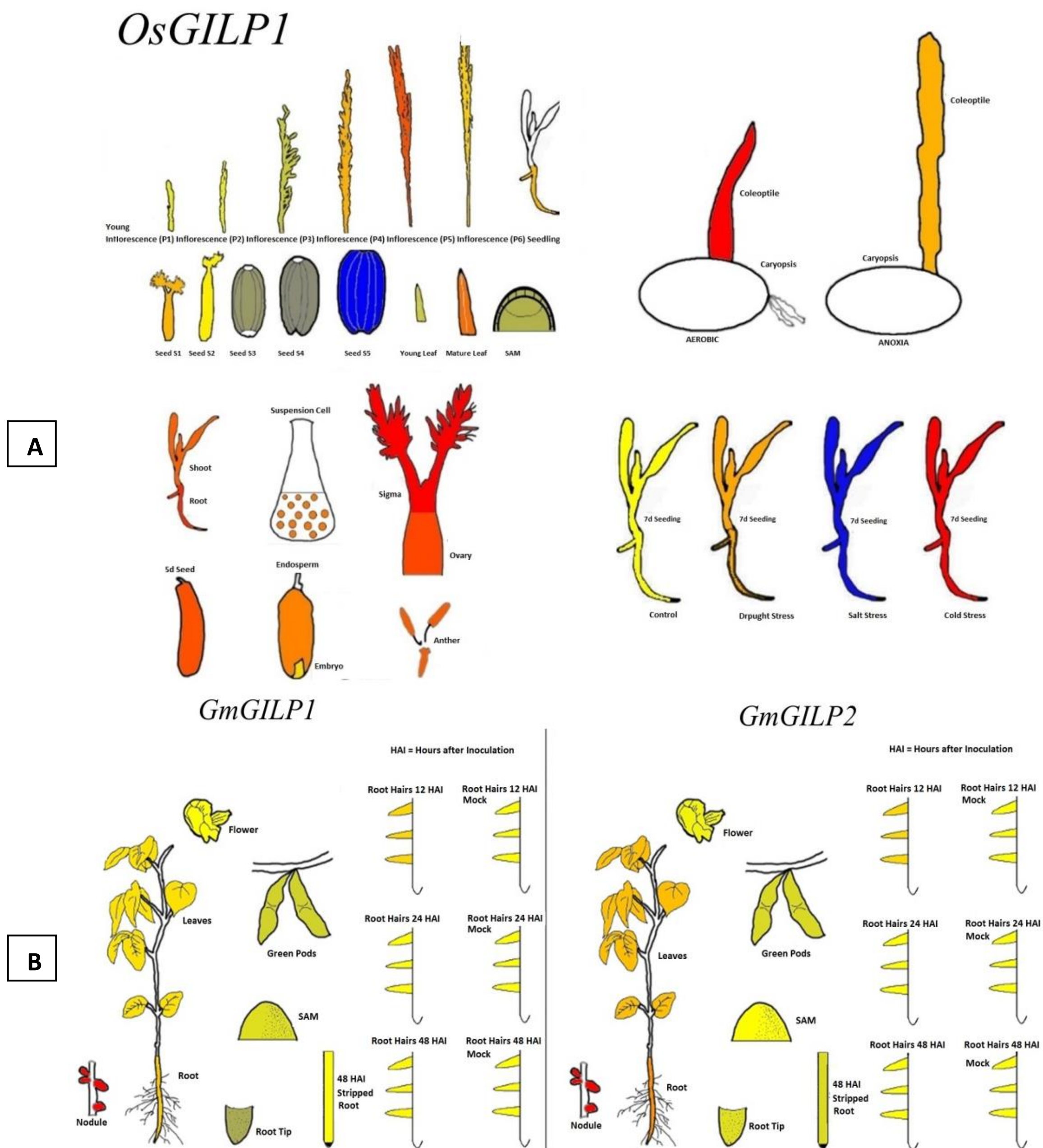


Figure 5: Expression profile of *GILP* genes conducted in the RNA-seq BAR database. Red color indicates up- and blue indicates downregulated genes (Log 2 ratio) under the described conditions, as plotted in the bar scale. A) *OsGILP1* expression during development in different tissues/organs. *OsGILP1* expression was also monitored in seedlings under drought, salt and cold stresses. B) *GmGILP1 and GmGILP2* expression different organs/tissues and in roots inoculated with *Bradyrhizobium japonicum*. HAI: Hours After Inoculation with soil bacteria.

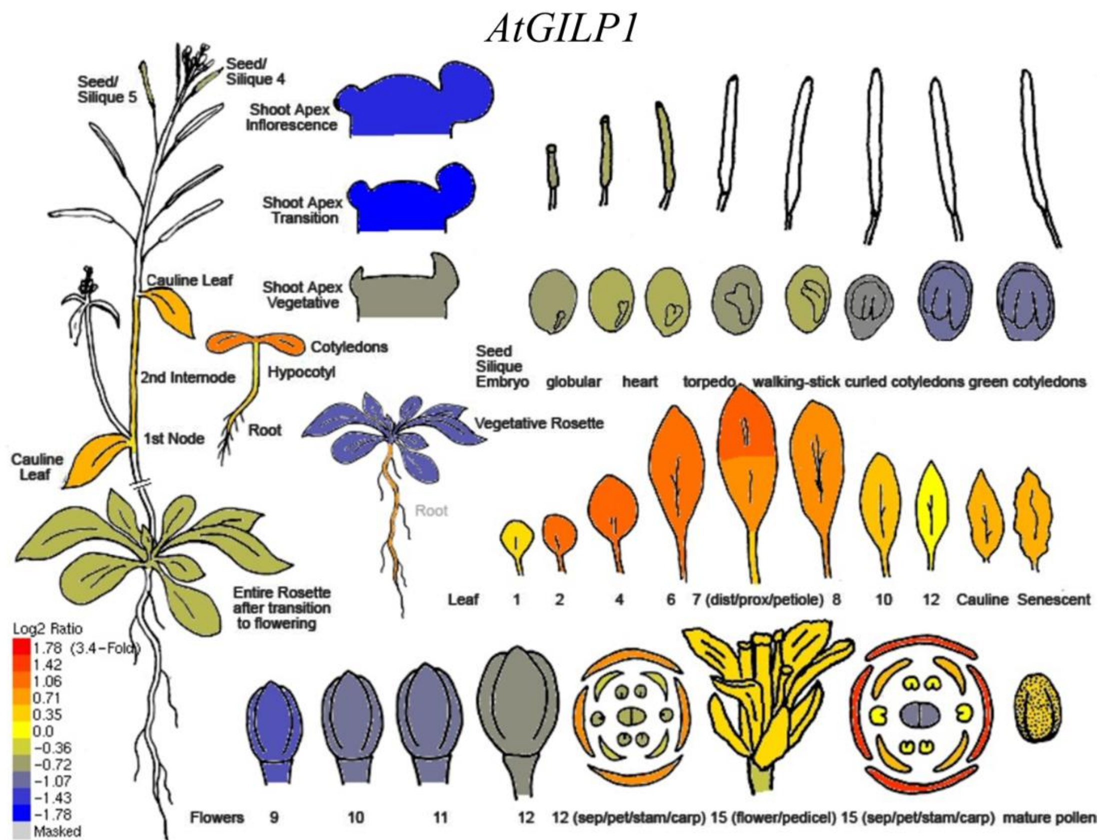


Figure 6: Expression profile of *AtGILP1* gene during plant development analyzed in the RNA-seq BAR database. Red color indicates up- and blue indicates downregulated genes (Log 2 ratio) under the described conditions, as plotted in the bar scale.

AtGILP1

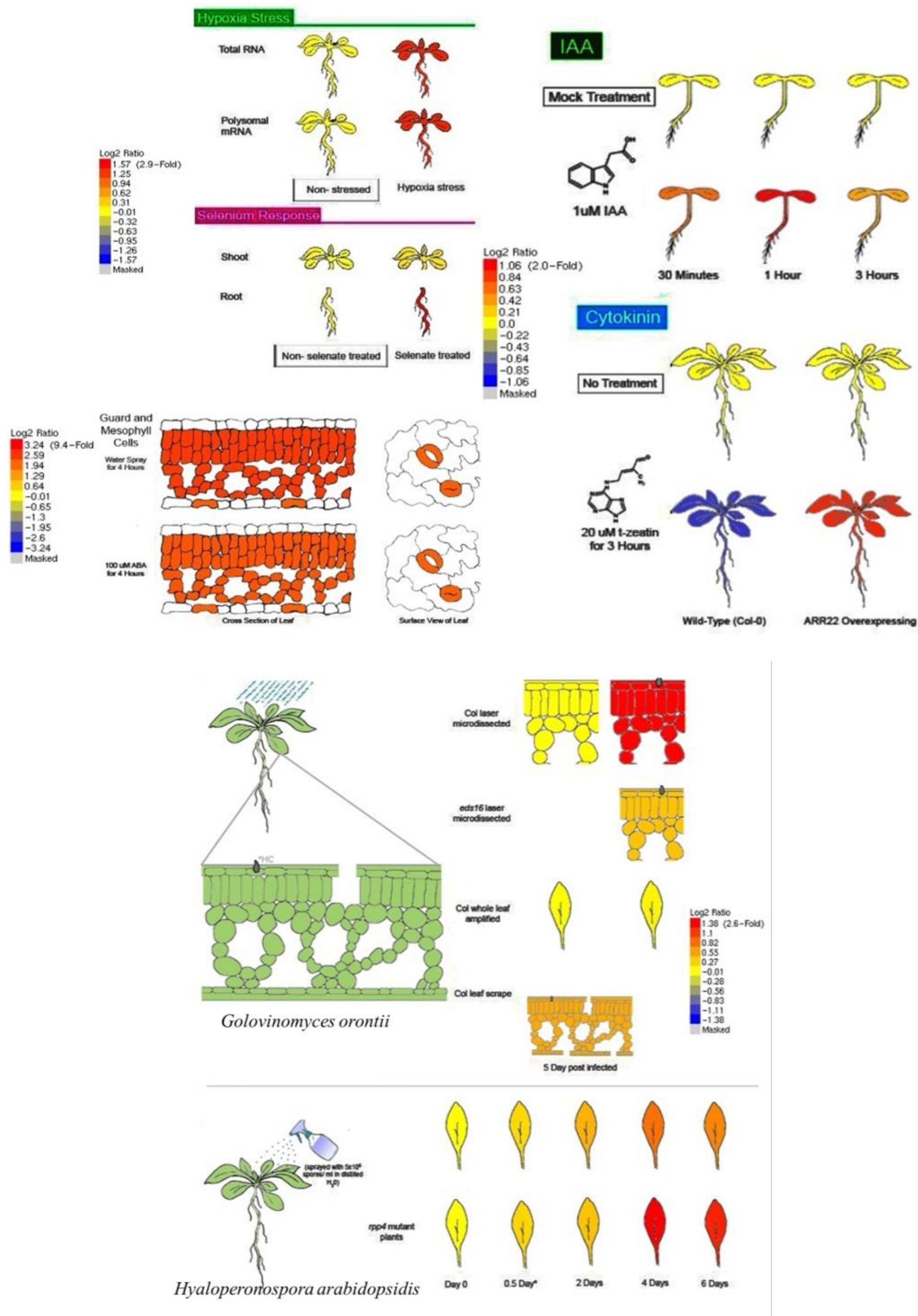


Figure 7: Expression profile of *AtGILP1* gene under biotic/abiotic stresses and under hormone treatment. Analyzes were performed in analyzed in the RNA-seq BAR database. Red color indicates up- and blue indicates downregulated genes (Log 2 ratio) under the described conditions, as plotted in the bar scale.

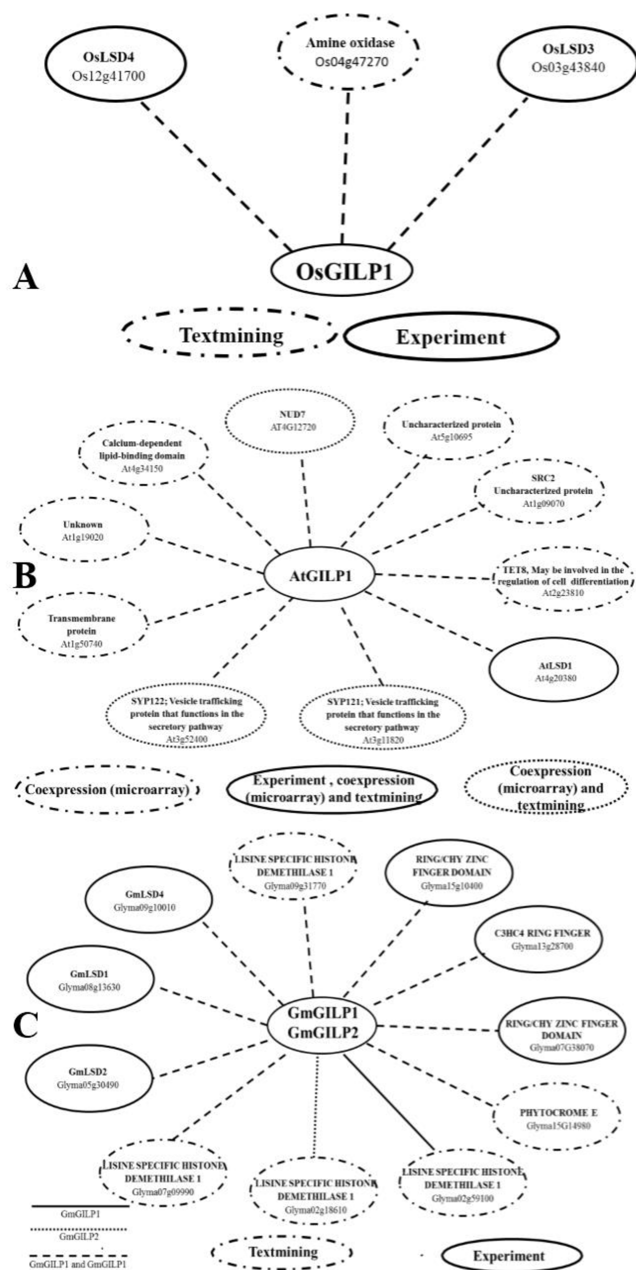


Figure 8: A) OsGILP1 protein-protein interaction network. Dotted circle indicates text-mining evidence, while continues circle shows experimental evidence. B)AtGILP1 interaction network. Dotted spaced circle indicates coexpression evidence, through microarray. Continues circle exhibits experimental, coexpression and text-mining evidence. Dotted circle shows coexpression and text-mining evidence. C)GmGILP1 and GmGILP2 protein-protein interaction network. Dotted circle indicates text-mining evidence, while closed circle shows experimental evidence. Continues line shows protein-protein interaction of GmGILP1, while dotted line indicates protein-

protein interaction of GmGILP2. Dotted spaced line represents the proteins of network which are common of GmGILP1 and GmGILP2 proteins.

GROUP	SPECIES	LOCUS	ACRONYM
EUDICOT	*Manihot esculenta*	Manes.08G103200	MsGILP1
		Manes.09G185800	MsGILP2
	Ricinus communis	30128.t000503	RcGILP1
	Salix purpurea	SapurV1A.1785s0030	SpuGILP1
		SapurV1A.1433s0060	SpuGILP2
	Linum usitatissimum	Lus10001805	LuGILP1
	Populus trichocarpa	Potri.001G061900	PtGILP1
		Potri.003G165900	PtGILP2
	Medicago truncatula	Medtr7g005440	MtGILP1
	Phaseolus vulgaris	Phvul.008G000500	PvuGILP1
	Glycine max	Glyma.18G301900	GmGILP1
		Glyma.08G360100	GmGILP2
	Cucumis sativus	Cucsa.34930	CsGILP1
	Prunus persica	Prupe.1G219100	PpGILP1
	Malus domestica	MDP0000365927	MdGILP1
		MDP0000273830	MdGILP2
	Arabidopsis thaliana	AT5G13190	AtGILP1
	Arabidopsis lyrata	488134	AlGILP1
	Arabidopsis halleri	Araha.8686s0002	AhGILP1
	Boechera stricta	Bostr.2902s0054	BstGILP1
	Capsella grandiflora	Cagra.0596s0011	CgGILP1
	Capsella rubella	Carubv10002238m	CrGILP1
	Eutrema salsugineum	Thhalv10014995m.g	EsGILP1
	Brassica rapa Chiifu-FPsc	Brara.B0047	BrGILP1
		Brara.J02111	BrGILP2
	Gossypium raimondii	Gorai.008G156500	GrGILP1
		Gorai.005G195900	GrGILP2
	Theobroma cacao	Thecc1EG008435	TcGILP1
	Citrus sinensis	orange1.1g032751m	CsiGILP1
	Eucalyptus grandis	Eucgr.B0354	EgGILP1
	Vitis vinífera	GSVIVT01000954001	VvGILP1
	Kalanchoe marnieriana	Kalax0070s0053	KmGILP1
		Kalax0133s0054	KmGILP2
	Solanum lycopersicum	Solyc01g009780	SlGILP1
	Mimulus guttatus	Migut.K0103	MgGILP1
	Aquilegia coeruela	Aquca_021_00021	AcGILP1
	Amaranthus hypochondria	AHYPO_004823	AhyGILP1
		AHYPO_000339	AhyGILP2
MONOCOT	*Sorghum bicolor*	Sobic.004G157800	SbGILP1
	Zea mayz	GRMZM2G010235	ZmGILP1
	Setaria italica	Seita.1G172800	SiGILP1
	Setaria viridis	Sevir.1G175600	SvGILP1
	Panicum virgatum	Pavir.Ab01504	PvGILP1
	Panicum hallii	Pahal.0269s0034	PhGILP1
	Oryza sativa	LOC_Os02g31100	OsGILP1
	Brachypodium	Bradi3g4432	BdGILP1
	Brachypodium stacei	Brast04G188700	BsGILP1
	Zostera marina	Zosma7g01360	ZmaGILP1
	Triticum aestivum	Traes_2DS_FD31C0778	TaGILP1
		Traes_2BS_7C849A3EB	TaGILP2
		Traes_2AS_4133C82FB	TaGILP3
	Ananas comosus	Aco017825	AcoGILP1
		Aco003618	AcoGILP2
LYCOPHYTA	*Selaginella moelledorffii*	121042	SmGILP1
	Amborella trichopoda	evm_27.model.AmTr	AtrGILP1
	Musa acuminata	GSMUA_Achr9G01170	MaGILP1
	Spirodela polyrhiza	Spipo0G0074500	SpGILP1
BRYOPHYTA	*Physicomitrella patens*	Pp3c26_1170	PpaGILP1

ESM 1. Detailed data on gene acronym and locus of *GILP* genes.

Gene	Subcellular localization prediction			Transmembrane domain	
	TargetP				
	Loc	Reliability	Cello	TMHMM	Phobius
MsGILP1	O	2	Plasma Membrane	x	(70-88)
MsGILP2	O	2	Plasma Membrane	x	(69-88)
RcGILP1	O	2	Plasma Membrane	x	(69-88)
SpuGILP1	O	3	Extracellular	(68-90)	(69-88)
SpuGILP2	O	3	Extracellular	(68-90)	(69-88)
LuGILP1	O	2	Plasma Membrane	x	(69-88)
PtGILP1	O	2	Plasma Membrane	x	(69-88)
PtGILP2	O	3	Plasma Membrane	(68-90)	(69-88)
MtGILP1	O	2	Plasma Membrane	(65-87)	(65-85)
PvuGILP1	O	1	Plasma Membrane	(69-91)	(69-89)
GmGILP1	O	2	Plasma Membrane	(73-95)	(74-92)
GmGILP2	O	2	Plasma Membrane	(73-95)	(74-92)
CsGILP1	O	2	Plasma Membrane	(68-90)	(68-98)
PpGILP1	O	2	Plasma Membrane	x	(69-88)
MdGILP1	O	4	Plasma Membrane	x	(97-116)
MdGILP2	O	4	Plasma Membrane	x	x
AtGILP1	O	3	Plasma Membrane	(68-90)	(69-88)
AlGILP1	O	3	Plasma Membrane	(68-90)	(69-88)
AhGILP1	O	3	Extracellular	(68-90)	(69-88)
CrGILP1	O	3	Plasma Membrane	(68-90)	(69-88)
BrGILP1	O	2	Plasma Membrane	(68-90)	(69-88)
BrGILP2	O	2	Plasma Membrane	x	(69-88)
BstGILP1	O	3	Extracellular	(69-91)	(69-88)
CgGILP1	O	3	Extracellular	(68-90)	(69-88)
EsGILP1	O	3	Extracellular	(69-91)	(69-88)
ThGILP1	O	3	Plasma Membrane	(69-91)	(69-88)
GrGILP1	O	2	Plasma Membrane	(68-90)	(69-88)
GrGILP2	O	3	Plasma Membrane	(70-92)	(71-90)
TcGILP1	O	3	Plasma Membrane	(68-90)	(69-88)
CsiGILP1	O	2	Plasma Membrane	x	(69-88)
EgGILP1	O	2	Plasma Membrane	x	(69-88)
VvGILP1	O	3	Plasma Membrane	(64-86)	(65-84)
KmGILP1	O	1	Extracellular	x	(72-90)
KmGILP2	O	2	Extracellular	x	(72-90)
SlGILP1	O	3	Plasma Membrane	x	(89-107)
MgGILP1	O	2	Plasma Membrane	(88-110)	(89-107)
AcGILP1	O	2	Plasma Membrane	(104-126)	(104-123)
AhyGILP1	O	2	Extracellular	x	(90-109)
AhyGILP2	O	1	Extracellular	(90-112)	(90-110)
SbGILP1	O	2	Plasma Membrane	x	(89-107)
ZmGILP1	O	1	Plasma Membrane	x	(90-108)
SiGILP1	O	2	Plasma Membrane	x	(89-107)
SvGILP1	O	2	Plasma Membrane	x	(89-107)
PvGILP1	O	1	Plasma Membrane	x	(90-108)
PhGILP1	O	4	Extracellular	x	x
OsGILP1	O	1	Plasma Membrane	x	(90-108)
BdGILP1	O	2	Plasma Membrane	x	(86-104)
BsGILP1	O	2	Extracellular	x	(86-104)
ZmaGILP1	O	3	Extracellular	x	(79-98)
TaGILP1	C	4	Extracellular	(22-44)	(24-42)
TaGILP2	S	2	Extracellular	(15-37)	(15-34)
TaGILP3	S	2	Extracellular	(15-37)	x
AcoGILP1	O	2	Plasma Membrane	x	(102-121)
AcoGILP2	O	5	Plasma Membrane	(45-67)	(46-65)
SmGILP1	O	4	Plasma Membrane	x	x
AtrGILP1	O	2	Extracellular	x	(83-102)
MaGILP1	O	3	Extracellular	(45-67)	(46-65)
SpGILP1	O	5	Extracellular	(67-89)	(67-87)
PpaGILP1	O	2	Plasma Membrane	x	x

ESM 2. Prediction of subcellular localization of the proteins encoded by *GILP* genes. The analysis of the existence of signal peptide was performed using the tools SignalP and TargetP. Phobius and TMHMM tools were employed to predict the existence of transmembrane domain.

DISCUSSÃO GERAL E PERSPECTIVAS

Todos os organismos vivos precisam lidar com vários fatores desencadeantes de dano ao longo de suas vidas, o que acarreta a necessidade do emprego de uma ampla variedade de estratégias defensivas para responder e superar a nocividade causada pela ocorrência de estresses. Para garantir respostas adequadas e apropriadas, as plantas utilizam sistemas de detecção de dano altamente sensíveis, bem como mecanismos de transdução de sinal robustos e interligados (Petrov *et al.,* 2015). Além disso, diferentes medidas de proteção a estresses são desencadeadas de uma forma altamente específica e rigorosamente controlada, de modo a garantir a ocorrência das reações adequadas sem desperdício excessivo de recursos energéticos. Após atingir certo limiar de alterações bioquímicas e moleculares, as células vegetais não podem mais manter processos metabólicos usuais e então a Morte Celular Programada (PCD) é induzida (Wituszyńska and Karpiński, 2013). A ativação de respostas múltiplas, envolvendo complexas interações gênicas e *crosstalk* com muitas vias moleculares é uma característica marcante na adaptação da planta a diferentes danos (Duque *et al.,* 2013). Embora a morte de células seja desfavorável à produção de biomassa, a indução de PCD em tecidos submetidos a estresses favorece o organismo como um todo. Ela permite a adaptação e sobrevivência da planta por meio da transdução de sinais de células afetadas para células sadias, levando à tolerância/resistência a estresses e aclimatação a condições adversas (Wituszyńska and Karpiński, 2013).

Em animais, a cascata de eventos e moléculas que regulam a PCD já foi bem descrita, enquanto que em plantas o mecanismo de PCD ainda permanece pouco compreendido (Wituszyńska and Karpiński, 2013; Minina *et al.,* 2014). Muitos estudos e descobertas em plantas são resultado da comparação direta com o mecanismo de PCD em animais e oriundos de estudos de genômica comparativa. Em animais, a caracterização de genes relacionados à indução ou repressão do mecanismo de apoptose vem sendo intensificada, devido à significância desse processo no estabelecimento de patologias, como diferentes tipos de câncer. Nesse sentido, a utilização da genômica comparativa para investigação da função gênica em animais e plantas pode esclarecer diversos mecanismos e processos conservados entre esses diferentes reinos.

A presente tese foi embasada no estudo de algumas famílias gênicas relacionadas a respostas a estresses e/ou constituintes do morteossomo, importante rede de interação e auto-regulação, fundamental para o mecanismo de PCD. A rede de genes pertencentes ao morteossomo apresenta papel fundamental para a ativação e posterior controle da extensão de morte celular, de forma que alguns genes apresentam papéis opostos, como *At*MC1 (regulador positivo de PCD) e *At*LSD1 (regulador negativo de PCD), necessitando atuar em conjunto (Coll *et al,.* 2010).

Em nosso grupo, a linha de pesquisa sobre PCD teve início com os estudos investigativos da família gênica *LSD*. Como desdobramento, a escolha das famílias gênicas a serem exploradas nessa tese levou em consideração seu envolvimento direto com o gene *At*LSD1, que ocupa papel central no morteossomo. Nesse sentido, além do estudo de algumas famílias gênicas reportadas como integrantes do morteossomo (*LSD*, *Metacaspase*, *NF-Y* e *PLAC8*), a caracterização da família a qual pertence *At*GILP (que também interage diretamente com *At*LSD1) foi igualmente conduzida.

A caracterização da família *LSD* demonstrou que esta é exclusiva de Viridiplantae. Um total de 113 genes foram identificados e a diversificação da família antecede a separação entre monocotiledôneas e dicotiledônias. A expansão da família parece ter ocorrido no clado embriófitas, que apresentam múltiplas copias gênicas, ao contrário das espécies basais, que apresentam apenas um gene *LSD*. A análise das sequências preditas de aminoácidos para os genes *LSD* permitiram a identificação de proteínas com um, dois e três domínios dedo de zinco LSD. Nossos resultados permitiram a proposição de um modelo de evolução para a família, no qual um ancestral comum daria origem a proteínas LSD com um, dois e três domínios. Proteínas com um e dois domínios podem não ter sido duplicadas, ou as cópias duplicadas teriam sido perdidas ao longo da evolução. Já as proteínas que possuem três domínios parecem ter seguido dois cenários: duplicação em tandem ou em bloco seguida por divergência funcional ou ausência de duplicação. A família LSD é a mais estudada entre as famílias pertencentes ao morteossomo. Trabalhos vêm mostrando seu envolvimento em PCD condicionada por estresse abiótico, como frio (Huang *et al.*, 2010), hipóxia (Muhlenbock *et al.*, 2007) e excesso de luz (Muhlenbock *et al.*, 2008), além do envolvimento na PCD ocasionada por estresse biótico (Cabreira *et al.,* 2013). O gene *At*LSD1 é o mais bem caracterizado, embora outros genes igualmente presentes no morteossomo, como *At*LOL1 e *At*LOL2 também mereçam atenção especial. Outra questão em aberto é a caracterização completa de proteínas que possuem um e dois domínios LSD, uma vez que apenas proteínas com três domínios vêm sendo exploradas.

No que se refere à caracterização da família Metacaspase, cisteína-proteases descritas em plantas (Vercammen *et al.,* 2004), fungos (Madeo *et al.,* 2002) e protistas (Lee *et al.*, 2002; Lee *et al.*, 2007), um total de 372 genes foram identificados em Viridiplantae. Assim como para a família LSD, diferentes tipos de proteínas foram identificadas e compreendem dois grupos distintos. Proteínas do tipo I apresentam o domínio metacaspase e podem apresentar ou não o domínio dedo de zinco LSD (também encontrado nessas proteínas) e uma região rica em prolina. Já as proteínas do tipo II não apresentam o domínio dedo de zinco LSD e a região rica em prolina, mas possuem uma região de ligação (*linker)* entre os domínios catalíticos p10 e p20. A condição sem domínio LSD parece ser a

ancestral, de forma que foi encontrada nos organismos basais. Tais resultados foram corroborados pela análise filogenética. Em plantas, poucos genes metacaspase foram caracterizados até o momento. Esses incluem genes de tomate (Hoeberichts *et al.,* 2003), trigo (Wang et al., 2012), *Nicotiana benthamiana* (Hao *et al*., 2007) *e Capsicum annuum* (Kim *et al.,* 2013). Em *Arabidopsis* existem nove genes metacaspase e apenas dois (*At*MC1 e *At*MC2) foram caracterizados até o momento. Nesse sentido, nosso trabalho representou uma fonte importante para investigação de genes metacaspase em diferentes espécies. Devido à importância da função do gene *At*MC1 durante PCD (Coll *et al.,* 2010), a investigação de outros genes metacaspase pode esclarecer o envolvimento dessa família no mecanismo de PCD, bem como permitir a investigação e expansão de tais resultados para espécies de importância econômica.

Com relação à família gênica NF-Y, a ampla caracterização dessa família vem sendo feita em diversas espécies, como *Brachypodium distachyon* (Cao *et al.* 2011), painço (Feng *et al.,* 2015), feijão (Rípodas *et al.,* 2015), tomate (Li *et al*., 2016), sorgo (Malviya et al., 2016), uva (Ren *et al.*, 2016), *Arabidopsis* (Zhao, *et al.,* 2016) e arroz (Yang *et al.* 2017). Sabe-se que essa família é ausente em algas e seu surgimento foi na espécie *Selaginella moellendorffii* (Cagliari et al., 2011). Em soja, a caracterização completa da família da família NF-Y não havia sido realizada. Por esse motivo, nosso trabalho objetivou a identificação e descrição dos genes NF-Y de soja. Foram identificados 68 genes, sendo 21 pertencentes à subunidade A (NF-YA), 23 à subunidade B (NF-YB) e 24 à subunidade C (NF-YC). A análise dos mecanismos de duplicação e evolução mostrou a ocorrência de duplicação em bloco e a atuação do processo de subfuncionalização na manutenção da família. Análises *in silico* dos genes NF-Y mostraram a modulação da expressão gênica em condições de estresses, especialmente durante a infecção por *Phakopsora pachyrhizi.* Desta forma, nossos resultados corroboram a proposição da família como constituinte do morteossomo, de modo que a expressão gênica de muitos genes NF-Y é modulada durante condições que disparam PCD. Trabalhos futuros, contemplando a validação experimental do papel dessa família no morteossomo poderão esclarecer o envolvimento da família nesse mecanismo, sendo que os genes podem estar relacionados à regulação positiva ou negativa de PCD.

A investigação da família de proteínas ricas em cisteinas PLAC8 permitiu a identificação de 445 genes, que apresentavam nomes distintos na literatura, sugerindo o não agrupamento em uma mesma família gênica. Nosso trabalho propôs uma nomenclatura unificada para os genes que possuem o domínio PLAC8. As variações encontradas no domínio PLAC8 permitiram a identificação de grupos de genes, que agruparam separadamente na análise filogenética, corroborando a existência de três

diferentes tipos de genes PLAC8. Os genes do tipo I são encontrados em mamíferos, fungos, plantas e algas, enquanto os genes do tipo II e III são exclusivos de plantas. Até o momento, apenas genes do tipo I e II foram investigados e são relacionados a diferentes mecanismos em plantas, que vão desde a determinação do tamanho de órgãos (Frary *et al.,* 2000; Cong *et al.* 2002; Cong *et al.,* 2006; Libault *et al*., 2010; Guo *et al.,* 2010; Liu *et al*., 2015), controle do fluxo de cálcio (Nakagawa *et al.,* 2007; Yamanaka *et al*., 2010; Kurusu *et al*., 2012) e resistência a elementos químicos (Song *et al.* 2004; Song *et al*., 2010). Tais resultados sugerem que os genes PLAC8 podem estar relacionados ao mecanismo de PCD em plantas, sendo no controle do desenvolvimento ou na resposta a estresses em plantas. Dessa forma, esses resultados corroboram a proposição de membros da família PLAC8 como componentes do morteossomo. Entretanto, muitas questões ainda permanecem em aberto. O papel exato dos genes PLAC8 no mecanismo de PCD ainda é desconhecido. Ainda, considerando o grande número de cópias de genes PLAC8, funções antagonistas podem ser encontradas para as diferentes cópias. De qualquer forma, como nosso trabalho foi pioneiro em identificar genes PLAC8 em Viridiplantae, o mesmo representa uma fonte importante de dados para serem profundamente investigados futuramente, especialmente, no que se refere aos genes do tipo III, nunca antes descritos.

A última família gênica estudada em nosso trabalho foi a classe que reúne proteínas de membrana que possuem domínio LITAF. Extensivamente estudadas em humanos, apenas uma proteína contendo o domínio LITAF havia sido descrita em *Arabidopsis* e o gene que a codifica nomeado como gene *At*GILP (*GSH-induced LITAF domain protei*n) (He et al. 2011). A interação física de *At*LSD1 e *At*GILP (He *et al.* 2011) motivou a investigação dessa família. Nosso trabalho objetivou uma ampla busca por genes GILP em Viridiplantae. Como resultado, foi observado que a família é ausente em algas e seu surgimento foi na espécie *Selaginella moellendorffii*. Foi identificado um total de 59 genes e o número de cópias variou de uma a três. Análises *in silico* mostraram a modulação da expressão gênica durante condições de estresse. Tais resultados reforçam a proposição dessa família como participante do morteossomo. A família GILP é a menos estudada entre todas as famílias exploradas nessa tese. Tendo em vista a existência de apenas um trabalho em plantas (He *et al*., 2011), muitas questões ainda permanecem desconhecidas e merecem ser exploradas.

Considerando todas as famílias aqui investigadas, nossos resultados mostraram a expansão das famílias em Viridiplantae, o que determinou a existência de múltiplas cópias gênicas. Além disso, a existência de diferentes grupos de proteínas em cada família foi uma característica comumente observada, sendo que muitos grupos foram descritos pela primeira vez neste estudo. Embora muitos

genes nessas famílias venham sendo investigados, a sua grande maioria ainda permanece sem caracterização funcional. Assim, embora a investigação dessas famílias tenha sido motivada pela presença no morteossomo, o papel dessas famílias em rotas diferentes e mecanismos distintos não podem ser descartados. Muitas questões permanecem em aberto em todas as famílias aqui apresentadas e nossos resultados representam apenas o ponto inicial para a compreensão da função dessas famílias em plantas. Seja como proteínas âncoras na membrana, (transdutoras de sinal), como fatores de transcrição ou atuando em redes de interação proteína-proteína, a investigação do papel dessas famílias na sinalização de estresses representa um importante desafio na compreensão do mecanismo de PCD em plantas.

Sabendo que *AtLSD1* e *AtLOL1* atuam de forma antagonista no controle da PCD (Epple *et al.*, 2003), o mesmo pode ocorrer entre membros das outras famílias gênicas, de forma que a maioria não se apresenta em cópia única nos genomas estudados. Uma vez que muitas espécies possuem mais de um gene de cada família, a manutenção de todas as cópias com a mesma função biológica pode não ser vantajosa. Além disso, as diferentes cópias podem interagir entre si, formando complexos a fim de controlar positiva ou negativamente a PCD. Embora interações entre genes *LSD, Metacaspase, PLAC8* e *GILP* 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 descritos na sinalização de PCD. A obtenção de plantas transgênicas ou mutantes 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. Estas análises certamente serão úteis para um melhor entendimento do papel desempenhado pelas famílias pertencentes ao morteossomo em vários processos biológicos, especialmente no controle de PCD.

As estratégias aplicadas durante o desenvolvimento dessa tese permitiram ampliar meus conhecimentos e delinear minha experiência na área de bioinformática, identificação e caracterização de genes/famílias gênicas. Durante o primeiro ano dessa tese, esforços significativos foram empregados no estudo da família de fatores de transcrição WRKY, resultando em um artigo científico, no qual divido a primeira autoria (Bencke-Malato *et al.*, 2014). Esse estudo não foi incluído nessa tese, já que não faz parte da linha de pesquisa com PCD. No entanto, contribuiu enormemente para minha formação, já que meu envolvimento incluiu a mesma linha de pesquisa na qual baseia-se essa tese (identificação e análise filogenética).

REFERENCIAS BIBLIOGRÁFICAS – INTRODUÇÃO, DISCUSSÃO GERAL E PERSPECTIVAS

Abuqamar S, Luo H, Laluk K, Mickelbart MV, Megiste T (2009) Crosstalk between biotic and abiotic stress responses in tomato is mediated by the AIM1 transcription factor. Plant J 58:p 347-360

Ahmad P, Latef AAHA, Rasool S, Akram NA, Ashraf M, Gucel S (2016) Role of Proteomics in Crop Stress Tolerance Front Plant Sci: 7-1336

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species Plant Molecular Biology Reporter 9:208-218

Ashraf M, Ahmad MSA, Öztürk M, Aksoy (2012) A Crop Improvement Through Different Means: Challenges and Prospects. Crop Production for Agricultural Improvement, Chapter 1:1-15

Atkinson NJ, Urwin PE (2012) The interaction of plant biotic and abiotic stresses: from genes to the field J Exp Bot 63:3523-3543

Atkinson NJ, Lilley CL, Urwin PE (2013) Identification of genes involved in the response of *Arabidopsis* to simultaneous biotic and abiotic stresses. Plant Physiol 162:2028-2041

Bencke-Malato M, Cabreira C, Wiebke-Strohm B, Bücker-Neto L, Mancini E, Osorio MB, Homrich MS, Turchetto-Zolet AC, De Carvalho MCCG, Stolf R, Weber RLM, Westergaard G, Castagnaro AP, Abdelnoor RV, Marcelino-Guimarães FC, Margis-Pinheiro M, Bodanese-Zanettini MH (2014) Genome-wide annotation of the soybean WRKY family and functional characterization of genes involved in response to *Phakopsora pachyrhizi* infection. BMC Plant Biol 14: 236

Bolger ME, Arsova B, Usadel B (2017) Plant genome and transcriptome annotations: from misconceptions to simple solutions. Briefings in Bioinformatics: 1–13

Bostock RM, Pye MF, Roubtsova TV (2014) Predisposition in plant disease: exploiting the nexus in abiotic and biotic stress perception and response. Annu Rev Phytopathol 52: 517-549

Bray EA, Bailey-Serres J, Weretilnyk E (2000) Responses to abiotic stresses, in Biochemistry and Molecular Biology of Plants, American Society of Plant Physiologists: 1158–1249

Cabreira C, Cagliari A, Bücker-Neto L, Wiebke-Strohm B, De Freitas L B, Marcelino-Guimarães FC, et al (2013) The Lesion Simulating Disease (LSD) gene family as a variable in soybean response to *Phakopsora pachyrhizi* infection and dehydration Funct Integr Genomics:13:323–38

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 1:1-26

Cao S, Kumimoto RW, Siriwardana CL, Risinger JR, Holt BF 3rd (2011) Identification and characterization of NF-Y transcription factor families in the monocot model plant *Brachypodium distachyon*. PLoS One 6: e21805

Coll NS, Vercammen D, Smidler A, Clover C, Breusegem FVan, Dangl J L et al (2010) Arabidopsis type I metacaspases control cell death Science 330:1393–7

Coll NS, Epple P, Dangl JL (2011) Programmed cell death in the plant immune system. Cell Death Differ 18(8):1247–56.

Cong B, Tanksley SD (2006) FW22 and cell cycle control in developing tomato fruit: A possible example of gene co-option in the evolution of a novel organ. Plant Mol Biol 62:867–80

Cong B, Liu J, Tanksley SD (2002) Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. Proc Natl Acad Sci U S A, 99:13606–11

Del Duca S, Serafini-Fracassini D, Cai G (2014) Senescence and programmed cell death in plants: polyamine action mediated by transglutaminase. Front Plant Sci 5: 120

Delatorre CA, Silva AA (2008) *Arabidopsis thaliana*: Uma pequena planta um Grande. Papel Revista de Ciências Agrárias,58-67

Dhanapal AP, Govindaraj M (2015) Unlimited Thirst for Genome Sequencing, Data Interpretation, and Database Usage in Genomic Era: The Road towards Fast-Track Crop Plant Improvement. Genet Res Int, 684321

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

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

Dita MA, Rispail N, Prats E, Rubiales D, Singh KB (2006) Biotechnology approaches to overcome biotic and abiotic stress constraints in legumes. Euphytica, 147: 1-24

Duque AS, Almeida AM, De Silva AB, Da Silva JM, Farinha AP, Santos D, et al (2013) Abiotic stress responses in plants: unraveling the complexity of genes and networks to survive. Abiotic Stress Plant Responses and Applications in Agriculture, Rijeka: InTech 3–23:105772/52779

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 U S A, 100:6831-6836

Feng ZJ, He GH, Zheng, WJ, Lu PP, Chen M, Gong YM, Ma YZ, Xu ZS (2015) Foxtail Millet NF-Y Families: Genome-Wide Survey and Evolution Analyses Identified Two Functional Genes Important in Abiotic Stresses. Front Plant Sci 6: 1142

Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, et al (2000) Fw22: a Quantitative Trait Locus Key To the Evolution of Tomato Fruit Size. Science, 289:85–8

Gepts P, Beavis WD, Brummer EC, Shoemaker RC, Stalker HT, Weeden NF, Young ND (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

Guo M, Rupe Ma, Dieter JA, Zou J, Spielbauer D, Duncan KE, et al (2010) Cell Number Regulator1 affects plant and organ size in maize: implications for crop yield enhancement and heterosis. Plant Cell, 22:1057–73

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

Hao L, Goodwin PH, Hsiang T (2007) Expression of a metacaspase gene of *Nicotiana benthamiana* after inoculation with *Colletotrichum destructivum* or *Pseudomonas syringae* pv tomato, and the effect of silencing the gene on the host response. Plant Cell Rep, 26:1879-88

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

Hoeberichts FA, ten Have A, Woltering EJ (2003) A tomato metacaspase gene is upregulated during programmed cell death in *Botrytis cinerea*-infected leaves. Planta, 217:517-522

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

Kaminaka H, Nake C, Epple P, Dittgen J, Schutze K, Chaban C, Holt BF 3rd, 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

Kim SM, Bae C, Oh SK, Choi D (2013) A pepper (*Capsicum annuum* L) metacaspase 9 (Camc9) plays a role in pathogen-induced cell death in plants. Mol Plant Pathol, 14:557-66

Kissoudis C, van de Wiel C, Visser RGF, Van Der Linden G (2014) Enhancing crop resilience to combined abiotic and biotic stress through the dissection of physiological and molecular crosstalk. Front Plant Sci:5

Lan TH, DelMonte TA, Reischmann KP, Hyman J, Kowalski S, McFerson J, Kresovich S, Paterson AH (2000) An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. Genome Res, 10:776–788

Lee N, Bertholet S, Debrabant A, Muller J, Duncan R, Nakhasi HL (2002) Programmed cell death in the unicellular protozoan parasite *Leishmania*. Cell Death Differ, 9:53–64

Lee N, Gannavaram S, Selvapandiyan A, Debrabant A (2007) Characterization of metacaspases with trypsin-like activity and their putative role in programmed cell death in the protozoan parasite *Leishmania*. Eukaryot Cell, 6:1745–1757

Li S, Li K, Ju Z, Cao D, Fu D, Zhu H, Zhu B, Luo Y (2016) Genome-wide analysis of tomato NF-Y factors and their role in fruit ripening. BMC Genomics, 17:36

Libault M, Stacey G (2010) Evolution of FW22-like (FWL) and PLAC8 genes in eukaryotes. Plant Signal Behav, 5:1226–8

Liu Z, Cheng Q, Sun Y, Dai H, Song G, Guo Z, et al (2015) A SNP in OsMCA1 responding for a plant architecture defect by deactivation of bioactive GA in rice. Plant Mol Biol, 87:17–30

Madeo F, Herker E, Maldener, C, Wissing S, Lächelt S, Herlan M, Fehr M, Lauber K, Sigrist SJ, Wesselborg S, Fröhlich KU (2002) A caspase-related protease regulates apoptosis in yeast. Mol Cell, 9:911–917

Malviya N, Jaiswal P, Yadav Y (2016) Genome- wide characterization of Nuclear Factor Y (NF-Y) gene family of sorghum [*Sorghum bicolor* (L) Moench]: a bioinformatics approach. Physiol Mol Biol Plants, 22:33–49

Mengiste T, Chen X, Salmeron J, Dietrich R (2003)The BOTRYTIS SUSCEPTIBLE1 gene encodes an R2R3MYB transcription factor protein that is required for biotic and abiotic stress responses in *Arabidopsis* Plant Cell, 15:2551–2565

Michaeli S, Galili G, Genschik P, Fernie AR, Avin-Wittenberg T (2016) Autophagy in plants–what's new on the menu? Trends Plant Sci, 21:134–144

Minina EA, Bozhkov P V and Hofius D (2014)Autophagy as initiator or executioner of cell death Trends Plant Sci 19, 692-697

Minina, E A,Filonova LH, Sanchez-Vera V, Suarez MF, Daniel G, Bozhko PV (2013) Detection and measurement of necrosis in plants. Methods Mol Biol, 1004:229-48

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

Mochida K, Shinozaki K (2011) Advances in Omics and Bioinformatics Tools for Systems Analyses of Plant Functions. Plant Cell Physiol, 52:2017–2038

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

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

Mur LA, Kenton P, Lloyd AJ, Ougham, He, Prats E (2008) The hypersensitive response, the centenary is upon us but how much do we know? J Exp Bot, 59:501-520

Paterson A H, Bowers JE, Burow MD, Draye X, Elsik CG, Jiang CX, Katsar CS, Lan TH, Lin, YR, Ming R, Wright RJ (2000) Comparative Genomics of Plant Chromosomes. Plant Cell, 12:1523–1540

Petrov V, Hille J, Mueller-Roeber B, Gechev TS (2015) ROS-mediated abiotic stress-induced programmed cell death in plants. Front Plant Sci, 6: 69

Prasch CM, Sonnewald U (2013) Simultaneous application of heat, drought, and virus to *Arabidopsis* plants reveals significant shifts in signaling networks. Plant Physiol, 162:1849–1866

Prasch CM, Sonnewald U (2014) Signaling events in plants: stress factors in combination change the picture. Environ Exp Bot, 114:4-14

Rasmussen S, Barah P, Suarez-Rodriguez MC, Bressendorff S, Friis P, Costantino P et al (2013) Transcriptome responses to combinations of stresses in *Arabidopsis*. Plant Physiol, 161:1783–1794

Reape TJ, Molony EM, McCabe PF (2008) Programmed cell death in plants: distinguishing between different modes. J Exp Bot, 59:435-44

Ren C, Wang Y, Li S, Liang Z (2016) Genome-wide identification and characterization of the NF-Y gene family in grape (*Vitis vinifera* L). BMC Genomics, 17:605

Ripodas C, Castaingts M, Clua J, Blanco F, Zanetti ME (2015) Annotation, phylogeny and expression analysis of the nuclear factor Y gene families in common bean (*Phaseolus vulgaris*). Front Plant Sci, 5:761

Rivero RM, Mestre TC, Mittler RON, Rubio F, Garcia-Sanchez F, Martinez V (2013) The combined effect of salinity and heat reveals a specific physiological, biochemical and molecular response in tomato plants. Plant Cell Environ, 37:1059–1073

Rockström J, Falkenmark M (2000) Semiarid crop production from a hydrological perspective: gap between potential and actual yields. CRC, 19:319–346

Rusterucci C, Aviv DH, Holt BF 3rd, Dangl JL, Parker JE (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

Shao HB, Chu LY, Shao MA, Jaleel CA, Hong-Mei M (2008) Higher plant antioxidants and redox signaling under environmental stresses. Comp Rend Biol, 331:433–441

Song WY, Choi KS, Kim DY, Geisler M, Park J, Vincenzetti V et al (2010) *Arabidopsis* PCR2 is a zinc exporter involved in both zinc extrusion and long-distance zinc transport. Plant Cell, 22:2237–52

Song WY, Martinoia E, Lee J, Kim D, Kim DY, Vogt E, et al (2004) A novel family of cys-rich membrane proteins mediates cadmium resistance in *Arabidopsis*. Plant Physiol, 135:1027–39

Su Z, Yang Z, Xu Y, Chen Y, Yu Q (2015) Apoptosis, autophagy, necroptosis, and cancer metastasis. Mol Cancer, 14: 48

Suzuki N, Rivero RM, Shulaev V, Blumwald E, Mittler R (2014) Abiotic and biotic stress combinations. New Phytol, 203:32–43

Suzuki N, Rizhsky L, Liang H, Shuman J, Shulaev V, Mittler R (2005) Enhanced tolerance to environmental stress in transgenic plants expressing the transcriptional coactivator multiprotein bridging factor 1c. Plant Physiol, 139:1313–1322

van Doorn WG, Beers EP, Dangl JL, Franklin-Tong V E, Gallois P, Hara-Nishimura I, et al (2011) Morphological classification of plant cell deaths. Cell Death Differ, 18:1241–1246

Vercammen D, van de Cotte B, De Jaeger G, Eeckhout D, Casteels P, Vandepoele K, Vandenberghe I, Van Beeumen J, Inzé D, Van Breusegem F (2004) Type II metacaspases Atmc4 and Atmc9 of *Arabidopsis thaliana* cleave substrates after arginine and lysine. J Biol Chem, 279:45329-36

Wang XD, Wang XJ, Feng H, Tang CL, Bai PF, Wei GR, Huang LL, Kang ZS (2012) TaMCA4, a novel wheat metacaspase gene functions in programmed cell death induced by the fungal pathogen *Puccinia striiformis* sp Tritici. Molecular Plant-Microbe Interactions, 25:755–764

Wituszyńska W, Karpiński S (2013) Programmed Cell Death as a Response to High Light, UV and Drought Stress in Plants. Abiotic Stress - Plant Responses and Applications in Agriculture, InTech, DOI: 105772/53127

Yang W, Lu Z, Xiong, Y, Yao J (2017) Genome-wide identification and co-expression network analysis of the *OsNF-Y* gene family in rice. The Crop Journal 5 (1), 21–31

Zhao C, Zhang Z, Xie S, Si 33 T, Li Y, Zh JK (2016) Mutational Evidence for the Critical Role of CBF Transcription Factors in Cold Acclimation in *Arabidopsis*. Plant Physiol, 171:2744-2759

Zurbriggen MD, Carrillo N, Hajirezaei MR (2010) ROS signaling in the hypersensitive response When, where and what for? Plant Signal Behav, 5:393–396