

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

**Diversidade de fitocistatinas em arroz e suas proteases cisteínicas alvo, com enfoque
em fitocistatinas carboxi-estendidas e inibição de legumainas**

Ana Paula Christoff

Porto Alegre, 14 setembro de 2015

Universidade Federal do Rio Grande do Sul
Instituto de Biociências
Departamento de Genética
Programa de Pós Graduação em Genética e Biologia Molecular

Diversidade de fitocistatinas em arroz e suas proteases cisteínicas alvo, com enfoque em fitocistatinas carboxi-estendidas e inibição de legumainas

Ana Paula Christoff

Orientador: Prof. Dr. Rogério Margis

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do título de Doutor em Genética e Biologia molecular

Porto Alegre, 14 setembro de 2015

INSTITUIÇÃO E FONTES FINANCIADORAS

Este trabalho foi desenvolvido no Laboratório de Genomas e Populações de Plantas do Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul e financiando pelas instituições listadas abaixo:

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico,
Ministério da Ciência e Tecnologia.

ÍNDICE

| | |
|--|--------------------------------------|
| RESUMO..... | 5 |
| ABSTRACT | 6 |
| Capítulo I - Introdução | 7 |
| 1. Inibidores de proteases cisteínicas | 8 |
| 1.1. Cistatinas | 8 |
| 1.2. Fitocistatinas | 9 |
| 2. Inibição de proteases cisteínicas por cistatinas e fitocistatinas..... | 12 |
| 2.1. Proteases do tipo Papaínas em plantas..... | 14 |
| 2.2. Proteases do tipo legumáinas em plantas..... | 15 |
| 3. O modelo vegetal..... | 17 |
| 4. Fitocistatinas em arroz (<i>Oryza sativa</i>) | 18 |
| Capítulo II - Objetivos..... | 21 |
| Capítulo III – The diversity of rice phytocystatins..... | Erro! Indicador não definido. |
| Capítulo IV – Uncovering legumain genes in rice | Erro! Indicador não definido. |
| Capítulo V – Rice bifunctional phytocystatin is a dual modulator of legumain and papain-like proteases..... | Erro! Indicador não definido. |
| Capítulo VI – Discussão | 23 |
| Capítulo VII - Perspectivas..... | 39 |
| Capítulo VIII - Bibliografia..... | 42 |
| Anexo I – Artigo em colaboração – The mitochondrial glutathione peroxidase GPX3 is essential for H ₂ O ₂ homeostasis and root and shoot development in rice..... | 52 |
| Anexo II – Artigo em colaboração – Identification of potential miRNAs and their targets in <i>Vrisea carinata</i> (Poales, Bromeliaceae)..... | 54 |
| Anexo III – Artigo em colaboração – isomiRID: a framework to identify microRNA isoforms ... | 56 |
| Anexo IV – Artigo em colaboração – The wall-associated kinase gene family in rice genomes.... | 58 |
| Anexo V – Artigo em colaboração – Comprehensive selection of reference genes for quantitative gene expression analysis during seed development in <i>Brasica napus</i> ... | 60 |

RESUMO

Fitocistatinas são inibidores competitivos de proteases cisteínicas em plantas que atuam principalmente na inibição de papaínas. Entretanto, há uma demonstração *in vitro* de uma fitocistatina carboxi-estendida com capacidade de inibir também uma protease do tipo legumaína. Em arroz existem 12 genes de fitocistatinas, mas apenas um deles (OcXII), possui a extensão C-terminal. Desta forma, o objetivo geral deste trabalho é compreender a diversidade e os mecanismos de interação entre as fitocistatinas e suas proteases cisteínicas alvo em arroz, com ênfase nas implicações funcionais da OcXII. Neste trabalho, nós demonstramos que, em arroz, os 12 genes de fitocistatinas têm um perfil de expressão gênica diferenciado na germinação e em respostas ambientais, sendo OcI, OcIII e OcXII os genes mais expressos. O peptídeo recombinante, correspondente à extremidade C-terminal da OcXII possui capacidade inibitória de legumaínas e não afeta a atividade das papaínas. Através do silenciamento transcricional da OcXII, via RNAi, foram obtidas plantas com atividades proteolíticas aumentadas tanto de papaínas quanto legumaínas, em adição a um fenótipo de crescimento inicial acelerado. O fenótipo oposto é observado em plântulas crescendo em condições alcalinas. Plantas expressando o promotor OcXII fusionado ao gene repórter *gus* demonstraram um perfil de ativação de OcXII durante a germinação, principalmente na região do escutelo das sementes. Esta ativação de OcXII permanece alta quando as sementes são mantidas com níveis elevados de ABA e sob estresse alcalino. Como alvos específicos de OcXII, as legumaínas estão relacionadas com a biossíntese de componentes vacuolares e degradação de proteínas de armazenamento. Em arroz, nós encontramos cinco diferentes *loci* para as legumaínas. Análises filogenéticas e estudos de expressão gênica demonstraram uma maior associação de OsaLeg2 e OsaLeg3 com tecidos de semente, e OsaLeg1, 4 e 5 com tecidos vegetativos. Também foram observadas formas de *splicing* alternativo, com potencial de originar diferentes isoformas ativas de legumaínas. Em geral, nossos dados demonstram o envolvimento bifuncional da OcXII nos processos de germinação e defesa interagindo e inibindo a atividade de proteases cisteínicas específicas, onde a região N-term de OcXII inibe papaínas, enquanto a C-term inibe legumaínas.

ABSTRACT

Phytocystatins are competitive inhibitors of cysteine proteases in plants, principally inhibiting papain-like proteases activity. However, there is one in vitro demonstration that a carboxy-extended phytocystatin can inhibit legumain-like proteases either. In rice there are 12 phytocystatins genes, but only one (OcXII) has the C-terminal extension. Thus, the aim of this study is to understand the diversity and interaction mechanisms among phytocystatins and its cysteine protease targets in rice, focusing on OcXII functional implications. In this work we demonstrated that in rice, the 12 phytocystatins genes have a different gene expression profile during the germination and environmental responses, while OcI OcIII and OcXII were the most expressed genes. A recombinant peptide corresponding to the different C-terminus of OcXII was produced and its inhibitory capacity was confirmed against legumains without affecting the activity of papains. OcXII transcriptional silencing, via RNAi, resulted in plants with increased proteolytic activity for both papain and legumains, in addition to an accelerated initial growth phenotype. The opposite phenotype is observed in seedlings growing under alkaline conditions. Plants expressing the OcXII promoter fused to *gus* reporter gene demonstrated an OcXII activation profile during germination, especially in the seed scutellum region. This OcXII activation remains high when the seeds are kept at high ABA levels or under alkaline conditions. As specific OcXII targets, legumains are related to the biosynthesis of vacuolar storage components and protein degradation. In rice, we found 5 different *loci* for legumains. Phylogenetic analyses and gene expression studies demonstrated a greater association of OsaLeg3 and OsaLeg2 with seed tissues, while OsaLeg1, 4 and 5 were more abundant in vegetative tissues. Also alternatively spliced forms were observed, with the potential to produce different isoforms of active legumains. Overall, our data demonstrate the involvement of the bifunctional OcXII in germination processes and defense, interacting and inhibiting the activity of specific cysteine proteases, where its N-term region inhibits papain, while the C-term inhibits legumains.

Capítulo I
- Introdução -

1. Inibidores de proteases cisteínicas

Existem diversos genes que codificam para proteases nos mais variados tipos de organismos conhecidos, os quais são equipados com uma grande maquinaria proteolítica, a fim de regular reversível ou irreversivelmente o destino das proteínas em suas células. As proteases são extremamente necessárias para os organismos, elas clivam ligações peptídicas, degradam proteínas em aminoácidos, removem proteínas não funcionais e participam na regulação de vias importantes em respostas ambientais, e de desenvolvimento. Isto faz com que haja uma alta especificidade e uma fina regulação da atividade da protease sobre seu substrato. A presença de uma cisteína (Cys), como nucleófilo, ativada por uma histidina (His), no sítio catalítico, caracteriza as proteases cisteínicas (van der Hoorn, 2008). A atividade destas proteases é eficientemente regulada através da inibição de seu sítio catalítico pela interação com inibidores de proteases cisteínicas. Naturalmente, os organismos possuem inibidores de proteases cisteínicas endógenos, produzidos de acordo com sua necessidade fisiológica. Tais inibidores foram caracterizados e denominados cistatinas (Barrett, 1981), sendo posteriormente incluídos no banco de dados MEROPS, no clã IH e na família I25.

1.1 Cistatinas

As cistatinas são inibidores competitivos de proteases cisteínicas que, interagindo com o sítio ativo das enzimas-alvo, bloqueiam o acesso ao substrato. Embora sejam consideradas inibidores competitivos, a afinidade por algumas proteases-alvo é extremamente elevada, tornando difícil a reversão da inibição (Barrett, 1987). Primeiramente isoladas e caracterizadas em animais (em clara de ovo de galinha (Fossum and Whitaker, 1968)), a superfamília das cistatinas está envolvida na proteção de células e tecidos contra a atividade proteolítica de peptidases lisossomais liberadas em condições normais de morte celular, ou ainda em situações de proliferação de células cancerígenas ou parasitas e patógenos (Abrahamson *et al.*, 2003).

A identificação de homólogos de cistatinas em outros metazoários levou-as a uma subdivisão em três famílias principais, de acordo com a homologia entre suas sequências, presença de pontes dissulfeto e a massa molecular da proteína. As cistatinas de tipo 1, ou stefinas, possuem uma massa molecular de aproximadamente 11kDa, com apenas um domínio inibitório e sem pontes dissulfeto; as de tipo 2, cistatinas propriamente ditas,

também apresentam um domínio inibitório único, porém possuem massa molecular de ~15kDa com quatro resíduos cisteínicos formando pontes dissulfeto; ainda as cistatinas de tipo 3, ou kininogênicos, possuem domínios inibitórios múltiplos, possivelmente resultantes de eventos de duplicação (Turk and Bode, 1991; Abrahamson *et al.*, 2003; Benchabane *et al.*, 2010), que originam proteínas de altas massas moleculares (60-120kDa). Estes inibidores de proteases caracterizam-se especificamente pela presença de pelo menos um domínio cistatina, o qual é formado por três motivos proteicos conservados: (1) um resíduo de glicina (G) na região amino-terminal; (2) o motivo QxVxG; e (3) um triptofano (W) na região carboxi-terminal, sendo que, estes dois últimos motivos estruturais localizam-se no primeiro e segundo *loops* entre as folhas- β da estrutura terciária. Estes motivos proteicos são os principais responsáveis pela atividade inibitória das cistatinas, pois eles interagem diretamente com o sítio ativo das enzimas-alvo, as proteases cisteínicas (Bode *et al.*, 1988) (Fig.1). Além disso, estes domínios são altamente conservados em termos de sequência e estrutura uma vez que, as cistatinas são proteínas bastante antigas em termos evolutivos, podendo ser encontradas em diversos organismos.

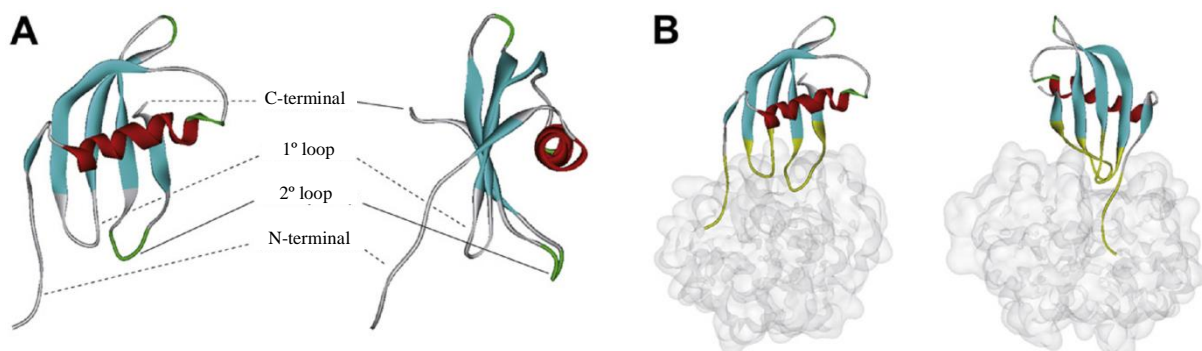


Figura 1. Modelo estrutural de uma cistatina e sua interação com uma protease-alvo. **A)** visão frontal e lateral demonstrando as cinco folhas- β , a α -hélice e os dois *loops* inibitórios. **B)** visão da cistatina interagindo com a protease cisteínica modelo, a papaína. Reproduzido de Benchabane *et al.*, 2010.

1.2 Fitocistatinas

Alguns anos após a identificação de cistatinas em animais verificou-se que estes inibidores de proteases cisteínicas também estavam presentes em plantas, através do isolamento e caracterização de uma cistatina em arroz, a oryzacistatina (Abe *et al.*, 1987a). Desta forma abriram-se precedentes para a busca destes inibidores de proteases nas mais diversas espécies de plantas e organismos distantemente relacionados, aumentando expressivamente o número de proteínas conhecidas. Com o aumento deste conhecimento,

pôde-se verificar que as cistatinas de plantas apresentavam uma peculiaridade, elas compartilhavam uma sequência característica não encontrada nas demais cistatinas de metazoários. Assim, os inibidores de proteases cisteínicas em plantas foram denominados fitocistatinas (PhyCys), diferindo das cistatinas animais pela presença do motivo conservado LARFAVDEHN em sua α -hélice amino-terminal, formando um cluster evolutivo distinto e monofilético, com a característica única de que este motivo específico encontra-se sempre à uma distância de 17 aminoácidos do domínio QxVxG (Margis *et al.*, 1998) (Fig.2).

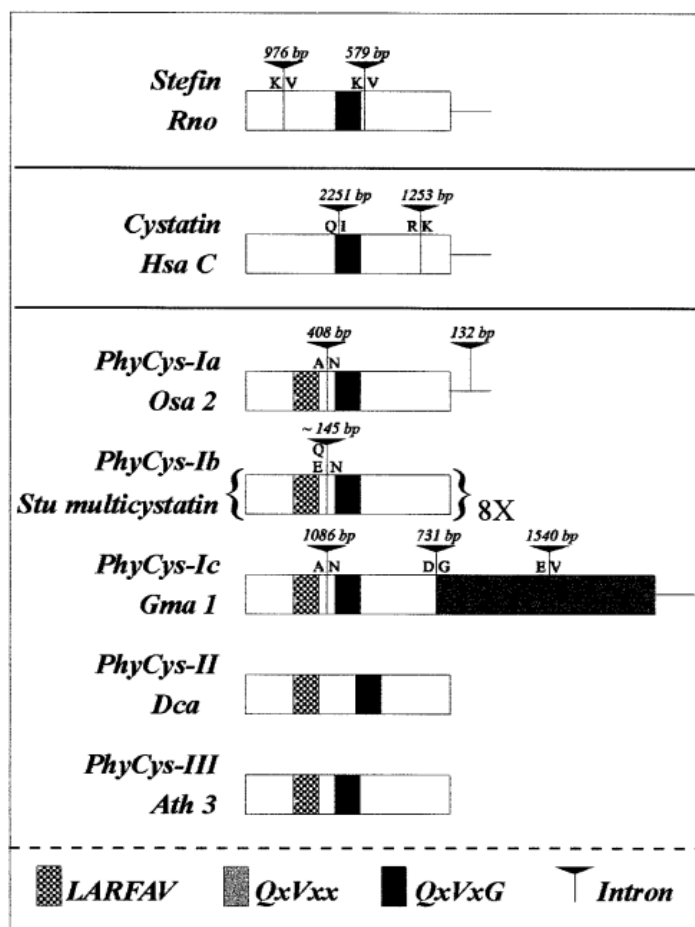


Figura 2. Organização genética das stefinas, cistatinas e fitocistatinas. A estrutura de aminoácidos está representada pelos boxes retangulares, com os motivos proteicos conservados identificados pelos retângulos hachurados, de acordo com a legenda especificada. Os aminoácidos cujos códons são separados por introns estão indicados acima dos retângulos, com seus respectivos tamanhos (bp). Ao lado esquerdo encontram-se identificadas as famílias de cistatinas (stefinas, cistatinas ou fitocistatinas (PhyCys)), assim como a espécie de origem da proteína: Rno (*Rattus norvegicus*), Hsa (*Homo sapiens*), Osa (*Oryza sativa*), Stu (*Solanum tuberosum*), Gma (*Glycine max*), Dca (*Daucus carota*) e Ath (*Arabidopsis thaliana*). Figura retirada de Margis *et al.* 1998.

As fitocistatinas são desprovidas de sítios putativos de glicosilação, de resíduos cisteínicos, e de pontes dissulfeto. Entretanto, as pontes dissulfeto podem ser

satisfatoriamente substituídas por outros tipos de interações moleculares que mantenham a estrutura funcional da proteína, como interações hidrofóbicas de resíduos aromáticos conservados que ocorrem entre as folhas antiparalelas β 4- β 5 e a conformação do *loop*2 (Margis *et al.*, 1998). Apesar destas restrições particulares às fitocistatinas, grande parte das características moleculares das cistatinas e fitocistatinas manteve-se através do processo evolutivo dos eucariotos, uma vez que estes inibidores de proteases cisteínicas estão amplamente distribuídos entre animais e plantas. Apesar das divergências entre as sequências de aminoácidos, as cistatinas de animais e as fitocistatinas possuem uma estrutura terciária conservada, formada por cinco folhas- β antiparalelas ao redor de uma α -hélice central (Turk and Bode, 1991). A conservação da estrutura terciária, assim como dos motivos QxVxG e W nas estruturas de *loop* do inibidor, tanto em cistatinas quanto em fitocistatinas, sugere um mecanismo inibitório conservado entre animais e plantas (Benchabane *et al.*, 2010).

As fitocistatinas podem ser subdivididas em três grupos distintos, de acordo com seu peso molecular e estrutura evolutiva. A maioria é incluída no grupo I, possuindo aproximadamente 12-16kDa e apenas um domínio cistatina, com aproximadamente 100 resíduos de aminoácidos. No grupo II, encontram-se as fitocistatinas com aproximadamente 23kDa e além do domínio cistatina característico na região amino-terminal, também possuem uma extensão carboxi-terminal com um domínio similar a cistatina, tendo assim ~200 resíduos de aminoácidos. O terceiro grupo, contém as fitocistatinas de ~80kDa, conhecidas como multicistatinas, uma vez que possuem diversos domínios cistatinas repetidos (Margis-Pinheiro *et al.*, 2008; Wang *et al.*, 2008; Chu *et al.*, 2011).

Desta forma, as fitocistatinas são os inibidores de proteases cisteínicas de plantas, equivalentes as cistatinas de animais. Grande parte de suas funções já são conhecidas, entretanto ainda há muito para ser descoberto e compreendido funcionalmente. Em plantas, os mecanismos mais estudados envolvem o controle de vários processos fisiológicos e celulares de proteólise endógena até a inibição de proteases cisteínicas exógenas, secretadas por parasitas e patógenos (Grudkowska and Zagdańska, 2004; Dutt *et al.*, 2010; Kunert *et al.*, 2015). Entretanto, ainda existem inúmeras lacunas sobre quais as proteases e qual a eficiência inibitória que as fitocistatinas exercem sobre elas em diferentes situações.

2. Inibição de proteases cisteínicas por cistatinas e fitocistatinas

Evolutivamente conservadas, cistatinas e fitocistatinas são inibidores de proteases cisteínicas (Kordis and Turk, 2009). Os alvos primários destes inibidores são proteases da família C1A, composta principalmente por proteases do tipo papaínas (papaínas, catepsinas, quimopapaína, bromelaína, ficaína, entre outras) (Bode *et al.*, 1988; Arai *et al.*, 1991; Martinez *et al.*, 2009). A inibição destas proteases resulta da interação forte e reversível com as cistatinas, através do contato direto de três partes do inibidor (os dois loops inibitórios e o segmento amino-terminal) (Fig1), conforme descrito no primeiro modelo de interação de cistatinas com suas proteases-alvo (Bode *et al.*, 1988). As enzimas do tipo papaínas tem uma estrutura tridimensional muito similar entre si, com dois lobos conectados em forma de “V”, compostos por α -hélices e folhas- β formando um sulco onde se localiza o sítio ativo da protease, contendo os resíduos catalíticos Cys-His (Turk and Turk, 2001). Os dois loops inibitórios das cistatinas interagem competindo e se ligando diretamente a este sítio ativo nas enzimas do tipo papaínas (Turk *et al.*, 1998). Este tipo de inibição tem sido bastante estudada, existindo uma infinidade de trabalhos sobre este mecanismo inibitório e efeitos em diferentes organismos (Barrett, 1981; Abe *et al.*, 1988; Machleidt *et al.*, 1989; Arai *et al.*, 1991, 2002; Björk *et al.*, 1994; Ohtsubo *et al.*, 2005; Hong *et al.*, 2008; Wang *et al.*, 2008; Hwang *et al.*, 2009; Martinez *et al.*, 2009; Martínez *et al.*, 2012; Diaz-Mendoza *et al.*, 2014).

Adicionalmente, demonstrou-se mais recentemente que algumas cistatinas animais também conseguiam inibir proteases de uma família bastante diferente das papaínas (família C1A), as legumaínas que pertencem à família C13. Estas duas famílias de proteases apresentam arranjos diferenciados em seus sítios catalíticos, além de motivos diferentes formando os sítios ativos, com estruturas tridimensionais divergentes. Alternativamente às proteases do tipo papaínas, as legumaínas estão mais relacionadas estruturalmente e funcionalmente com enzimas relacionadas às caspases, que são proteases aspartato-específicas (Chen *et al.*, 1998). Outra evidência sugere um mecanismo inibitório diferente para as legumaínas, uma vez que elas não são inibidas pelo maior inibidor das enzimas do tipo papaínas, o E-64 (Chen *et al.*, 1997), suportando a teoria de que a topologia do sítio ativo das legumaínas é diferente da observada em papaínas. Assim, em 1991, um primeiro estudo demonstrou que a cistatina C de humanos era capaz de inibir

legumaínas da mesma forma que ela inibe proteases da família C1, com alta afinidade em uma ligação competitiva e reversível (Alvarez-Fernandez *et al.*, 1999). Entretanto, o mecanismo de inibição de legumaínas por cistatinas é bastante diferente da inibição de papaínas, sendo que, para que as legumaínas sejam inibidas, é necessária a interação com o resíduo de asparagina (Asn) localizado no segmento de loop que conecta a α -hélice principal da estrutura da cistatina com a primeira folha- β (Fig3). Desta forma, constatou-se que existiam algumas cistatinas com capacidade bifuncional de atuação, inibindo proteases do tipo papaínas e também proteases do tipo legumaínas, através de dois sítios inibitórios independentes da cistatina (Alvarez-Fernandez *et al.*, 1999). Em algumas situações pode até ocorrer uma inibição simultânea dos dois tipos de proteases e formando um complexo ternário papaína-cistatina-legumaína.

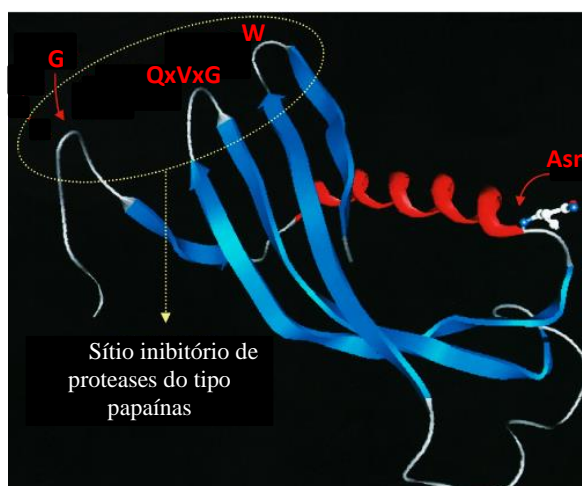


Figura 3. Representação da estrutura da cistatina C animal, com seus domínios inibitórios. A estrutura tridimensional é baseada na estrutura descrita para a cistatina da clara de ovo de galinha. A região da cistatina que constitui o sítio inibitório responsável pela inibição das proteases do tipo papaínas está identificada pela região circutada, com os motivos G, QxVxG e W. Próximo à α -hélice da estrutura estão identificados os átomos que compõe o resíduo de asparagina (Asn), responsável pela inibição de proteases do tipo legumaínas. Figura adaptada de (Alvarez-Fernandez *et al.*, 1999).

Assim como as cistatinas, algumas fitocistatinas são diferenciadas dentro da mesma família inibidores de proteases cisteínicas. As plantas possuem cópias variadas de genes de fitocistatinas, mas na maioria das espécies de Viridiplantae analisadas até o momento, pode-se encontrar pelo menos uma cópia de fitocistatina que possua uma extensão carboxi-terminal, com o motivo SNSL conservado (Martinez and Diaz, 2008). Em 2007, Martinez *et al.*, descreveram a similaridade deste domínio SNSL com àquele domínio da cistatina

humana, importante para a inibição de proteases do tipo legumainas em plantas. De forma bastante conservada com as cistatinas bifuncionais, a extensão carboxi-terminal das fitocistatinas possui o resíduo de asparagina (Asn) na região do loop logo após a α -hélice principal da estrutura terciária, e é este resíduo de asparagina que confere a capacidade inibitória de legumainas à extensão C-terminal das fitocistatinas. Isto tudo foi demonstrado *in vitro*, onde a fitocistatina carboxi-estendida de cevada foi capaz de inibir a atividade da legumaina humana, além de inibir a atividade de proteases tipo legumainas em extratos proteicos de cevada (Martinez *et al.*, 2007).

Os estudos realizados até o momento serviram para estabelecer a funcionalidade das cistatinas e fitocistatinas como inibidores de proteases cisteínicas do tipo papaínas, levando a um amplo conhecimento sobre estas interações. Entretanto, os estudos sobre a capacidade inibitória de legumainas ainda são escassos, limitando-se a poucos estudos *in vitro* (Martinez *et al.*, 2007; Santos-Silva *et al.*, 2012). Tendo em vista a importância das proteases cisteínicas de ambas as famílias C1A e C13 para os organismos, aqui com interesse específico em plantas, torna-se importante a ampliação do conhecimento sobre as fitocistatinas como inibidores bifuncionais endógenos destes dois tipos de proteases cisteínicas.

2.1 Proteases do tipo Papaínas em plantas

O número de proteases cisteínicas da família C1A em plantas é bastante elevado, estando em quantidades mais reduzidas apenas em algas e outras plantas mais basais (Martinez and Diaz, 2008). As atividades desempenhadas pelas proteases do tipo papaínas estão relacionadas a diversos processos fisiológicos importantes nas plantas vasculares (Beers *et al.*, 2004), destacando-se sua atuação sobre proteínas intracelulares, processando, modificando e remobilizando proteínas de armazenamento durante a germinação das sementes (Watanabe *et al.*, 1991; Lepelley *et al.*, 2012), durante a senescência ou morte celular programada (Beers *et al.*, 2000; Hayashi *et al.*, 2001; Wan *et al.*, 2002; Diaz-Mendoza *et al.*, 2014), ou ainda, em situações de estresse (Khanna-Chopra *et al.*, 1999).

Na defesa contra patógenos e insetos, as proteases C1A também são importantes, atuando contra as proteínas por eles produzidas e secretadas (Chen *et al.*, 1992; van der Hoorn and Jones, 2004; Dutt *et al.*, 2010). Organismos que se alimentam de plantas podem facilmente ser intoxicados pelo látex que é secretado pelas plantas, uma vez que esse látex é composto por diversas proteases do tipo papaínas (Konno *et al.*, 2003). As papaínas de plantas foram classificadas em nove subfamílias (RD21A, CEP1, XCP2, XBCP3, THI1, SAG12, RD19A, AALP, CTB3) e análises bioquímicas revelaram que estas proteínas existem em diferentes isoformas ativas, em pH variados e com sensibilidade alterada para diferentes inibidores (Richau *et al.*, 2012). Mais especificamente, as proteases da família C1A de plantas estão proximamente relacionadas às catepsinas de mamíferos que pertencem aos grupos B, F, H e L (Martinez and Diaz, 2008), possuindo mecanismos de ação e estruturas tridimensionais semelhantes (Kamphuis *et al.*, 1985). Desde a descoberta destas proteases, até os dias atuais, os avanços em bioquímica e biologia molecular proporcionaram a caracterização das mais diversificadas proteases cisteínicas entre a família C1A, com distribuições específicas entre tecidos ou características únicas. Entretanto, o elevado número destas proteases sugere que possa existir sobreposição de funções, com possível redundância funcional, sendo que ainda há muito a ser explorado sobre as proteínas-alvo destas enzimas.

2.2 Proteases do tipo legumainas em plantas

As legumainas são proteases cisteínicas que pertencem à família C13 (EC 3.4.22.34, MEROPS), e estão presentes tanto em plantas quanto em animais. Previamente descritas em plantas, estas proteases também foram denominadas de *Vacuolar Processing Enzymes* (VPEs), sendo sintetizadas como pro-peptídeos, que por sua vez, são co-transcricionalmente direcionados para o lúmen do retículo endoplasmático rugoso, onde, na forma de uma pro-legumaina inativa é transportada para o vacúolo ou parede celular (Hara-Nishimura *et al.*, 1991; Müntz and Shutov, 2002). Ao alcançar estes compartimentos celulares, sob condições de pH ácido, a legumaina é auto-cataliticamente ativada (Hiraiwa *et al.*, 1999; Fischer *et al.*, 2000; Kuroyanagi *et al.*, 2002). Legumainas são também conhecidas como *asparaginil endopeptidases*, devido a sua habilidade de reconhecer e clivar ligações peptídicas contendo resíduos de asparagina (Asn), ou ácido aspártico (Asp) (Abe *et al.*, 1993; Ishii, 1994; Becker *et al.*, 1995; Hiraiwa *et al.*, 1999). Tais

especificidades catalíticas tornam as legumaínas importantes no processamento de outras proteínas em suas respectivas formas maduras, para ativá-las, mobilizá-las, ou ainda, estocá-las de acordo com as necessidades metabólicas e fisiológicas da planta (Hara-Nishimura *et al.*, 1991).

Nas sementes de plantas, as legumaínas são responsáveis pelo processamento pós-traducional de diversas proteínas, atuando em vias de biogênese ou armazenamento das mesmas. Desta forma, as legumaínas contribuem para a biossíntese de diversos componentes vacuolares nas sementes, além de outros tecidos, como também no endosperma, nos cotilédones, na raiz, no hipocótilo e nas folhas (Hara-Nishimura *et al.*, 1995). Durante a germinação e o crescimento da plântula as legumaínas são importantes na ativação de outras proteases cisteínicas (como as da família C1A, tipo papaínas), para a degradação de proteínas de armazenamento e desenvolvimento da planta (Okamoto and Minamikawa, 1999; Kato *et al.*, 2003). Na defesa contra patógenos também há uma atuação das legumaínas nas respostas de morte celular programada e resposta hipersensível, uma vez que legumaínas possuem atividade similar às caspases (Hatsugai *et al.*, 2004; Rojo *et al.*, 2004). Em plantas, não são conhecidos genes específicos de caspases, portanto, outros genes, tais como as legumaínas são capazes de desempenhar estas funções na morte celular (Bonneau *et al.*, 2008). As caspases são endopeptidases aspartato específicas, extremamente importantes para o processo de apoptose em células animais, mas que são estruturalmente relacionadas com as legumaínas de plantas, possuindo especificidade semelhante pelo sítio de clivagem de suas proteínas-alvo (Chen *et al.*, 1998; Shutov *et al.*, 2012).

Usualmente, as legumaínas de plantas são separadas em dois grandes grupos, (1) semente e (2) vegetativo, devido ao seu padrão de expressão gênica em diferentes tecidos e pela homologia entre suas sequências peptídicas (Müntz *et al.*, 2002; Müntz and Shutov, 2002). Estas características estão provavelmente relacionadas com a especificidade dos papéis desempenhados pelas legumaínas durante o ciclo de vida da planta. Entretanto, nem todas as legumaínas se encaixam nesta classificação. Em *Arabidopsis thaliana*, quatro genes homólogos de legumaínas estão relativamente bem descritos, α VPE e γ VPE são mais específicas em órgãos vegetativos, β VPE em sementes e δ VPE, diferente das duas famílias propostas, é encontrada em estágios centrais da embriogênese, antes da síntese das proteínas de armazenamento (Yamada *et al.*, 2005). A inclusão das legumaínas nestes dois

grandes grupos pode ser relacionada com a classificação dos tipos vacuolares de plantas: em vacúolos de armazenamento e vacúolos líticos. Desta forma α VPE e γ VPE possuem expressão gênica detectada em vacúolos líticos de órgãos vegetativos, e os alvos destas duas legumainas provavelmente são enzimas hidrolíticas vacuolares, induzidas pela senescência. As legumainas geralmente são pouco detectadas em tecidos de folhas saudáveis, mas aparecem de forma marcante nos tecidos em senescência, seguidas pela ativação de proteases cisteínicas em resposta à senescência e degradação do tecido (Hara-Nishimura *et al.*, 1998; Kinoshita *et al.*, 1999). Em plantas, os homólogos de β VPE são detectados em vacúolos de armazenamento proteicos, desempenhando funções na maturação de algumas proteínas (Hara-Nishimura and Nishimura, 1987; Hara-Nishimura *et al.*, 1993, 1995; Kinoshita *et al.*, 1999). Também, em *Arabidopsis*, a expressão do gene β VPE diminui após a germinação da semente e continua diminuindo gradativamente com o crescimento da plântula (Kinoshita *et al.*, 1999). As diferenças entre o padrão de expressão e funções destes dois grupos de legumainas sugerem que existam diferentes elementos regulatórios na região promotora destes genes, os quais são responsáveis pela expressão espacial e temporal diferenciada das legumainas (Kinoshita *et al.*, 1999; Müntz and Shutov, 2002). Desta forma, sabe-se da importância dos papéis desempenhados pelas legumainas nos organismos, entretanto existem poucos estudos moleculares e funcionais sobre a diversidade e especificidade desta família de proteases.

3. O modelo vegetal

O arroz (*Oryza sativa* L.) é o alimento diário de pelo menos metade da população de nosso Planeta, proporcionando também uma grande fonte de renda em função de sua produção. O Brasil é um dos maiores produtores mundiais de arroz, sendo esta produção destacada na região Sul, onde totaliza 60% da produção nacional (<http://sistemasdeproducao.cnptia.embrapa.br>). Entre as monocotiledôneas o arroz pode ser considerado uma planta modelo de estudos moleculares e funcionais, sendo que diversos protocolos de manipulação genética estão bem estabelecidos, além de muitas informações moleculares, tais como sequência completa do genoma, bancos de ESTs, microarranjos, transcriptomas, entre inúmeras outras informações. Assim, o arroz torna-se um excelente modelo para estudos, uma vez que, queira-se entender mecanismos biológicos básicos das plantas, necessários à sua sobrevivência e evolução, para

futuramente modificar geneticamente plantas a fim de melhorar cada vez mais seu cultivo e utilização.

4. Fitocistatinas em arroz (*Oryza sativa*)

A primeira fitocistatina a ser identificada em plantas, na década de 80, foi a oryzacistatina-I (OcI), encontrada em sementes de arroz. Sua caracterização demonstrou que ela era capaz de inibir efetivamente proteases cisteínicas do tipo papaínas (Abe *et al.*, 1987b), através dos domínios cistatina característicos, mas sem a necessidade do resíduo de glicina (Gly) na região amino-terminal (Abe *et al.*, 1988). Posteriormente a sequência genômica codificadora da OcI foi identificada e descrita, onde o gene OcI possuía cerca de 1.4kb de tamanho, com três exons e dois introns (Hiroto *et al.*, 1989). Em 1990, uma nova fitocistatina foi isolada em arroz, a OcII, também encontrada em sementes e com aproximadamente 55% de identidade com a OcI em nível proteico (Kondo *et al.*, 1990). Neste estudo, as comparações realizadas entre OcI e OcII revelaram algumas peculiaridades de cada gene, como a diferença entre a especificidade inibitória de cada um, onde OcI inibe mais eficientemente papaínas do que OcII. A localização de OcI e OcII em sementes maduras de arroz sugeriu seus envolvimento na regulação da proteólise desencadeada por diversos tipos de proteases (Kondo *et al.*, 1990). Posteriormente, a descrição da organização genômica de OcII revelou também uma grande similaridade estrutural em relação à OcI (Kondo *et al.*, 1991). Estas descobertas e caracterizações iniciais desencadearam uma busca por fitocistatinas em outros organismos do reino vegetal, além de uma contínua caracterização e validação da atividade inibitória das fitocistatinas já descritas de arroz (Arai *et al.*, 1991; Turk and Bode, 1991; Urwin *et al.*, 1995a). A utilização destes inibidores de proteases cisteínicas também começou a ser avaliada de forma recombinante, com efeito sobre proteases de patógenos e parasitas (Chen *et al.*, 1992; Urwin *et al.*, 1995b). A OcI foi a primeira fitocistatina a ter sua estrutura tridimensional avaliada por cristalografia, com uma resolução de 2.8 Å (Kudo *et al.*, 1998), sendo posteriormente melhor resolvida e mais detalhada por RMN (Ressonância Magnética Nuclear), revelando uma maior similaridade desta fitocistatina com as cistatinas (tipo 2) de animais (Nagata *et al.*, 2000). Mais recentemente, em calos de arroz da variedade *nipponbare*, foi realizada a caracterização de uma terceira fitocistatina em arroz, a OcIII, altamente homóloga à OcII, mas com algumas diferenças na região amino-

terminal e na sua organização genômica. A OcIII assim como as outras duas fitocistatinas, OcI e OcII, possui capacidade *in vitro* de inibição de proteases do tipo papaínas, sendo que a constante de inibição da OcIII difere de OcI e OcII. A eficiência de inibição de papaínas obedece a ordem OcI > OcIII > OcII, além da OcIII ser a única das três com capacidade de inibir eficientemente a catepsina B (Ohtsubo *et al.*, 2005). Assim, grande parte das descobertas iniciais, relacionadas a características e funções das fitocistatinas, deve-se aos estudos aprofundados com as fitocistatinas de arroz.

O sequenciamento do genoma do arroz proporcionou uma ampla identificação dos genes homólogos de fitocistatinas, totalizando 12 genes encontrados e anotados como OcI, OcII, OcIII, OcIV, OcV, OcVI, OcVII, OcVIII, OcIX, OcX, OcXI e OcXII (Martínez *et al.*, 2005). Estes genes encontram-se nos respectivos *loci* gênicos: Os01g58890, Os05g41460, Os05g33880, Os01g68660, Os01g68670, Os03g11180, Os03g11170, Os03g31510, Os03g11160, Os04g28250, Os09g08100 e Os01g16430, e foram encontrados tanto no genoma da sub-espécie *japônica*, quanto na subespécie *indica*, com apenas algumas variações de aminoácidos na região do peptídeo sinal. Ainda, segundo Martinez *et al.*, 2005, a organização genômica das fitocistatinas em arroz demonstra que três genes (OcI, OcII e OcIII) possuem um intron entre as sequências que codificam o motivo conservado LARFAV e o sítio ativo QxVxG (Fig 4). Algumas das sequências de arroz não possuem introns em suas ORFs (OcIV- OcXI), e apenas um dos genes, OcXII, possui três introns. Estas sequências estão distribuídas em cinco cromossomos diferentes, sendo que algumas delas estão arranjadas de forma consecutiva, o que poderia inferir a existência de eventos de duplicação entre as fitocistatinas de arroz (Fig 5).

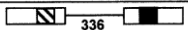
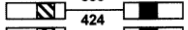
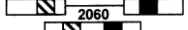


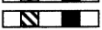




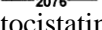
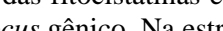
| Gene name | Locus id | Predicted gene structure | Amino acids |
|-----------|------------|--|-------------|
| OC-I | Os01g58890 |  | 139 |
| OC-II | Os05g41460 |  | 156 |
| OC-III | Os05g33880 |  | 150 |
| OC-IV | Os01g68660 |  | 158 |
| OC-V | Os01g68670 |  | 147 |
| OC-VI | Os03g11180 |  | 113 |
| OC-VII | Os03g11170 |  | 117 |
| OC-VIII | Os03g31510 |  | 123 |
| OC-IX | Os03g11160 |  | 114 |
| OC-X | Os04g28250 |  | 156 |
| OC-XI | Os09g08100 |  | 120 |
| OC-XII | Os01g16430 |  | 250 |

Figura 4. Estrutura e organização genômica das fitocistatinas em arroz. O acrônimo referente ao nome de cada gene é indicado antes de cada *locus* gênico. Na estrutura predita para cada gene, o tamanho dos introns é indicado abaixo de cada linha representativa, e nos exons, representados pelas caixas, o motivo LARFAV é indicado pelas linhas diagonais e o motivo QxVxG pelas regiões em preto. Ao lado de cada estrutura encontra-se indicado o tamanho predito da proteína resultante em aminoácidos. Figura reproduzida de Martinez *et al.*, 2005.

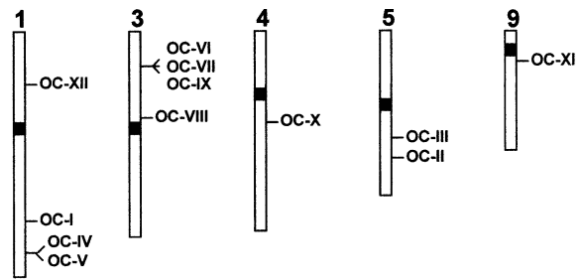


Figura 5. Organização cromossômica dos genes de fitocistatinas em arroz. Os genes estão distribuídos em 5 cromossomos. As fitocistatinas OcIV e OcV apresentam um padrão de duplicação no cromossomo 1, enquanto que as OcVI, OcVII e OcIX apresentam este padrão no cromossomo 3. Figura reproduzida de Martinez *et al.*, 2005.

A sequência proteica predita para as fitocistatinas de arroz, juntamente com os motivos proteicos encontrados, sugere que elas possuem uma estrutura tridimensional similar à descrita para OcI (Nagata *et al.*, 2000). A maior parte dos resíduos que constituem a estrutura das cinco folhas- β e a α -hélice são conservados entre todas as sequências proteicas identificadas em arroz, além de uma grande conservação dos motivos envolvidos na interação com as proteases cisteínicas. Entretanto, três sequências de fitocistatinas em arroz, OcVII, OcIX e OcXI, não possuem o aminoácido triptofano (W), importante no segundo *loop* da estrutura da proteína para a inibição das proteases-alvo (Koiwa *et al.*, 2001; Martínez *et al.*, 2005). Adicionalmente, em arroz, existe uma única fitocistatina com a extensão carboxi-terminal, a OcXII, e a provável capacidade de inibir proteases cisteínicas do tipo legumainas, além de proteases tipo papaínas (Martínez *et al.*, 2005; Martinez *et al.*, 2007; Margis-Pinheiro *et al.*, 2008).

Alguns estudos iniciais levaram em consideração o perfil de expressão gênica das fitocistatinas em arroz avaliando bancos de ESTs disponíveis, para alguns tecidos como folha, raiz, flor e semente. A expressão de OcI foi encontrada nestes quatro tecidos avaliados, enquanto outras fitocistatinas (OcIV, OcV, OcVI, OcVII, OcIX e OcXI) não foram detectadas nas bibliotecas de ESTs analisadas. As demais fitocistatinas (OcII, OcIII, OcVIII, OcX e OcXII) apresentam um padrão mais restrito de expressão gênica, a OcII sendo expressa em folha, raiz e flor, a OcIII apenas em folha e flor, a OcVI em folha e semente, a OcX em folha e raiz e a OcXII sendo expressa em folha e semente (Martínez *et al.*, 2005). Desta forma, sabe-se que as doze fitocistatinas de arroz possuem um padrão diverso de expressão gênica e atividade, sendo que poucas delas foram caracterizadas detalhada e extensivamente, através de novas técnicas de biologia molecular. Conhecendo melhor cada gene de fitocistatina em arroz, será possível compreender melhor os mecanismos de controle de proteólise locais e específicos da planta.

Capítulo II

- Objetivos -

As fitocistatinas são caracteristicamente inibidores de proteases cisteínicas do tipo papaínas, entretanto a descoberta de que algumas fitocistatinas possuem uma extensão carboxi-terminal levantou a possibilidade de uma nova função para estes genes. Alguns estudos com genes homólogos sugerem a bifuncionalidade para estas fitocistatinas, as quais teriam a capacidade de inibir tanto proteases do tipo papaínas quanto legumaínas. Muitas vezes os genomas de plantas apresentam redundância gênica, dificultando a compreensão da função de um gene específico. Excepcionalmente, o arroz (*Oryza sativa*) apresenta uma única cópia da fitocistatina carboxi-estendida (OcXII). Assim o objetivo geral deste trabalho, é compreender a diversidade e os mecanismos de interação entre as fitocistatinas e suas proteases cisteínicas alvo em arroz, com ênfase nas implicações funcionais da OcXII para o desenvolvimento vegetal. Para isto, os objetivos específicos incluem:

- 1) Determinar o perfil de expressão dos genes de fitocistatinas e proteases cisteínicas em diferentes tecidos, durante o desenvolvimento vegetal e sob condições de estresse ambiental;
- 2) Caracterizar genes de legumaínas em plantas sob aspectos evolutivos, relacionando-os com a diversidade funcional das mesmas;
- 3) Analisar linhagens de plantas silenciadas para o gene OcXII, verificando alterações fenotípicas e moleculares;
- 4) Avaliar o perfil da atividade de proteases cisteínicas em plantas OcXII silenciadas, identificar espacial e temporalmente as proteases com atividade alterada;
- 5) Relacionar a atividade da fitocistatina carboxi-estendida com funções biológicas da planta durante o desenvolvimento, ou frente à respostas ambientais.

Capítulo VI

- Discussão-

1. A família das fitocistatinas em arroz, diversidade e especificidades

As fitocistatinas, funcionalmente importantes como inibidores reversíveis e de alta afinidade por proteases cisteínicas, começaram a ser descobertas em plantas na década de 80 (Abe *et al.*, 1987b). Mais recentemente, o sequenciamento de genomas completos possibilitou a busca por genes homólogos em diferentes organismos, revelando assim uma grande variedade de genes de cistatinas e fitocistatinas conservados entre eucariotos e procariotos (Margis *et al.*, 1998; Martinez and Diaz, 2008; Kordis and Turk, 2009). Em arroz (*Oryza sativa*), 12 genes de fitocistatinas foram preditos, anotados e denominados de OcI a OcXII (Martínez *et al.*, 2005), em referência às oryzacistatinas (cistatinas de arroz) previamente descritas e caracterizadas (OcI, OcII e OcIII) (Ohtsubo *et al.*, 2005). Nesta identificação inicial, as 12 fitocistatinas de arroz foram analisadas de forma estrutural, com avaliação de alguns resultados funcionais obtidos em bancos de dados de expressão gênica (EST–*Expressed sequence tags*) (Martínez *et al.*, 2005).

De forma geral, neste trabalho, nós demonstramos que a família gênica das fitocistatinas em arroz é evolutivamente bem conservada, sendo que apenas um dos 12 genes não possui o domínio inibitório cistatina (OcIX), e outro gene (OcXII), possui uma extensão carboxi-terminal com o domínio similar a cistatina. Outros 3 genes (OcVII, OcIX e OcXI), não apresentam o resíduo de triptofano que compõe o domínio cistatina, conforme já detectado anteriormente (Martínez *et al.*, 2005). O agrupamento filogenético das fitocistatinas em arroz reflete várias tendências observadas na expressão destes genes, relacionando diretamente o processo evolutivo e as funções particulares de cada fitocistatina em arroz. Analisando os genes de fitocistatinas de forma global, incluindo os genes homólogos encontrados em 26 genomas de plantas, é possível observar a existência de três grandes grupos filogenéticos. O primeiro grupo de sequências, bem suportado, possui uma relação mais estreita com as fitocistatinas de plantas basais, o que sugere uma relação mais conservada entre eles, contendo sequências que não divergiram muito de suas formas ancestrais. Um segundo e terceiro grupo de sequências são bem suportados internamente, mas a relação global com as demais fitocistatinas, não está bem estabelecida. Isto sugere de forma clara que, os genes destes dois grupos, são genes com sequências mais divergentes, que estão menos evolutivamente relacionados com as sequências ancestrais de plantas basais, podendo ter se originado em eventos de duplicação singulares

de cada espécie ou grupo mais proximamente relacionado (ex. gramíneas). Esta hipótese baseia-se conjuntamente nas informações estruturais e funcionais obtidas para as fitocistatinas de arroz, onde OcI, OcII, OcIII e OcXII (grupo 1), são os únicos genes com introns e com níveis de expressão gênica mais elevado. As demais fitocistatinas, OcIV – OcXI, são genes estruturalmente sem introns, sendo que alguns deles demonstram um perfil evidente de duplicação gênica, detectado tanto pelo sinal filogenético, quanto pela posição cromossômica em *loci* sequenciais (Martínez *et al.*, 2005), além da perda de domínios funcionais em alguns destes genes.

Em arroz, as fitocistatinas apresentam um padrão específico e variado de expressão gênica em diferentes órgãos, tecidos ou estágios de desenvolvimento. Entretanto, ainda que haja variações, é possível detectar uma expressão gênica mais significativa entre as fitocistatinas que compõe o grupo 1, o que reforça a hipótese de que estes genes em particular são os mais conservados evolutivamente, retendo as funções primárias de inibidores de proteases cisteínicas de forma mais ampla. A análise de dados de microarranjos, RNAseq e RT-qPCR revelam que os genes mais ubiquamente expressos são os que compõe o grupo 1 (OcI, OcIII e OcXII), enquanto que os genes dos grupos 2 e 3 tem uma expressão gênica mais basal e tecido específica, ou mesmo sem expressão detectada (OcVII e OcIX). Este padrão de expressão gênica, função e conservação evolutiva também é observado em cevada, onde a maioria dos genes homólogos aos do grupo 3 não tem níveis de mRNA detectáveis, com baixa expressão proteica e sem capacidade inibitória contra proteases cisteínicas endógenas (Martinez *et al.*, 2009). Desta forma, pode-se verificar que apesar do número elevado de genes de fitocistatinas em plantas, nem todos parecem ser funcionais, existindo genes parálogos originados em eventos de duplicação, ou homólogos que perderam sua função. Adicionalmente, alguns destes genes, ortólogos ou parálogos, parecem ter adquirido uma subfuncionalização para atividades específicas mais restritas em alguns tecidos ou algumas situações fisiológicas. Entretanto, o grupo de fitocistatinas mais conservado e com níveis de expressão gênica mais significativa (grupo 1), ao que tudo indica, parece conter os genes majoritariamente responsáveis pelo controle da proteólise nas plantas (OcI, OcIII e OcXII).

1.1. Fitocistatinas carboxi-estendidas, diversificação e evolução funcional

A descoberta de fitocistatinas com sequências maiores, devido à existência de um domínio adicional similar à cistatina em suas extremidades carboxi-terminais, levou a uma diferenciação destes genes nos genomas de plantas e os mesmos são conhecidos como fitocistatinas carboxi-estendidas (Martínez *et al.*, 2005; Margis-Pinheiro *et al.*, 2008). Estes inibidores carboxi-estendidos possuem dois domínios principais, o primeiro deles é o domínio cistatina canônico, localizado na parte amino-terminal da proteína, ele possui os motivos (QxVxG e W) nos *loops* inibitórios da estrutura tridimensional. O segundo domínio, na região carboxi-terminal é um domínio similar a cistatina, mas que de fato, não possui similaridade de sequência com as cistatinas, ao invés disto, ele possui o domínio SNSL conservado (Martinez *et al.*, 2007). Nossas análises de modelagem tridimensional realizadas com a fitocistatina carboxi-estendida de arroz, a OcXII, demonstram que a estrutura desta proteína têm duas regiões principais separadas espacialmente em dois lobos, de acordo com os domínios inibitórios preditos N-terminal e C-terminal.

As fitocistatinas carboxi-estendidas puderam ser detectadas nos genomas de todas as 26 plantas com genoma sequenciado analisadas neste estudo, com algumas diferenças entre número de cópias por genoma. As relações filogenéticas encontradas sugerem uma única origem ancestral, com divergência posterior, durante a especiação das famílias ou gêneros de plantas. Previamente, já havia sido proposto que a extensão carboxi-terminal das fitocistatinas originou-se a partir da duplicação de uma cistatina, seguida de evolução divergente (Martinez and Diaz, 2008; Margis-Pinheiro *et al.*, 2008; Benchabane *et al.*, 2010). Interessantemente, plantas mais basais como musgos, possuem duas cópias das fitocistatinas carboxi-estendidas e este padrão se repete em muitas espécies de dicotiledôneas. Além disso, as duas cópias gênicas detectadas em dicotiledôneas diferem entre si por várias substituições de aminoácidos, inclusive em regiões próximas aos sítios inibitórios. Se esta segunda cópia gênica, mais divergente, mantém o mesmo padrão funcional da cópia mais conservada, ainda é uma questão em aberto para ser testada. Diferentemente, as monocotiledôneas analisadas possuem apenas uma cópia da fitocistatina estendida, o que sugere a ocorrência de uma perda gênica provavelmente em algum ancestral das monocotiledôneas.

Comprovadamente, algumas cistatinas de animais (ex. cistatinas humanas C, E/M e F) são aptas a inibir consecutivamente proteases C1A (papaínas) e C13 (legumaínas),

podendo inclusive formar um complexo ternário (papaína-cistatina-legumaína). Entretanto, esta inibição é realizada por dois sítios diferentes, mas espacialmente próximos na estrutura do inibidor: o domínio cistatina (QxVxG e W), inibidor de proteases do tipo papaínas, e o resíduo de Asn logo após a α -hélice principal da estrutura, inibindo legumaínas (Alvarez-Fernandez *et al.*, 1999). Em plantas, o que podemos observar é uma possível reestruturação das fitocistatinas, onde a principal função inibitória de proteases C1A permaneceu em todas as fitocistatinas, enquanto que a habilidade de inibição de proteases C13 parece ter sido transferida para a extensão C-terminal, onde encontra-se o resíduo de asparagina. Esta separação entre os dois domínios inibitórios, em plantas, provavelmente deve ter ocorrido em algum ancestral das embriófitas, logo após a duplicação do domínio cistatina. Este domínio duplicado, por sua vez, começou a divergir e reestruturar-se de forma específica, sofrendo diversas substituições de aminoácidos, perdendo o domínio QxVxG, porém mantendo sua estrutura tridimensional similar as cistatinas, mas apenas com domínio SNSL, contendo o resíduo de asparagina (N) conservado. Desta forma, o processo evolutivo em plantas parece ter direcionado as atividades inibitórias de diferentes grupos de proteases para regiões espacialmente mais distantes no inibidor, assim originando uma fitocistatina carboxi-estendida. A eficiência de inibição de diferentes proteases pelas cistatinas animais ou pelas fitocistatinas carboxi-estendidas de plantas, ainda é um ponto a ser revisto para futuras comparações dos mecanismos inibitórios.

No genoma de arroz, existe apenas uma cópia gênica da fitocistatina carboxi-estendida, denominada OcXII. Direcionando nossas análises para este gene, foi possível obter uma estrutura tridimensional hipotética através de modelagem por homologia. Esta estrutura demonstra claramente que ambos os domínios amino-terminal e carboxi-terminal são estruturalmente similares, apesar das diferenças existentes em termos de sequência nucleotídica e de aminoácidos, conforme predito anteriormente. No modelo é possível verificar que ambos os domínios estão espacialmente separados na estrutura da proteína, levantando-se então à hipótese de que OcXII pode ser o inibidor bifuncional de proteases cisteínicas em arroz, atuando sobre proteases do tipo papaína (C1A) com a região amino-terminal, e também sobre legumaínas (C13) com o domínio carboxi-terminal.

1.2. Variações das fitocistatinas de arroz em tecidos específicos e respostas a fatores ambientais

O envolvimento das fitocistatinas em respostas a variações e estresses ambientais é um fato amplamente estabelecido, uma vez que, durante situações adversas as atividades proteolíticas estão envolvidas no controle de inúmeras funções fisiológicas (Pesquet, 2012). Sendo assim, muitas correlações e inferências funcionais podem ser obtidas analisando-se os perfis de atividade das fitocistatinas com suas proteases-alvo. Em arroz, a família gênica das fitocistatinas apresenta perfis variados de expressão gênica, considerando os 12 genes em diferentes tecidos e situações ambientais.

Algumas fitocistatinas possuem atividade antifúngica, inibindo o crescimento de micélios, a germinação de esporos e fungos fitopatogênicos (Yang and Yeh, 2005; Martinez *et al.*, 2005; Wang *et al.*, 2008; Pirovani *et al.*, 2010; Carrillo *et al.*, 2011), proporcionando resistência a insetos e fitopatógenos (Senthilkumar *et al.*, 2010; Jung *et al.*, 2010). Em arroz, através de RT-qPCR nós pudemos detectar que a utilização de quitosan como um elicitor fúngico desencadeia um aumento na expressão gênica das fitocistatinas, principalmente de OcI e OcXII, o que também já havia sido previamente detectado em nível de proteínas para as fitocistatinas de maneira geral (Agrawal *et al.*, 2002). Adicionalmente, as fitocistatinas têm a capacidade de inibir proteases exógenas, participando dos mecanismos de defesa das plantas. Em *Theobroma cacao*, a fitocistatina TcCYS4 auxilia na prevenção da morte celular em folhas inibindo a atividade das proteases (Pirovani *et al.*, 2010), assim como a AtCYS1 de *A. thaliana* é induzida após ferimentos mecânicos, suprimindo a morte celular e resposta hipersensível (Belenghi *et al.*, 2003). Quando as plantas de arroz sofrem alguma injúria em suas folhas, todas as fitocistatinas têm um decréscimo momentâneo em seus níveis de mRNA, entretanto, OcIII e OcXII são as mais responsivas em estágios posteriores demonstrando um considerável aumento de expressão gênica após 24h. Assim, de forma mais específica, a OcI de arroz é mais responsiva a infecção por patógenos, enquanto OcIII e OcXII parecem estar mais envolvidas em respostas a ferimentos mecânicos. Em outros processos como a senescência, onde várias proteases cisteínicas são ativadas, algumas fitocistatinas tem um declínio significativo em seus níveis de mRNA (OcI, OcIV, OcVI, OcX e OcXI), o que faz-se necessário para que haja a ativação das proteases cisteínicas.

Considerando as inúmeras situações adversas às quais uma planta pode ser exposta, alguns estudos apontam que genes homólogos a fitocistatinas em *Arabidopsis*, tem um aumento de expressão gênica em condições de estresse salino (Zhang *et al.*, 2008). Em arroz, as fitocistatinas OcI, OcIII e OcXII são induzidas quando há um aumento de salinidade no meio. Outras inúmeras respostas fisiológicas são mediadas pelo fitormônio ABA (ácido abscísico), que é uma molécula sinalizadora altamente importante para as plantas (Lee and Luan, 2012). Algumas fitocistatinas de arroz são responsivas a este hormônio vegetal, entre elas OcI, OcX e OcXII têm um aumento de expressão gênica, enquanto OcIV, OcVI e OcXI têm um decréscimo significativo nos níveis de mRNA. Similarmente, alguns genes homólogos em *Arabidopsis* (AtCYS3 e AtCYS6), são genes tipicamente induzíveis por ABA, e capazes de controlar as taxas de germinação e crescimento da plântula (Zhang *et al.*, 2008).

Estas análises, de forma geral, estão relacionadas com o que se sabe sobre os genes homólogos em outras plantas, como em *Arabidopsis*, uma planta modelo de estudos em dicotiledôneas. Aqui, a demonstração da variedade funcional das fitocistatinas em monocotiledôneas, adiciona informações importantes sobre as fitocistatinas. Em arroz, assim como em outras plantas, as fitocistatinas também apresentam importantes papéis no controle das proteases, refletindo no tempo de germinação das sementes e crescimento das plântulas. Nós demonstramos ainda, que os inibidores de proteases cisteínicas em arroz têm padrões de expressão gênica específicos em diferentes tecidos, e em resposta a diferentes condições ambientais, as quais podem ser futuramente relacionadas diretamente às proteases-alvo e suas especificidades de interações, que podem diferir entre tecidos e condições ambientais.

1.3. O importante papel das fitocistatinas durante a germinação e o crescimento das plantas

Durante a germinação, as fitocistatinas desempenham um papel importante na inibição de proteases cisteínicas, controlando assim a germinação através da regulação de atividade de papaínas que, por exemplo, mobilizam outras proteínas de armazenamento provendo os nutrientes necessários para a planta em formação (Arai *et al.*, 1998, 2002; Hwang *et al.*, 2009; Benchabane *et al.*, 2010; Martínez *et al.*, 2012). Os primeiros

trabalhos com as fitocistatinas já demonstravam sua importância em sementes de arroz (Abe *et al.*, 1987a). Grande parte das fitocistatinas que detectamos aqui apresenta níveis bastante elevados de expressão gênica nas sementes dormentes, entretanto logo após 12h do início do processo de germinação estes níveis caem drasticamente. Isto ocorre principalmente com as fitocistatinas OcI, OcIII, OcVI, OcVIII e OcXII, e esta queda é inversamente proporcional à atividade das proteases cisteínicas avaliadas, as quais têm um aumento. Esta correlação inversa corrobora a hipótese de inibição pelas fitocistatinas e diminuição da atividade das proteases-alvo, onde, o decréscimo da expressão gênica das fitocistatinas, faz com que estes inibidores estejam presentes em menor quantidade nas células, permitindo que as proteases-alvo exerçam as suas funções proteolíticas. A degradação das proteínas de reserva e remobilização de nutrientes é um dos eventos mais importantes para a germinação e crescimento inicial das plântulas. Em um modelo proposto, durante a formação das sementes há um aumento na quantidade das fitocistatinas, enquanto ocorre a deposição das proteínas de armazenamento. Esta quantidade de fitocistatinas permanece alta durante o período de dormência da semente, para a proteção destas proteínas de armazenamento. Então, no momento em que inicia a germinação, há uma diminuição do nível de fitocistatinas, de forma a aumentar a atividade das proteases necessárias para desencadear a hidrólise e mobilização das proteínas armazenadas (Benchabane *et al.*, 2010).

2. Proteases cisteínicas do tipo legumaínas em arroz, identificação e caracterização gênica

Legumaínas, ou VPEs (*vacuolar processing enzymes*), são proteases cisteínicas que pertencem à família C13 (EC 3.4.22.34) e são amplamente encontradas em plantas e animais. Estas proteases possuem a habilidade de realizar proteólise limitada do substrato após resíduos de asparagina (Asn), e em alguns casos após resíduos de ácido aspártico (Asp) (Abe *et al.*, 1993; Becker *et al.*, 1995; Hiraiwa *et al.*, 1999), mas também são capazes de desempenhar proteólise ilimitada, dependendo da conformação do substrato. No genoma do arroz, existem 5 genes de legumaínas que podem ser agrupados filogeneticamente nos grupos pré-estabelecidos de legumaínas “vegetativas” e legumaínas de “semente”. Estes grupos foram estabelecidos previamente com base em estudos de expressão gênica e similaridade de sequências de legumaínas de plantas (Hara-Nishimura

et al., 1991; Hara-Nishimura and Hatsugai, 2011; Shutov *et al.*, 2012). Adicionalmente um grupo de sequências a parte é encontrado apenas em dicotiledôneas, contendo genes de legumainas detectados durante a embriogênese (Yamada *et al.*, 2005). A análise filogenética aqui apresentada sugere que todas estas diversificações funcionais de subgrupos de legumainas tiveram uma origem ancestral, durante o processo de especiação de Viridiplantae. As legumainas de arroz foram nomeadas e classificadas de acordo com seu *locus* gênico e posição filogenética. OsaLeg1, OsaLeg4 e OsaLeg5 são genes caracteristicamente do grupo “vegetativo”, enquanto que OsaLeg2 e OsaLeg3 estão incluídas no grupo de legumainas características de sementes. Com exceção de OsaLeg5 que apresenta um padrão de duplicação gênica a partir de OsaLeg4, os demais genes de arroz apresentam sequência e estrutura conservada com os homólogos descritos em *Arabidopsis*. Esta visão inicial propiciou um ponto de partida para a caracterização mais específica dos genes de legumainas em arroz.

As análises de expressão gênica demonstram que há uma correlação entre o perfil de mRNA detectado e a classificação nos grupos de tecidos propostos. OsaLeg1 e OsaLeg4 tem expressão gênica mais elevada em tecidos vegetativos, enquanto OsaLeg2 e OsaLeg3 tem seus maiores níveis de expressão gênica detectados em sementes. Entretanto, esta correlação é apenas superficial e deve ser utilizada com cuidado, uma vez que todas as legumainas expressas foram detectadas em todos os tecidos e regiões da planta analisados, mas nem sempre com alto nível de expressão gênica. OsaLeg3 por exemplo, é inclusive mais expressa em tecidos vegetativos do que OsaLeg4, entretanto, os maiores níveis de mRNA de OsaLeg3 encontram-se em sementes, enquanto que os maiores níveis de OsaLeg4 encontram-se em tecidos vegetativos. Apenas OsaLeg5 não teve sua expressão gênica detectada. Desta forma, esta classificação em “vegetativo” e “semente” aplica-se apenas quando consideramos o tecido com o maior nível de expressão das legumainas, mas sem excluir que a mesma também seja transcrita nos demais tecidos da planta. Assim como, já haviam indícios prévios de que a expressão das legumainas não é mutualmente exclusiva em diferentes tecidos, podendo, inclusive, uma legumaina complementar a função de outra (Gruis *et al.*, 2004).

Em arroz, dois trabalhos prévios já haviam identificado a existência da atividade de legumainas em sementes, tanto de OsaLeg2 quanto de OsaLeg3, levando em consideração suas semelhanças com a β -VPE de *Arabidopsis*. OsaLeg2 foi reportada como

importante na ativação de uma protease cisteínica do tipo papaína, para a digestão de proteínas de armazenamento (Kato *et al.*, 2003). Já a OsaLeg3, da mesma forma relacionada com a β -VPE, teve sua importância estabelecida no processamento eficiente da gluteína em sementes de arroz (Wang *et al.*, 2009).

Algumas legumainas “vegetativas” já foram amplamente caracterizadas pela sua importância em processos como a senescência, ativando outras proteases ou degradando outras proteínas (Rojo *et al.*, 2003; Donnison *et al.*, 2007; Hara-Nishimura and Hatsugai, 2011). Estas proteases encontram-se no mesmo grupo filogenético das legumainas “vegetativas” de arroz, OsaLeg1, OsaLeg4 e OsaLeg5, as quais possuem um nível de expressão gênica ainda mais acentuado em folhas senescentes (exceto OsaLeg5), sugerindo o envolvimento destas proteases no metabolismo de senescência em folhas de arroz. De forma similar, estas legumainas “vegetativas” podem estar correlacionadas com o processo de morte celular nas plantas, na senescência natural ou induzida por alguns tipos de patógenos (Beers *et al.*, 2000). Em tabaco foi claramente demonstrada a importância das legumainas na indução de morte celular através de sua atividade de caspase-1, desencadeando uma resposta hipersensível após infecção por vírus, prevenindo sua dispersão pela planta (Hatsugai *et al.*, 2004). Em arroz, as legumainas apresentaram um aumento de expressão gênica em plantas que sofreram injúrias mecânicas, revelando um potencial de ativação em resposta a mecanismos de morte celular, assim como também reportado em outras plantas (Yamada *et al.*, 2004; Hara-Nishimura *et al.*, 2005).

A influência do ácido abscísico na ativação da transcrição de genes de legumainas é bastante relevante (Santos-Silva *et al.*, 2012; Julián *et al.*, 2013), da mesma forma, as legumainas de arroz apresentam respostas variadas ao tratamento com ABA, sendo induzidas de maneira geral. Entretanto, utilizando-se nicotinamida para bloquear a síntese endógena de ABA, em um experimento de estresse por sal, pode-se perceber que apenas OsaLeg2 é diretamente sensível aos níveis de ABA endógenos. Para as demais legumainas de arroz este efeito não foi evidente e direto, podendo sugerir mecanismos indiretos de resposta ao ABA. A alta sensibilidade de OsaLeg2 ao ABA pode ser relacionada ao fato de que esta é a legumaina mais expressa em sementes secas. Ao mesmo tempo em que o ABA controla a dormência das sementes, quando a germinação inicia os níveis de ABA diminuem e conseqüentemente a expressão gênica de OsaLeg2 também diminui drasticamente. Isto ocorre de forma inversa para as demais legumainas durante a

germinação, onde a diminuição gradativa dos níveis de ABA leva ao aumento dos níveis de transcrição dos genes de OsaLeg1, OsaLeg4 e inclusive OsaLeg3 ao longo do período de germinação e crescimento da plântula.

Algumas sequências de legumaínas anotadas no genoma de arroz apresentam diferentes isoformas transcritas para um mesmo gene, configurando possíveis sequências com *splicing* alternativo. As variações de *splicing* de um gene resultam em importantes fontes de diversificação proteômica, podendo aumentar a variabilidade dos organismos. As legumaínas OsaLeg1 e OsaLeg2 apresentam duas versões de transcritos para cada gene (1.1 e 1.2, 2.1 e 2.2), e sob condições ambientais estáveis a isoforma 1 (maior e mais completa) é sempre a mais abundante para ambos os genes. Contudo, quando as plantas sofrem algum ferimento em suas folhas, a isoforma OsaLeg1.2 passa a ser mais abundante. Esta isoforma não possui a sequência do peptídeo sinal, o que faz com que seu mRNA correspondente provavelmente não atinja os compartimentos celulares de destino (retículo endoplasmático e vacúolo), sugerindo que talvez esta proteína acabe ficando restrita ao citoplasma da célula. Estruturalmente, as isoformas de *splicing* de OsaLeg1 não parecem afetar drasticamente a funcionalidade das legumaínas em termos de especificidades para reconhecimento do substrato, mantendo uma estrutura tridimensional bastante similar à forma canônica. Os transcritos de OsaLeg2 não apresentam alterações significativas, sendo que nas situações analisadas a isoforma 2.2 nunca chega a ser mais expressa do que a isoforma 2.1, apesar de perceber-se leves alterações nas suas proporções relativas. A isoforma 2.2 é desprovida de um segmento carboxi-terminal, originando uma proteína pré-ativada, hipoteticamente sem a necessidade de contato com pH ácido para sua ativação. Entretanto, mais estudos são necessários para demonstrar de fato a funcionalidade destas isoformas de legumaínas.

Em conclusão, as legumaínas de arroz são amplamente conservadas filogeneticamente e estruturalmente, quando comparadas com outros genes de legumaínas previamente descritos, mantendo inclusive uma semelhança de estrutura terciária com as legumaínas animais. Assim como em outras plantas, as legumaínas de arroz subdividem-se em agrupamentos filogenéticos um pouco distintos que têm correlação com seus perfis transcricionais, em diferentes situações fisiológicas, tempo de desenvolvimento da planta, ou variações ambientais. Embora os ramos filogenéticos apresentem um bom suporte para a manutenção da classificação “semente/vegetativa”, deve ficar claro que as legumaínas

não são expressas somente nestes tecidos específicos. Assim como demonstrado aqui, as legumaínas foram detectadas em todos os tecidos analisados, com diferentes quantidades de mRNA, podendo refletir diferentes especificidades seus pelos substratos. Assim como, a detecção de diferentes isoformas para um mesmo transcrito de legumaína pode sugerir a existência de neofuncionalizações e diferentes especificidades por substratos, refletindo diretamente no modo de atuação e funções destas proteases.

3. Fitocistatinas bifuncionais em plantas, a importância da fitocistatina carboxi-estendida de arroz

Atualmente, o conhecimento que existe acerca das fitocistatinas estabelece que a função melhor caracterizada destas proteínas é a inibição de proteases cisteínicas do tipo papaínas (família C1A) (Hwang *et al.*, 2009; Benchabane *et al.*, 2010; Diaz-Mendoza *et al.*, 2014; Kunert *et al.*, 2015). Entretanto, existe um estudo *in vitro* que descreve a possível interação das fitocistatinas carboxi-estendidas com uma nova família de proteases, as legumaínas (C13). Este estudo demonstra a habilidade que o domínio carboxi-terminal tem de inibir a legumaína humana, além de inibir a atividade de legumaínas em extratos proteicos de cevada, através do domínio inibitório SNSL (Martinez *et al.*, 2007). Contudo, este foi o único trabalho demonstrando esta nova função das fitocistatinas, o que eleva a necessidade de mais estudos funcionais a fim de aumentar o conhecimento sobre as proteases e seus inibidores endógenos específicos. Outros estudos avaliando fitocistatinas carboxi-estendidas, apenas levaram em consideração seu papel inibitório contra papaínas, sem considerar as legumaínas (Misaka *et al.*, 1996; Gianotti *et al.*, 2006; Wang *et al.*, 2008; Hwang *et al.*, 2009), além da tentativa de cristalização de uma fitocistatina carboxi-estendida em um complexo com papaínas, apenas reafirmando a relação de interação entre o domínio cistatina (N-terminal) e as proteases da família C1A, sem adicionar novas informações sobre a extensão carboxi-terminal (Chu *et al.*, 2011; Cheng *et al.*, 2014).

Em arroz, existe apenas um gene de fitocistatina carboxi-estendida, a OcXII. A fim de estabelecer a funcionalidade da região carboxi-terminal da OcXII em arroz, um peptídeo recombinante correspondente à esta região foi produzido em bactéria. Podemos detectar a capacidade deste peptídeo em inibir de forma específica a atividade de proteases do tipo legumaínas, sem interferir na atividade das papaínas em extratos proteicos de folhas de arroz. Conforme observamos no modelo tridimensional da proteína OcXII, a

região amino-terminal encontra-se bem separada da região carboxi-terminal. Isso tudo indica um claro potencial de bifuncionalidade para este inibidor de proteases cisteínicas, atuando em proteases C1A através do domínio N-terminal e em proteases C13 com o domínio C-terminal. De fato, o domínio amino-terminal de uma fitocistatina estendida já havia sido previamente caracterizado como inibidor de proteases do tipo papaínas (Cheng *et al.*, 2014). Diferentemente, algumas cistatinas animais possuem estas duas funções concentradas em proteínas com um único domínio, apenas com uma pequena separação espacial entre os dois sítios inibitórios (Alvarez-Fernandez *et al.*, 1999). Nas plantas, as fitocistatinas carboxi-estendidas originaram-se de forma que estas funções inibitórias ficaram separadas em dois domínios independentes, e todas as plantas analisadas até o momento possuem pelo menos uma cópia deste gene.

Após a caracterização da família das fitocistatinas em arroz, uma questão remanescente é a elucidação dos papéis funcionais da única fitocistatina carboxi-estendida no genoma de arroz, e suas implicações na regulação das proteases do tipo legumaínas para a planta. Para analisar de forma mais detalhada o papel funcional da OcXII, foi realizado seu silenciamento gênico em arroz. Com uma redução de aproximadamente 98% no nível dos transcritos de OcXII, este silenciamento foi mantido por pelo menos três gerações de plantas de forma bastante específica. A diminuição considerável de OcXII nas plantas, leva a um aumento significativo na atividade das proteases cisteínicas, tanto papaínas quanto legumaínas, corroborando *in vivo* a hipótese de inibição bifuncional por fitocistatinas carboxi-estendidas. Apesar destas atividades de proteases alteradas, não existem outras alterações drásticas visíveis nas plantas adultas silenciadas, em condições ambientais normais. Entretanto, quando as plantas são submetidas a situações mais críticas, como em presença de quitosana, mimetizando uma infecção fúngica, há uma detecção ainda maior da atividade de proteases do tipo legumaínas. As legumaínas são importantes em processos que envolvem senescência e morte celular programada em resposta à infecções (Hatsugai *et al.*, 2004; Hara-Nishimura and Hatsugai, 2011; Pierre *et al.*, 2014). Assim, visivelmente há um envolvimento específico entre OcXII e o controle da atividade de legumaínas em respostas a infecções por fungos.

O silenciamento gênico de OcXII altera o equilíbrio da atividade das proteases, influenciando na fisiologia das plantas de arroz, principalmente no controle da germinação e crescimento inicial das plântulas, uma vez que OcXII é a terceira fitocistatina mais

expressa em arroz. Inicialmente a germinação das sementes de arroz ocorre com emissão de radícula de forma similar entre as plantas silenciadas e plantas WT (*wild-type*). Contudo, em estágios mais avançados de crescimento, aproximadamente com 8 dias, as plantas silenciadas estão mais desenvolvidas que as plantas WT. Da mesma forma, quando apenas um dos dois genes estendidos de *Arabidopsis* é nocauteado, as plantas apresentam um fenótipo de crescimento ligeiramente acelerado (Hwang *et al.*, 2009). O fenótipo oposto é encontrado quando uma das cópias de fitocistatinas estendidas de *Arabidopsis* é superexpresso, apresentando um atraso na germinação e crescimento da plântula (Hong *et al.*, 2007; Hwang *et al.*, 2009). Porém, os estudos citados acima avaliam apenas a atividade de proteases do tipo papaínas, desconsiderando a capacidade bifuncional e os efeitos na inibição de legumaínas. Os dados de nosso estudo indicam uma correlação direta entre as plantas silenciadas, as taxas de crescimento das plântulas e as atividades de ambas as famílias de proteases cisteínicas: papaínas e legumaínas. Naturalmente, as plantas OcXII silenciadas apresentam níveis elevados de atividade de legumaínas e papaínas, mas enquanto as legumaínas permanecem aumentadas durante toda a germinação, as papaínas possuem um pico de atividade deslocado 12h antes nas plantas silenciadas. Esta desregulação fisiológica altera os níveis de expressão dos genes envolvidos em germinação, incluindo outras fitocistatinas, legumaínas e papaínas, provavelmente em uma tentativa de estabilizar o metabolismo de proteólise desregulado.

Baseando-se nestes experimentos e resultados, é possível propor um modelo que estabeleça a importância de fitocistatinas carboxi-estendidas, como OcXII, no processo de germinação e crescimento da plântula. No modelo que sugerimos, a fitocistatina bifuncional é capaz de inibir tanto legumaínas quanto papaínas e alterar as taxas nas quais os nutrientes se tornam disponíveis para a planta, levando a um crescimento inicial mais rápido. Isto tudo, porque, fitocistatinas tem um duplo poder de inibição da atividade das proteases cisteínicas C1A, direta e indiretamente. A forma indireta de inibição ocorre justamente através da inibição das legumaínas, que são as proteases responsáveis pela ativação de inúmeras papaínas. Desta forma, o efeito inibitório potencializa-se, fazendo com que a maior atividade das proteases C13 desencadeie uma atividade ainda maior das proteases C1A, que acabam não sendo eficientemente inibidas pela ausência da fitocistatina carboxi-estendida.

A atividade de OcXII nas sementes de arroz foi elucidada através de estudos com o promotor deste gene fusionado ao gene repórter *gus*. A expressão do mesmo foi detectada na região do escutelo das sementes secas, e após o início e o progresso da germinação esta expressão vai diminuindo, assim como detectado nas análises de expressão gênica das fitocistatinas de arroz. Contudo, OcXII tem expressão gênica controlada por ABA, assim, quando as sementes de arroz são tratadas com este fitormônio, há um atraso na germinação das sementes (resposta característica das plantas ao ABA), mas ao mesmo tempo a detecção de OcXII na região do escutelo das sementes permanece elevada. As proteases cisteínicas estão entre os maiores grupos responsáveis pela hidrólise de outras proteínas, e durante a germinação estas proteases são sintetizadas nas camadas do escutelo e aleurona da semente, em resposta à sinalização por giberelina (contrária a ABA), para que ocorra a germinação (Zhang and Jones, 1996; Mikkonen *et al.*, 1996). Sendo assim, a presença da fitocistatina bifuncional na região do escutelo das sementes, fazendo a interface entre embrião-endosperma, torna-se extremamente importante para a regulação da proteólise endógena necessária à germinação.

Além das atividades já descritas para OcXII na germinação e crescimento de plantas sob condições fisiológicas normais, quando as plantas encontram-se em situações adversas, a atividade do inibidor bifuncional pode ser ainda mais necessária para a sobrevivência do organismo. Por exemplo, a alcalinização é um fator desfavorável para a germinação e crescimento de arroz, afetando diversas vias metabólicas, principalmente em raízes, o que é transmitido para o desenvolvimento do restante da plântula (Wang *et al.*, 2012; Lv *et al.*, 2013). Ao mesmo tempo em que o estresse alcalino reduz as taxas de germinação, ele mantém OcXII altamente ativa na região do escutelo das sementes por períodos prolongados. Além disso, OcXII possui um papel importante na germinação e crescimento de plântulas sob condições alcalinas. As plantas silenciadas para o gene OcXII apresentam um comprometimento do crescimento, com fenótipos mais severos de desenvolvimento das raízes, que provavelmente dificultam a captação de água e nutrientes, levando à um desenvolvimento comprometido da parte aérea das plantas também. Desta forma, a expressão de OcXII no escutelo da semente em germinação é importante para a manutenção do metabolismo da planta e taxas adequadas de crescimento, regulando as atividades proteolíticas, principalmente sob condições ambientais adversas.

Em conclusão, este estudo demonstra a caracterização geral da família das fitocistatinas em plantas, revelando informações importantes sobre cada um dos 12 genes presentes no genoma de arroz, a fim de direcionar estudos futuros com a intenção de especificar mais à fundo a atividade das fitocistatinas, principalmente estudando as proteases-alvo específicas de cada membro da família. Particularmente importante, o papel bifuncional da fitocistatina carboxi-estendida de arroz, OcXII foi estabelecido e pode ser investigado em outras plantas, uma vez que estes genes são evolutivamente bem conservados entre o reino vegetal e também devem ter a capacidade de inibir papaínas e legumaínas. A utilização do arroz como planta modelo foi de fundamental importância para a obtenção de resultados mais claros e diretamente relacionados com a função de OcXII, uma vez que este gene está presente em cópia única no genoma de arroz, portanto não deixando dúvidas sobre possíveis mecanismos compensatórios da atividade deste gene. De forma geral, alterando o balanço de genes reguladores de proteases em plantas, inúmeras vias fisiológicas serão alteradas, as quais ainda não são amplamente estudadas, e isto abre um vasto campo de estudos para a compreensão dos mecanismos relacionados com o desenvolvimento das plantas e respostas fisiológicas.

Capítulo VII

- Perspectivas-

No presente trabalho, nós demonstramos que as fitocistatinas de arroz são evolutivamente conservadas, mantendo uma relação de homologia com as demais fitocistatinas e plantas e cistatinas animais. Os perfis de expressão gênica detectados são bastante variados, sugerindo uma ampla diversidade de funções e proteases-alvo, para cada fitocistatina. Isto amplia a necessidade de mais estudos caracterizando também a relação dos inibidores e das proteases específicas com as quais eles interagem em diferentes momentos. Algumas fitocistatinas, por exemplo, aparentemente não são funcionais, outras parecem ter uma atividade redundante, enquanto outras ainda parecem ser as mais relevantes em termos funcionais. Apenas a caracterização específica de cada uma destas proteínas com suas especificidades inibitórias sobre determinadas proteases poderá demonstrar sua real funcionalidade, revelando um longo caminho a ser percorrido para que as rotas metabólicas e as proteases envolvidas sejam estabelecidas. Tal conhecimento poderá ser aplicado futuramente na melhoria de plantas com interesse agrônômico, como o próprio arroz, além de estender este conhecimento para outras plantas.

A demonstração de que a região C-terminal da OcXII é um bom inibidor de proteases do tipo legumaínas, que não interage com as proteases do tipo papaínas, é uma informação importante que pode ser utilizada para a investigação detalhada de vias proteolíticas em plantas. Adicionalmente o produto gênico completo da fitocistatina carboxi-estendida de arroz, é um inibidor bifuncional, que é capaz de inibir tanto papaínas quanto legumaínas, o que ainda é bastante subavaliado, uma vez que só tem-se considerado a inibição de fitocistatinas sobre papaínas. As interações das fitocistatinas carboxi-estendidas precisam ser profundamente avaliadas no contexto da planta, considerando a inibição de ambas as proteases-alvo, e quais seriam estas proteases de fato.

Inúmeros estudos ainda devem ser realizados para a melhor compreensão dos inibidores de proteases cisteínicas, levando em consideração as diferentes especificidades inibitórias das fitocistatinas, em relação à quais proteases elas são capazes de inibir. Qual a função que as proteases inibidas exercem na planta? A existência de diferentes formas de *splicing* das proteases como as legumaínas é de fato relevante, originando diferentes proteínas para cada situação? Como ocorre a interação tridimensional das fitocistatinas estendidas com as legumaínas? Existe diferença na eficiência de inibição das proteases pelas cistatinas animais ou pelas fitocistatinas estendidas de plantas, resultando da separação do domínio inibitório em plantas? Tais conhecimentos refletirão diretamente na

compreensão das vias proteolíticas dos organismos, as quais ainda são bastante obscuras, necessitando de mais estudos para elucidar quais proteases interagem com substratos específicos e quais os efeitos destas interações para os organismos. Uma vez que estes mecanismos se tornem cada vez mais claros, há uma importante aplicação biotecnológica tanto para os genes de fitocistatinas, atuando como inibidores, quanto para os genes das proteases-alvo, nas plantas. A possibilidade de obter-se plantas mais tolerantes, resistentes ou ainda com melhor desenvolvimento, é uma aplicação bastante promissora dos estudos de mecanismos proteolíticos e inibitórios em plantas.

Capítulo VIII

- Bibliografía -

Abe K, Emori Y, Kondo H, Arai S, Suzuki K. 1988. The NH₂-terminal 21 Amino Acid Residues Are Not Essential for the Papain-inhibitory Activity of Oryzacystatin , a Member of the Cystatin Superfamily. *The Journal of biological chemistry* **263**, 7655–7659.

Abe K, Emori Y, Kondo H, Suzuki K, Arai S. 1987*a*. Molecular cloning of a cysteine proteinase inhibitor of rice (oryzacystatin). Homology with animal cystatins and transient expression in the ripening process of rice. *Journal of Biological Chemistry* **262**, 16793–16797.

Abe K, Kondo H, Arai S. 1987*b*. Purification and characterization of a rice cysteine proteinase inhibitor. *Agricultural and biological chemistry* **51**, 2763–2768.

Abe Y, Shirane K, Yokosawa H, Matsushita H, Mitta M, Kato I, Ishiis S. 1993. Asparaginyl Endopeptidase of Jack Bean Seeds. *Biological Chemistry* **268**, 3525–3529.

Abrahamson M, Alvarez-Fernandez M, Nathanson C-M. 2003. Cystatins. *Biochemical Society symposium* **99**, 179–199.

Agrawal GK, Rakwal R, Tamogami S, Yonekura M, Akihiro K, Hikaru S. 2002. Chitosan activates defense/stress response (s) in the leaves of *Oryza sativa* seedlings. *Plant physiology and biochemistry* **40**, 1061–1069.

Alvarez-Fernandez M, Barrett a J, Gerhartz B, Dando PM, Ni J, Abrahamson M. 1999. Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. *The Journal of biological chemistry* **274**, 19195–203.

Arai S, Matsumoto I, Abe K. 1998. Phytocystatins and their target enzymes: from molecular biology to practical application: a review. *Journal of food biochemistry* **22**, 287–299.

Arai S, Matsumoto I, Emori Y, Abe K. 2002. Plant seed cystatins and their target enzymes of endogenous and exogenous origin. *Journal of agricultural and food chemistry* **50**, 6612–7.

Arai S, Watanabe H, Kondo H. 1991. Papain activity of oryzacystatin, a rice seed cysteine proteinase inhibitor, depends on the central Gln-Val-Val-Ala-Gly region conserved among cystatin superfamily members. *Journal of biochemistry* **109**, 294–298.

Barrett AJ. 1981. Cystatin, the egg white inhibitor of cysteine proteinases. *Methods in enzymology* **80**, 771–778.

Barrett AJ. 1987. The cystatins: a new class of peptidase inhibitors. *Trends in Biochemical Sciences* **12**, 193–196.

Becker C, Shutov a D, Nong VH, Senyuk VI, Jung R, Horstmann C, Fischer J, Nielsen NC, Müntz K. 1995. Purification, cDNA cloning and characterization of proteinase B, an asparagine-specific endopeptidase from germinating vetch (*Vicia sativa* L.) seeds. *European journal of biochemistry / FEBS* **228**, 456–62.

Beers EP, Jones AM, Dickerman AW. 2004. The S8 serine, C1A cysteine and A1 aspartic protease families in Arabidopsis. *Phytochemistry* **65**, 43–58.

Beers EP, Woffenden BJ, Zhao C. 2000. Plant proteolytic enzymes: possible roles during programmed cell death. *Plant molecular biology* **44**, 399–415.

Belenghi B, Acconcia F, Trovato M, Perazzolli M, Bocedi A, Polticelli F, Ascenzi P, Delledonne M. 2003. AtCYS1, a cystatin from Arabidopsis thaliana, suppresses hypersensitive cell death. *European Journal of Biochemistry* **270**, 2593–2604.

Benchabane M, Schlüter U, Vorster J, Goulet M-C, Michaud D. 2010. Plant cystatins. *Biochimie* **92**, 1657–66.

Björk I, Pol E, Raub-Segall E, Abrahamson M, Rowan a D, Mort JS. 1994. Differential changes in the association and dissociation rate constants for binding of cystatins to target proteinases occurring on N-terminal truncation of the inhibitors indicate that the interaction mechanism varies with different enzymes. *The Biochemical journal* **299** (Pt 1, 219–25.

Bode W, Engh R, Musil D, Thiele U, Huber R, Karshikov A, Brzin J, Kos J, Turk V. 1988. The 2.0 Å X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. *The EMBO journal* **7**, 2593.

Bonneau L, Ge Y, Drury GE, Gallois P. 2008. What happened to plant caspases? *Journal of experimental botany* **59**, 491–9.

Carrillo L, Martínez M, Ramessar K, Cambra I, Castañera P, Ortego F, Díaz I. 2011. Expression of a barley cystatin gene in maize enhances resistance against phytophagous mites by altering their cysteine-proteases. *Plant cell reports* **30**, 101–12.

Chen J, Dando PM, Rawlings ND, Brown MA, Young NE, Stevens RA, Hewitt E, Watts C, Barrett AJ. 1997. Cloning, Isolation, and Characterization of Mammalian Legumain, an Asparaginyl Endopeptidase. *The Journal of Biological Chemistry* **272**, 8090–8098.

Chen MS, Johnson B, Wen L, Muthukrishnan S, Kramer KJ, Morgan TD, Reeck GR. 1992. Rice cystatin: bacterial expression, purification, cysteine proteinase inhibitory activity, and insect growth suppressing activity of a truncated form of the protein. *Protein expression and purification* **3**, 41–9.

Chen J, Rawlings N, Stevens R. 1998. Identification of the active site of legumain links it to caspases, clostripain and gingipains in a new clan of cysteine endopeptidases. *FEBS letters* **441**, 361–365.

Cheng M-L, Tzen JTC, Shyu DJH, Chou W-M. 2014. Functional characterization of the N-terminal and C-terminal domains of a sesame group II phytocystatin. *Botanical Studies* **55**, 18.

Chu M-H, Liu K-L, Wu H-Y, Yeh K-W, Cheng Y-S. 2011. Crystal structure of tarocystatin-papain complex: implications for the inhibition property of group-2 phytocystatins. *Planta* **234**, 243–54.

Diaz-Mendoza M, Velasco-Arroyo B, Gonzalez-Melendi P, Martinez M, Diaz I. 2014. C1A cysteine protease-cystatin interactions in leaf senescence. *Journal of experimental botany*, 1–9.

Donnison IS, Gay AP, Thomas H, Edwards KJ, Edwards D, James CL, Thomas AM, Ougham HJ. 2007. Modification of nitrogen remobilization, grain fill and leaf senescence in maize (*Zea mays*) by transposon insertional mutagenesis in a protease gene. *The New phytologist* **173**, 481–94.

Dutt S, Singh VK, Marla SS, Kumar A. 2010. In silico analysis of sequential, structural and functional diversity of wheat cystatins and its implication in plant defense. *Genomics, proteomics & bioinformatics / Beijing Genomics Institute* **8**, 42–56.

Fischer J, Becker C, Hillmer S, Horstmann C, Neubohn B, Schlereth A, Senyuk V, Shutov A, Müntz K. 2000. The families of papain- and legumain-like cysteine proteinases from embryonic axes and cotyledons of *Vicia* seeds: developmental patterns, intracellular localization and functions in globulin proteolysis. *Plant molecular biology* **43**, 83–101.

Fossum K, Whitaker JR. 1968. Ficin and Papain from Chicken Egg White. *Archives of Biochemistry and Biophysics* **125**, 367–375.

Gianotti A, Rios WM, Soares-Costa A, Nogaroto V, Carmona AK, Oliva ML V, Andrade SS, Henrique-Silva F. 2006. Recombinant expression, purification, and functional analysis of two novel cystatins from sugarcane (*Saccharum officinarum*). *Protein expression and purification* **47**, 483–9.

Grudkowska M, Zagdańska B. 2004. Multifunctional role of plant cysteine proteinases. *Acta biochimica Polonica* **51**, 609–24.

Gruis D, Schulze J, Jung R. 2004. Storage protein accumulation in the absence of the vacuolar processing enzyme family of cysteine proteases. *The Plant Cell* **16**, 270–290.

Hara-Nishimura I, Hatsugai N. 2011. The role of vacuole in plant cell death. *Cell death and differentiation* **18**, 1298–304.

Hara-Nishimura I, Hatsugai N, Nakaune S, Kuroyanagi M, Nishimura M. 2005. Vacuolar processing enzyme: an executor of plant cell death. *Current opinion in plant biology* **8**, 404–8.

Hara-Nishimura I, Inoue K, Nishimura M. 1991. A unique vacuolar processing enzyme responsible for conversion of several proprotein precursors into the mature forms. *FEBS letters* **294**, 89–93.

Hara-Nishimura I, Kinoshita T, Hiraiwa N, Nishimura M. 1998. Vacuolar processing enzymes in protein-storage vacuoles and lytic vacuoles. *Journal of Plant Physiology* **152**, 668–674.

Hara-Nishimura I, Nishimura M. 1987. Proglobulin processing enzyme in vacuoles isolated from developing pumpkin cotyledons. *Plant physiology* **85**, 440–5.

Hara-Nishimura I, Shimada T, Hiraiwa N, Nishimura M. 1995. Vacuolar Processing Enzyme Responsible for Maturation of Seed Proteins. *Journal of Plant Physiology* **145**, 632–640.

Hara-Nishimura I, Takeuchi Y, Nishimura M. 1993. Molecular characterization of a vacuolar processing enzyme related to a putative cysteine proteinase of *Schistosoma mansoni*. *The Plant cell* **5**, 1651–9.

Hatsugai N, Kuroyanagi M, Yamada K, Meshi T, Tsuda S, Kondo M, Nishimura M, Hara-Nishimura I. 2004. A plant vacuolar protease, VPE, mediates virus-induced hypersensitive cell death. *Science (New York, N.Y.)* **305**, 855–8.

Hayashi Y, Yamada K, Shimada T, Matsushima R, Nishizawa NK, Nishimura M, Hara-Nishimura I. 2001. A proteinase-storing body that prepares for cell death or stresses in the epidermal cells of *Arabidopsis*. *Plant & cell physiology* **42**, 894–9.

Hiraiwa N, Nishimura M, Hara-Nishimura I. 1999. Vacuolar processing enzyme is self-catalytically activated by sequential removal of the C-terminal and N-terminal propeptides. *FEBS letters* **447**, 213–6.

Hiroto K, Yasufumi E, Keiko A, Koichi S. 1989. Cloning and sequence analysis of the genomic DNA fragment encoding oryzacystatin. *Gene* **81**, 259–265.

Hong JK, Hwang JE, Chung WS, Lee KO, Choi YJ, Gal SW, Park BS, Lim CO. 2008. Expression of a Chinese cabbage cysteine proteinase inhibitor, BrCYS1, retards seed germination and plant growth in transgenic Tobacco plant. *Journal of Plant Biology* **51**, 347–353.

Hong JK, Hwang JE, Lim CJ, et al. 2007. Over-expression of Chinese cabbage phytolectin 1 retards seed germination in *Arabidopsis*. *Plant Science* **172**, 556–563.

Van der Hoorn R a L. 2008. Plant proteases: from phenotypes to molecular mechanisms. *Annual review of plant biology* **59**, 191–223.

Van der Hoorn R a L, Jones JDG. 2004. The plant proteolytic machinery and its role in defence. *Current opinion in plant biology* **7**, 400–7.

Hwang JE, Hong JK, Je JH, Lee KO, Kim DY, Lee SY, Lim CO. 2009. Regulation of seed germination and seedling growth by an *Arabidopsis* phytolectin isoform, AtCYS6. *Plant cell reports* **28**, 1623–32.

Ishii SI. 1994. Legumain: Asparaginyl endopeptidase. *Methods in enzymology* **244**, 604–615.

Julián I, Gandullo J, Santos-Silva LK, Diaz I, Martinez M. 2013. Phylogenetically distant barley legumains have a role in both seed and vegetative tissues. *Journal of experimental botany*.

Jung D-Y, Lee OR, Kim Y-J, Lee J-H, Pulla RK, Sathiyaraj G, Shim J-S, Yang D-C. 2010. Molecular characterization of a cysteine proteinase inhibitor, PgCPI, from *Panax ginseng* C. A. Meyer. *Acta Physiologiae Plantarum* **32**, 961–970.

Kamphuis IG, Drenth J, Baker EN. 1985. Thiol proteases. Comparative studies based on the high-resolution structures of papain and actinidin, and on amino acid sequence information for cathepsins B and H, and stem bromelain. *Journal of molecular biology* **182**, 317–29.

Kato H, Sutoh K, Minamikawa T. 2003. Identification, cDNA cloning and possible roles of seed-specific rice asparaginyl endopeptidase, REP-2. *Planta* **217**, 676–85.

Khanna-Chopra R, Srivalli B, Ahlawat YS. 1999. Drought induces many forms of cysteine proteases not observed during natural senescence. *Biochemical and biophysical research communications* **255**, 324–7.

Kinoshita T, Yamada K, Hiraiwa N, Kondo M, Nishimura M, Hara-Nishimura I. 1999. Vacuolar processing enzyme is up-regulated in the lytic vacuoles of vegetative tissues during senescence and under various stressed conditions. *The Plant journal : for cell and molecular biology* **19**, 43–53.

Koiwa H, D'Urzo MP, Assfalg-Machleidt I, Zhu-Salzman K, Shade RE, An H, Murdock LL, Machleidt W, Bressan R a, Hasegawa PM. 2001. Phage display selection of hairpin loop soyacystatin variants that mediate high affinity inhibition of a cysteine proteinase. *The Plant journal : for cell and molecular biology* **27**, 383–91.

Kondo H, Abe K, Emori Y, Arai S. 1991. Gene organization of oryzacystatin-II, a new cystatin superfamily member of plant origin, is closely related to that of oryzacystatin-I but different from those of animal cystatins. *FEBS letters* **278**, 87–90.

Kondo H, Abe K, Nishimura I, Watanabe H, Emori Y, Arai S. 1990. Two Distinct Cystatin Species in Rice Seeds with Different Specificities against Cysteine Proteinases. *Journal of Biological Chemistry* **265**, 15832–15837.

Konno K, Hirayama C, Nakamura M, Tateishi K, Tamura Y, Hattori M, Kohno K. 2003. Papain protects papaya trees from herbivorous insects: role of cysteine proteases in latex. *The Plant Journal* **37**, 370–378.

Kordis D, Turk V. 2009. Phylogenomic analysis of the cystatin superfamily in eukaryotes and prokaryotes. *BMC evolutionary biology* **9**, 266.

- Kudo N, Nishiyama M, Sasaki H, Abe K, Arai S, Tanokura M.** 1998. Crystallization and preliminary X-ray diffraction studies of a rice cysteine proteinase inhibitor, oryzacystatin-I. *Journal of biochemistry* **123**, 568.
- Kunert KJ, van Wyk SG, Cullis C a., Vorster BJ, Foyer CH.** 2015. Potential use of phytocystatins in crop improvement, with a particular focus on legumes. *Journal of Experimental Botany*, 1–12.
- Kuroyanagi M, Nishimura M, Hara-Nishimura I.** 2002. Activation of Arabidopsis vacuolar processing enzyme by self-catalytic removal of an auto-inhibitory domain of the C-terminal propeptide. *Plant & cell physiology* **43**, 143–51.
- Lee SC, Luan S.** 2012. ABA signal transduction at the crossroad of biotic and abiotic stress responses. *Plant, cell & environment* **35**, 53–60.
- Lepelley M, Amor M Ben, Martineau N, Cheminade G, Caillet V, McCarthy J.** 2012. Coffee cysteine proteinases and related inhibitors with high expression during grain maturation and germination. *BMC Plant Biology* **12**, 31.
- Lv BS, Li XW, Ma HY, Sun Y, Wei LX, Jiang CJ, Liang ZW.** 2013. Differences in growth and physiology of rice in response to different saline-alkaline stress factors. *Agronomy Journal* **105**, 1119–1128.
- Machleidt W, Thiele U, Laber B, Assfalg-Machleidt I, Esterl A, Wiegand G, Kos J, Turk V, Bode W.** 1989. Mechanism of inhibition of papain by chicken egg white cystatin. Inhibition constants of N-terminally truncated forms and cyanogen bromide fragments of the inhibitor. *FEBS letters* **243**, 234–8.
- Margis R, Reis EM, Villeret V.** 1998. Structural and phylogenetic relationships among plant and animal cystatins. *Archives of biochemistry and biophysics* **359**, 24–30.
- Margis-Pinheiro M, Zolet ACT, Loss G, Pasquali G, Margis R.** 2008. Molecular evolution and diversification of plant cysteine proteinase inhibitors: new insights after the poplar genome. *Molecular phylogenetics and evolution* **49**, 349–55.
- Martínez M, Abraham Z, Carbonero P, Díaz I.** 2005. Comparative phylogenetic analysis of cystatin gene families from arabidopsis, rice and barley. *Molecular genetics and genomics* : MGG **273**, 423–32.
- Martinez M, Abraham Z, Gambardella M, Echaide M, Carbonero P, Diaz I.** 2005. The strawberry gene *Cyf1* encodes a phytocystatin with antifungal properties. *Journal of experimental botany* **56**, 1821–9.
- Martinez M, Cambra I, Carrillo L, Diaz-Mendoza M, Diaz I.** 2009. Characterization of the entire cystatin gene family in barley and their target cathepsin L-like cysteine-proteases, partners in the hordein mobilization during seed germination. *Plant physiology* **151**, 1531–45.

Martínez M, Cambra I, González-Melendi P, Santamaría ME, Díaz I. 2012. C1A cysteine-proteases and their inhibitors in plants. *Physiologia plantarum*, 85–94.

Martinez M, Diaz I. 2008. The origin and evolution of plant cystatins and their target cysteine proteinases indicate a complex functional relationship. *BMC evolutionary biology* **8**, 198.

Martinez M, Diaz-Mendoza M, Carrillo L, Diaz I. 2007. Carboxy terminal extended phytocystatins are bifunctional inhibitors of papain and legumain cysteine proteinases. *FEBS letters* **581**, 2914–8.

Mikkonen A, Porali I, Cercos M, Ho TH. 1996. A major cysteine proteinase, EPB, in germinating barley seeds: structure of two intronless genes and regulation of expression. *Plant molecular biology* **31**, 239–54.

Misaka T, Kuroda M, Iwabuchi K, Abe K, Arai S. 1996. Soyacystatin, a novel cysteine proteinase inhibitor in soybean, is distinct in protein structure and gene organization from other cystatins of animal and plant origin. *European journal of biochemistry / FEBS* **240**, 609–14.

Müntz K, Blattner FR, Shutov AD. 2002. Legumains-a family of asparagine-specific cysteine endopeptidases involved in propolypeptide processing and protein breakdown in plants. *Journal of plant physiology* **159**, 1281–1293.

Müntz K, Shutov AD. 2002. Legumains and their functions in plants. *Trends in Plant Science* **7**, 340–344.

Nagata K, Kudo N, Abe K, Arai S. 2000. Three-dimensional solution structure of oryzacystatin-I, a cysteine proteinase inhibitor of the rice, *Oryza sativa* L. japonica. *Biochemistry* **39**, 14753–14760.

Ohtsubo S, Kobayashi H, Noro W, Taniguchi M, Saitoh E. 2005. Molecular cloning and characterization of oryzacystatin-III, a novel member of phytocystatin in rice (*Oryza sativa* L. japonica). *Journal of agricultural and food chemistry* **53**, 5218–24.

Okamoto T, Minamikawa T. 1999. Molecular cloning and characterization of *Vigna mungo* processing enzyme 1 (VmPE-1), an asparaginyl endopeptidase possibly involved in post-translational processing of a vacuolar cysteine endopeptidase (SH-EP). *Plant molecular biology* **39**, 63–73.

Pesquet E. 2012. Plant proteases - from detection to function. *Physiologia plantarum* **145**, 1–4.

Pierre O, Hopkins J, Combiér M, Baldacci F, Engler G, Brouquisse R, Hérouart D, Boncompagni E. 2014. Involvement of papain and legumain proteinase in the senescence process of *Medicago truncatula* nodules. *The New phytologist* **202**, 849–63.

Pirovani CP, da Silva Santiago A, Dos Santos LS, Micheli F, Margis R, da Silva Gesteira A, Alvim FC, Pereira GAG, de Mattos Cascardo JC. 2010. Theobroma cacao cystatins impair Moniliophthora perniciosa mycelial growth and are involved in postponing cell death symptoms. *Planta* **232**, 1485–97.

Richau KH, Kaschani F, Verdoes M, Pansuriya TC, Niessen S, Stüber K, Colby T, Overkleeft HS, Bogyo M, van der Hoorn RAL. 2012. Subclassification and Biochemical Analysis of Plant Papain-like Cysteine Proteases displays Subfamily-specific Characteristics. *Plant physiology* **158**, 1583–1599.

Rojo E, Martín R, Carter C, et al. 2004. VPEgamma exhibits a caspase-like activity that contributes to defense against pathogens. *Current biology* **14**, 1897–906.

Rojo E, Zouhar J, Carter C, Kovaleva V, Raikhel N V. 2003. A unique mechanism for protein processing and degradation in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 7389–94.

Santos-Silva LK, Soares-Costa A, Gerald LTS, Meneghin SP, Henrique-Silva F. 2012. Recombinant expression and biochemical characterization of sugarcane legumain. *Plant physiology and biochemistry : PPB / Societe francaise de physiologie vegetale* **57C**, 181–192.

Senthilkumar R, Cheng C-P, Yeh K-W. 2010. Genetically pyramiding protease-inhibitor genes for dual broad-spectrum resistance against insect and phytopathogens in transgenic tobacco. *Plant biotechnology journal* **8**, 65–75.

Shutov AD, Blattner FR, Kakhovskaya I a, Müntz K. 2012. New aspects of the molecular evolution of legumains, Asn-specific cysteine proteinases. *Journal of plant physiology* **169**, 319–21.

Turk V, Bode W. 1991. The cystatins: protein inhibitors of cysteine proteinases. *FEBS letters* **285**, 213–219.

Turk D, Guncar G, Podobnik M, Turk B. 1998. Revised definition of substrate binding sites of papain-like cysteine proteases. *Biological chemistry* **379**, 137–147.

Turk V, Turk B. 2001. Lysosomal cysteine proteases: facts and opportunities. *The EMBO journal* **20**, 4629–4633.

Urwin PE, Atkinson HJ, McPherson MJ. 1995a. Involvement of the NH₂-terminal region of oryzacystatin-I in cysteine proteinase inhibition. *Protein engineering* **8**, 1303–7.

Urwin P, Atkinson H, Waller D. 1995b. Engineered oryzacystatin I expressed in transgenic hairy roots confers resistance to Globodera pallida. *The Plant Journal* **8**, 121–131.

Wan L, Xia Q, Qiu X, Selvaraj G. 2002. Early stages of seed development in *Brassica napus*: a seed coat-specific cysteine proteinase associated with programmed cell death of the inner integument. *The Plant journal : for cell and molecular biology* **30**, 1–10.

Wang H, Ahan J, Wu Z, Shi D, Liu B, Yang C. 2012. Alteration of nitrogen metabolism in rice variety ‘Nipponbare’ induced by alkali stress. *Plant and Soil* **355**, 131–147.

Wang K-M, Kumar S, Cheng Y-S, Venkatagiri S, Yang A-H, Yeh K-W. 2008. Characterization of inhibitory mechanism and antifungal activity between group-1 and group-2 phytocystatins from taro (*Colocasia esculenta*). *The FEBS journal* **275**, 4980–9.

Wang Y, Zhu S, Liu S, et al. 2009. The vacuolar processing enzyme OsVPE1 is required for efficient glutelin processing in rice. *The Plant journal : for cell and molecular biology* **58**, 606–17.

Watanabe H, Abe K, Emori Y, Hosoyama H. 1991. Molecular cloning and gibberellin-induced expression of multiple cysteine proteinases of rice seeds (oryzains). *Journal of Biological* **266**, 16897–16902.

Yamada K, Nishimura M, Hara-Nishimura I. 2004. The slow wound-response of gammaVPE is regulated by endogenous salicylic acid in *Arabidopsis*. *Planta* **218**, 599–605.

Yamada K, Shimada T, Nishimura M, Hara-Nishimura I. 2005. A VPE family supporting various vacuolar functions in plants. *Physiologia Plantarum* **123**, 369–375.

Yang a H, Yeh KW. 2005. Molecular cloning, recombinant gene expression, and antifungal activity of cystatin from taro (*Colocasia esculenta* cv. Kaosiung no. 1). *Planta* **221**, 493–501.

Zhang N, Jones BL. 1996. Purification and partial characterization of a 31-kDa cysteine endopeptidase from germinated barley. *Planta* **199**, 565–72.

Zhang X, Liu S, Takano T. 2008. Two cysteine proteinase inhibitors from *Arabidopsis thaliana*, AtCYSa and AtCYSb, increasing the salt, drought, oxidation and cold tolerance. *Plant Molecular Biology* **68**, 131–143.

Anexo I

- Artigo científico publicado em colaboração 1 -



The mitochondrial glutathione peroxidase GPX3 is essential for H₂O₂ homeostasis and root and shoot development in rice

Gisele Passaia^a, Leila Spagnolo Fonini^{a,b}, Andréia Caverzan^{a,b}, Douglas Jardim-Messeder^a, Ana Paula Christoff^{a,b}, Marcos Letaif Gaeta^c, Jorge Ernesto de Araujo Mariath^c, Rogério Margis^b, Marcia Margis-Pinheiro^{a,b,*}

^a Department of Genetics, Federal University of Rio Grande do Sul, RS, Brazil

^b Biotechnology Center, Federal University of Rio Grande do Sul, RS, Brazil

^c Department of Botany, Federal University of Rio Grande do Sul, RS, Brazil

ARTICLE INFO

Article history:

Received 21 January 2013

Received in revised form 11 March 2013

Accepted 17 March 2013

Available online 8 April 2013

Keywords:

Oryza sativa

Gramineae

Rice

GPX

Glutathione peroxidase

Oxidative stress

ABSTRACT

Glutathione (GSH) peroxidases (GPXs: EC 1.11.1.9 and EC1.11.1.12) are non-heme thiol peroxidases that catalyze the reduction of H₂O₂ or organic hydroperoxides to water, and they have been identified in almost all kingdoms of life. The rice glutathione peroxidase (*OsGPX*) gene family is comprised of 5 members spread throughout a range of sub cellular compartments. The *OsGPX* gene family is induced in response to exogenous H₂O₂ and cold stress. In contrast, they are down regulated in response to drought and UV-B light treatments. Transgenic rice plants have been generated that lack mitochondrial *OsGPX3*. These *GPX3s* plants showed shorter roots and shoots compared to non-transformed (NT) plants, and higher amounts of H₂O₂ mitochondrial release were observed in the roots of these plants cultivated under normal conditions. This accumulation of H₂O₂ is positively associated with shorter root length in *GPX3s* plants compared to NT ones. Moreover, *GPX3* promoter analysis indicated that it is mainly expressed in root tissue. These results suggest that silencing the mitochondrial *OsGPX3* gene impairs normal plant development and leads to a stress-induced morphogenic response via H₂O₂ accumulation.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Whereas animals can move away from environmental adversities such as water deficit, high temperature or heavy metal pollution beyond others, plants are sessile organisms. As a result, the anti-oxidant and repair system of plants is considerably enhanced and diversified compared to bacterial or animal systems because plants also have to cope with a second source of reactive oxygen species (ROS) from chloroplasts, which is beyond the mitochondrial electron transport chain [1]. The most abundant ROS produced during stress is hydrogen peroxide (H₂O₂). This compound is not only an oxidant but also a signal generated from superoxide ions through the action of superoxide dismutases (SOD) [2]. It is important that H₂O₂ levels are properly controlled under stress or normal conditions, and this is regulated by the orchestrated participation of a range of different enzymatic and

non-enzymatic antioxidant systems. The non-enzymatic antioxidant compounds are represented by carotenoids, tocopherols, reduced glutathione (GSH) and ascorbate (AsA), while the enzymatic systems rely on superoxide dismutases (SOD), catalases (CAT), ascorbate peroxidases (APX), peroxiredoxins (PRXs) and glutathione peroxidases (GPXs) [3]. Glutathione (GSH) Peroxidases (GPXs: EC 1.11.1.9 and EC1.11.1.12) are non-heme thiol peroxidases that catalyze the reduction of H₂O₂ or organic hydroperoxides to water, and they have been identified in almost all kingdoms of life [4]. In animals, GPXs are selenoproteins that contain selenocysteine at the catalytic site, whereas the plant enzymes do not bind selenium. The active site is composed of a catalytic triad formed by selenocysteine/cysteine, glutamine and tryptophan. This difference allows plants to preferentially reduce thioredoxin instead of GSH, which is opposite the preference of the enzymes in animals [4,5]. The rice glutathione peroxidase (*OsGPX*) gene family consists of 5 members. Phylogenetic analysis has predicted that GPX is localized to the mitochondria (GPX1 and GPX3), chloroplasts (GPX4) and cytosol/endoplasmic reticulum (GPX2 and GPX5) [6].

There is increasing evidence that GPXs are able to interact with other proteins, conferring peroxide-induced oxidation in addition to their well-established ROS scavenging function. For example, in yeast, GPX3 activates the Yeast Activation Protein 1

* Corresponding author at: Universidade Federal do Rio Grande do Sul, Avenida Bento Gonçalves, 9500, CEP: 91591-970, Porto Alegre, RS, Brasil.
Tel.: +55 51 3308 9814; fax: +55 51 3308 7311.

E-mail addresses: marcia.margis@ufrgs.br, marcia.p.margis@gmail.com (M. Margis-Pinheiro).

Anexo II

- Artigo científico publicado em colaboração 2 –



Identification of potential miRNAs and their targets in *Vriesea carinata* (Poales, Bromeliaceae)



Frank Guzman^a, Mauricio Pereira Almerão^b, Ana Paula Korbes^a, Ana Paula Christoff^a,
Camila Martini Zanella^a, Fernanda Bered^a, Rogério Margis^{a,b,*}

^a PPGGBM at Federal University of Rio Grande do Sul – UFRGS, Porto Alegre, RS, Brazil

^b Centre of Biotechnology and PPGBCM, Laboratory of Genomes and Plant Population, Federal University of Rio Grande do Sul – UFRGS, Porto Alegre, RS, Brazil

ARTICLE INFO

Article history:

Received 8 September 2012

Received in revised form 24 April 2013

Accepted 23 May 2013

Available online 2 June 2013

Keywords:

Vriesea carinata

miRNAs

High-throughput sequencing

Stress response

ABSTRACT

The miRNAs play important roles in regulation of gene expression at the post-transcriptional level. A small RNA and RNA-seq of libraries were constructed to identify miRNAs in *Vriesea carinata*, a native bromeliad species from Brazilian Atlantic Rainforest. Illumina technology was used to perform high throughput sequencing and data was analyzed using bioinformatics tools. We obtained 2,191,509 mature miRNAs sequences representing 54 conserved families in plant species. Further analysis allowed the prediction of secondary structures for 19 conserved and 16 novel miRNAs. Potential targets were predicted from pre-miRNAs by sequence homology and validated using RTqPCR approach. This study provides the first identification of miRNAs and their potential targets of a bromeliad species.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

MicroRNAs (miRNAs) are small non-coding regulatory RNAs widely found in unicellular and multicellular organisms that act as regulators of gene expression at the post-transcriptional level on genes containing miRNA binding sites [1]. Mature miRNAs are single-stranded RNA molecules of approximately 21 nucleotides (nt) in length processed from a precursor molecule (pre-miRNA) [2]. To regulate protein-coding genes the mature miRNA binds in the mRNA target site leading to mRNA degradation or translation repression [3]. In plants, miRNAs have diverse biological functions and are involved in the regulation of optimal growth and development, as well as other physiological processes, including abiotic and biotic stress responses [4]. Several studies showed that many miRNAs are conserved across different plant families [5]. However, it was also reported species and family specific miRNAs that are expressed in low levels and probably have evolved recently [6].

Bromeliaceae family, with 3248 species distributed in 58 genera [7], is an example of a large and well described adaptive radiation

of plant families in the Neotropics. The family is composed of terrestrial xerophytes and both facultative and obligatory epiphytes species which acquired, throughout their evolution, interesting and different adaptive mechanisms such as central tanks, CAM photosynthetic pathway, and bear absorptive trichomes, characteristics that allowed them to occupy a wide range of habitats [8,9]. Consequently, Bromeliaceae constitutes one of the most ecologically diverse and species-rich clades of flowering plants native to the New World [10]. In Brazil, the Atlantic Rainforest is considered one of the main centers of diversity and endemism of Bromeliaceae, showing 31 genera and 803 species of which 10 genera and 653 species are endemic [11]. *Vriesea carinata* is an epiphytic or terrestrial species distributed along the Brazilian Atlantic Rainforest. As a typical species of this biome [12], *V. carinata* is an interesting model for studying the expression of miRNAs in Bromeliaceae. The first step to study the expression of miRNAs is to identify miRNAs and its targets in different natural conditions. For this purpose, we performed a high-throughput sequencing analysis (Solexa technology) of small RNAs (sRNAs) from the endemic Brazilian Atlantic Rainforest species *V. carinata*.

2. Materials and methods

2.1. Plant material and RNA isolation

Total RNA was isolated from *V. carinata* leaves using Trizol reagent (Invitrogen, CA, USA), according to manufacturer's protocol. The RNA quality was evaluated by electrophoresis on a 1%

Abbreviations: miRNA, microRNA; sRNA, small RNA; pre-miRNA, microRNA precursor; MFEI, minimal folding energy index; nt, nucleotides; vca, *Vriesea carinata*.

* Corresponding author at: Centre of Biotechnology and PPGBCM, Laboratory of Genomes and Plant Population, Building 43431, Federal University of Rio Grande do Sul – UFRGS, P.O. Box 15005, CEP 91501-970, Porto Alegre, RS, Brazil.
Tel.: +55 51 33087766; fax: +55 51 33087309.

E-mail addresses: rogerio.margis@ufrgs.br, rogerio.margis@gmail.com (R. Margis).

Anexo III

- Artigo científico publicado em colaboração 3 –

isomiRID: a framework to identify microRNA isoforms

Luiz Felipe Valter de Oliveira¹, Ana Paula Christoff¹ and Rogerio Margis^{1,2,*}

¹Genetics and Molecular Biology Graduation Program and ²Department of Biophysics, Center of Biotechnology, Universidade Federal do Rio Grande do Sul – UFRGS, Avenida Bento Goncalves 9500, Predio 43431, Sala 213, Porto Alegre, Brasil, CEP 91501-970

Associate Editor: Ivo Hofacker

ABSTRACT

Summary: MicroRNAs (miRNAs) have been extensively studied owing to their important regulatory roles in genic expression. An increasingly number of reports are performing extensive data mining in small RNA sequencing libraries to detect miRNAs isoforms and also 5' and 3' post-transcriptional nucleotide additions, as well as edited miRNAs sequences. A ready to use pipeline, isomiRID, was developed to standardize and automatize the search for miRNAs isoforms in high-throughput small RNA sequencing libraries.

Availability: isomiRID is a command line Python script available at <http://www.ufrgs.br/RNAi/isomiRID/>.

Contact: rogerio.margis@ufrgs.br

Supplementary information: Supplementary Data are available at *Bioinformatics* online.

Received on January 31, 2013; revised on July 16, 2013; accepted on July 17, 2013

1 INTRODUCTION

Several high-throughput small RNA (sRNA) sequencing projects have increased the discovery and number of small non-coding RNAs related to gene expression regulation. MicroRNAs (miRNAs) are endogenous sRNAs with 19–24 nt in length (Bartel, 2004). To regulate protein-coding genes, mature miRNA binds in the mRNA target sites leading to their degradation or repression of translation (Pasquinelli, 2012; Voinnet, 2009).

Recently, reports have demonstrated the existence of isomiRs (Cloonan *et al.*, 2011; Guo and Lu, 2010; Körbes *et al.*, 2012). The isomiR classification relies in three major categories: 5', 3' and polymorphic isomiRs, with the sub classification of 5' and 3' isomiRs into templated or non-templated modifications, according to the miRNA precursor sequence (Neilsen *et al.*, 2012). Furthermore, efforts have been made to identify and understand the biological processes where non-templated nucleotides are added in the 3' end of mature miRNAs, altering the miRNA stability and efficiency (Burroughs *et al.*, 2010; Ebhardt *et al.*, 2009; Lu *et al.*, 2009; Wyman *et al.*, 2011). Nucleotide additions at the 5' end have also been reported altering the miRNA seed region and consequently its functionality (Bizuyehu *et al.*, 2012; Ebhardt *et al.*, 2010). Several isomiR variants have been discovered with deep sequencing technologies. Some of these variations may have biological origin and functions, once the modifications were seen repeatedly at the same sites with higher frequencies than random errors (Ebhardt *et al.*, 2009).

*To whom correspondence should be addressed.

To improve and automate the search for isomiRs in sRNA sequenced libraries, we developed a simplified workflow. isomiRID allows the identification of 5', 3' and polymorphic isomiRs based on the canonical miRNA known sequence and also from other regions on the same miRNA precursor. Additionally, the program can also identify non-templated 5' or 3' end variations by mapping the sRNAs in the known pre-miRNAs. This framework is a simple method to identify and compare isomiRs abundance in different sequencing libraries, providing an initial view for biological relevant isomiRs. We also highlight that this pipeline could be applied to study isomiRs from plants, animals or any other living organisms.

2 METHODS

The isomiRID workflow (Fig. 1) has four main steps:

The first step generates the file of sRNAs with perfect matches on the pre-miRNA (R0 = round 0) and create the dataset for the subsequent analysis with the pre-miRNA unmapped reads. Optionally, it is possible to use a reference genome (or transcriptome), to filter the unmapped sRNAs from other genomic regions.

On the second step, sRNAs mapping is performed, allowing a single mismatch in the sequence. This first part generates the R1 file, containing sequences with one mismatch in the 5' end (5'MM), in the middle of sequence (MM) and in the 3' end (3'MM). Posteriorly, the small reads, which still do not match with the pre-miRNA sequence, are submitted to trimming rounds in their 5' and 3' ends. Each trim is performed once in a round, where a single 5' or 3' nt is removed for subsequent sRNA mapping on the pre-miRNA reference file. This procedure can be performed N times, according to the number of nucleotide trimmings determined by the user. Reads that match to the pre-miRNA reference after the trimming will be separated in R2, R3...RN files. These files contain a raw table with the mapped sequences, the name of miRNA precursor from which it potentially originated, their lengths, the variant nucleotide(s) and the abundance of such variations in each sequence on the analyzed library(ies) (Supplementary File 1).

Additionally, in a third step, the files with mapped reads can undergo an abundance cutoff filter, to restrict the read values, specified in the configuration file. Finally, the fourth step relies on concatenation of the isomiRs mapping results in a sub folder (MapResults). A graphical output is built based on the mappings with or without abundance cutoff. In graphical outputs, the pre-miRNA sequence is aligned with the miRNA and its isomiRs sequences, also detailing the individual sRNA variation and the abundance of each sequence in different libraries that were analyzed.

3 IMPLEMENTATION

The isomiRID was implemented with python 2.7 and use Bowtie for mapping the reads (Langmead *et al.*, 2009). This framework

Anexo IV

- Artigo científico publicado em colaboração 4 –



The Wall-associated Kinase gene family in rice genomes



Luiz Felipe Valter de Oliveira^{a,b,1}, Ana Paula Christoff^{a,b,1}, Júlio Cesar de Lima^{a,b}, Bruno Comparsi Feijó de Ross^b, Gilberto Sachetto-Martins^c, Marcia Margis-Pinheiro^a, Rogerio Margis^{a,b,d,*}

^a Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Brazil

^b Centro de Biotecnologia e Programa de Pós-Graduação em Biologia Celular e Molecular, Universidade Federal do Rio Grande do Sul, Brazil

^c Departamento de Genética, Universidade Federal do Rio de Janeiro, Brazil

^d Departamento de Biofísica, Universidade Federal do Rio Grande do Sul, Brazil

ARTICLE INFO

Article history:

Received 19 July 2014

Received in revised form

10 September 2014

Accepted 12 September 2014

Available online 22 September 2014

Keywords:

WAK

Wall-associated Kinase

RLK

Rice

Oryza sativa

ABSTRACT

The environment is a dynamic system in which life forms adapt. Wall-Associated Kinases (WAK) are a subfamily of receptor-like kinases associated with the cell wall. These genes have been suggested as sensors of the extracellular environment and triggers of intracellular signals. They belong to the ePK superfamily with or without a conserved arginine before the catalytic subdomain VIB, which characterizes RD and non-RD WAKs. WAK is a large subfamily in rice. We performed an extensive comparison of WAK genes from *A. thaliana* (*AtWAK*), *O. sativa japonica* and *indica* subspecies (*OsWAK*). Phylogenetic studies and WAK domain characterization allowed for the identification of two distinct groups of WAK genes in *Arabidopsis* and rice. One group corresponds to a cluster containing only *OsWAKs* that most likely expanded after the *monocot–dicot* separation, which evolved into a non-RD kinase class. The other group comprises classical RD-kinases with both *AtWAK* and *OsWAK* representatives. Clusterization analysis using extracellular and kinase domains demonstrated putative functional redundancy for some genes, but also highlighted genes that could recognize similar extracellular stimuli and activate different cascades. The gene expression pattern of WAKs in response to cold suggests differences in the regulation of the *OsWAK* genes in the *indica* and *japonica* subspecies. Our results also confirm the hypothesis of functional diversification between *A. thaliana* and *O. sativa* WAK genes. Furthermore, we propose that plant WAKs constitute two evolutionarily related but independent subfamilies: WAK-RD and WAK-nonRD. Recognition of this structural division will further provide insights to understanding WAK functions and regulations.

© 2014 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Plants use several environmental factors, such as light, temperature, gravity and water as signals for different physiological processes, including growth, flowering and dormancy. However, changes in the environment can perturb the metabolic homeostasis of plants, inducing abiotic stresses [1]. The ability to recognize signals from the environment and other adjacent cells and activate appropriate and specific downstream signaling cascades is important to plant responses. Until the late 80s, it was not clear how

plant cells communicate with the environment. In 1990, the first receptor kinase protein (RLK – Receptor-Like Kinase) was described in plants [2]. Since then, several different plant RLK gene subfamilies have been identified, with more than 600 and 1100 genes in *Arabidopsis thaliana* and *Oryza sativa*, respectively [3–6].

The RLK family belongs to a protein kinase superfamily (ePK) that comprises all of the protein kinases identified in eukaryotes. The ePK superfamily plays an important role in post-translational protein modification through phosphorylation activity [7]. Members of the ePK superfamily possess 12 conserved subdomains. The subdomains I to IV are part of the amino-terminal lobe, associated with ATP binding. Subdomain V corresponds to a link between the amino and the carboxyl-terminal lobe. The subdomains VIA, VIB, VII, VIII, IX, X and XI are part of the carboxyl-terminal lobe, which is associated with peptide-substrate binding and phosphorylation activity [7–10]. The subdomain VIB contains the catalytic motif DxxxxN, which directly participates in substrate phosphorylation, and the presence of a conserved arginine (R) residue preceding the

* Corresponding author at: Centro de Biotecnologia, sala 213, prédio 43431, Universidade Federal do Rio Grande do Sul – UFRGS, P.O. Box 15005, CEP 91501-970 Porto Alegre, RS, Brazil. Tel.: +55 51 3308 7766; fax: +55 51 3308 6072.

E-mail addresses: rogerio.margis@ufrgs.br, rogerio.margis@gmail.com (R. Margis).

¹ These authors contributed equally to this manuscript.

Anexo V

- Artigo científico publicado em colaboração 5 –

Comprehensive selection of reference genes for quantitative gene expression analysis during seed development in *Brassica napus*

Ronei Dorneles Machado · Ana Paula Christoff ·
Guilherme Loss-Morais · Márcia Margis-Pinheiro ·
Rogério Margis · Ana Paula Körbes

Received: 17 October 2014/Revised: 19 January 2015/Accepted: 16 February 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract

Key message MicroRNAs have higher expression stability than protein-coding genes in *B. napus* seeds and are therefore good reference genes for miRNA and mRNA RT-qPCR analysis.

Abstract Reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR) has become the “gold standard” to gain insight into function of genes. However, the accuracy of the technique depends on appropriate reference genes for quantification analysis in different experimental conditions. Accumulation of microRNAs (miRNAs) has also been studied by RT-qPCR, but there are no reference genes currently validated for normalization of *Brassica napus* miRNA expression data.

Communicated by M. Menossi.

Electronic supplementary material The online version of this article (doi:10.1007/s00299-015-1773-1) contains supplementary material, which is available to authorized users.

R. D. Machado · A. P. Christoff · M. Margis-Pinheiro ·
R. Margis · A. P. Körbes (✉)
Departamento de Genética, PPGGBM, Universidade Federal do
Rio Grande do Sul, UFRGS, Av. Bento Gonçalves 9500,
Predio 43323 M, Sala 204/NGFP, Agronomia, Porto Alegre,
RS 91501-970, Brazil
e-mail: anapkorbes@yahoo.com.br; apkorbes@gmail.com

R. D. Machado
e-mail: dornelesronei@gmail.com

A. P. Christoff
e-mail: anachff@gmail.com

M. Margis-Pinheiro
e-mail: marcia.margis@ufrgs.br

R. Margis
e-mail: rogerio.margis@ufrgs.br

In this study, we selected 43 *B. napus* miRNAs and 18 previously validated mRNA reference genes. The expression stability of the candidate reference genes was evaluated in different tissue samples (stages of seed development, flowers, and leaves) using geNorm, NormFinder, and RefFinder analysis. The best-ranked reference genes for expression studies during seed development (miR167-1_2, miR11-1, miR159-1 and miR168-1) were used to assess the expression of miR03-1. Since candidate miRNAs showed higher expression stability than protein-coding genes in most of the tested conditions, the expression profile of DGAT1 gene was compared when normalized by the four most stable miRNAs reference genes and by the four most stable mRNA reference genes. The expected expression pattern of DGAT1 during seed development was achieved with the use of miRNA as reference genes. In conclusion, the most stable miRNA reference genes can be employed in the normalization of RT-qPCR quantification of miRNAs and protein-coding genes. This work is the first to perform a comprehensive survey of the

G. Loss-Morais · M. Margis-Pinheiro · R. Margis
Centro de Biotecnologia, PPGBCM, Universidade Federal do
Rio Grande do Sul, UFRGS, Porto Alegre, RS, Brazil
e-mail: guilherme.loss@gmail.com

G. Loss-Morais
LNCC, Laboratório Nacional de Computação Científica,
Labinfo, Laboratório de Bioinformática, Petrópolis, RJ, Brazil

R. Margis
Departamento de Biofísica, Universidade Federal do Rio Grande
do Sul, UFRGS, Porto Alegre, RS, Brazil

