UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO DE BIOTECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

IDENTIFICAÇÃO E ANÁLISE DE EXPRESSÃO DE microRNAS EM SOJA SOB ESTRESSE BIÓTICO E ABIÓTICO

Franceli Rodrigues Kulcheski

Porto Alegre, RS, Brasil
Fevereiro de 2013

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO DE BIOTECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

IDENTIFICAÇÃO E ANÁLISE DE EXPRESSÃO DE microRNAS EM SOJA SOB ESTRESSE BIÓTICO E ABIÓTICO

Franceli Rodrigues Kulcheski

Orientador: Dr. Rogério Margis

Tese submetida ao Programa de Pós-Graduação em Biologia Celular e Molecular do Centro de Biotecnologia da UFRGS como requisito parcial para a obtenção do título de Doutor em Ciências.

Porto Alegre, RS, Brasil
Fevereiro de 2013

Instituições e Fontes Financiadoras

As atividades de pesquisa cujos resultados estão reunidos nesta tese de doutorado foram desenvolvidas no Laboratório de Genomas e Populações de Plantas do Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, sob a orientação do Prof. Dr. Rogério Margis.

O estágio sanduíche de doutorado (09/2011 a 08/2012) foi desenvolvido no Laboratório de Biologia Molecular do Instituto Max-Planck para Biologia do Desenvolvimento, Tübingen, Alemanha, sob supervisão do Prof. Dr. Detlef Weigel.

Parte deste trabalho esteve inserido no projeto GenoSoja, financiado pelo CNPq dentro do plano de transcriptômica referente à pesquisa de microRNAs.

Este trabalho contou com recursos financeiros do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

“Por vezes sentimos que aquilo que fazemos não é senão uma gota no mar.
Mas o mar seria menor se lhe faltasse uma gota”.

Madre Teresa de Calcutá

Aos meus amados pais, Ivone e Sílvio;
E amada irmã, Patrícia;
Pelo amor incondicional, incentivo e apoio em todos os momentos.

Dedico

Agradecimentos

Agradecer é, antes de qualquer coisa, reconhecer o esforço ou mesmo um gesto das pessoas que de uma forma ou outra, ajudaram-me na realização deste estudo. Desta forma, relato o meu muito obrigada:

- Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela bolsa concedida.

- Ao Prof. Dr. Rogério Margis, pela oportunidade, orientação, confiança e apoio durante toda a etapa da realização deste trabalho.

- Ao Dr. Alexandre Nepomuceno, Drª. Francismar Marcelino-Guimarães, Dr. Ricardo Abdelnoor, Drª. Fabiana Rodrigues da EMBRAPA Soja pela colaboração ativa durante o desenvolvimento do projeto.

- Ao Prof. Dr. Detlef Weigel, que me aceitou para o período de "doutorado-sanduíche", e ao Dr. Pablo Manavella que me inseriu em um dos seus projetos durante minha estada no laboratório de Biologia Molecular do Instituto Max-Planck para Biologia do Desenvolvimento.

- A todos os co-autores dos artigos produzidos durante o doutorado.

- Aos professores da minha comissão de acompanhamento durante o curso de doutorado Dr. Arthur Germano Fett Neto e Dr. Guido Lenz.

- Aos professores que formaram a banca para a minha qualificação de doutorado: Dr. Giancarlo Pasquali, Dr. Guido Lenz e Drª. Luciane Passaglia.

- Aos professores: Dr. Arthur Fett Neto, Drª. Claudia Balbinotti de Andrade e Dr. Régis Correa, por aceitarem participar da banca para a defesa desta tese de doutorado.

- Ao professor Dr. Giancarlo Pasquali por aceitar o convite para revisor e membro suplente desta tese.

- Aos funcionários do Programa de Pós-Graduação em Biologia Celular e Molecular: Sílvia e Luciano, exemplos de competência e dedicação.

- Ao grupo do Laboratório de Genomas e Populações de Plantas (LGPP), aos que já passaram e aos atuais. Mas em especial aos queridos colegas com os quais tive mais convívio durante este período: Ana, Andréia, Cláudia, Felipe, Fernanda, G. Loss, G. Cordenonsi, Júlio e Lorrayne.

- Ao meu amigo Vinicius Galvão, grande parceria durante o período sanduíche no Max-Planck.

- Meu agradecimento especial às amigas que foram companheiras durante esta jornada. Déia e Cláu obrigada pela amizade, carinho e por sempre estenderem a mão quando eu precisei.

- A todos os demais amigos, que sempre estiveram torcendo por mim, obrigada pelo carinho.

- A minha família, que é a base de tudo para mim: meus amados pais, e minha amada irmã e cunhado. Obrigada pelo inestimável apoio, amor e confiança dedicados ao longo de todos os anos de minha vida!

- A Deus, por sempre estar presente em minha vida e pelos anjos que coloca em meu caminho!

Identificação e análise de expressão de microRNAs em soja sob estresse biótico e abiótico[1]

Autor: Franceli Rodrigues Kulcheski

Orientador: Rogério Margis

Resumo

Seca e ferrugem asiática da soja (FAS) são dois dentre os principais estresses abióticos e bióticos que afetam negativamente a produtividade da soja (*Glycine max* L. Merrill) no mundo inteiro. A base genética da tolerância à seca e da resistência à FAS não são bem conhecidas e esclarecer como ocorre a resposta a estes estresses em soja é ainda um desafio. Atualmente, sabe-se que as plantas adaptam-se a estes estresses por meio da regulação da expressão gênica em nível transcricional e pós-transcricional. Na via da regulação pós-transcricional, microRNAs (miRNAs) têm sido apontados como importantes reguladores em várias plantas sob estresse biótico e abiótico. Entretanto, em soja, não havia sido relatado qualquer miRNA responsivo a estas condições. Neste contexto, nosso objetivo foi identificar novos miRNAs em soja e, também, caracterizar o padrão de expressão de alguns destes miRNAs durante ambos os estresses, além de buscar detectar genes alvos para estes miRNAs. Deste modo, esta tese foi dividida em capítulos, os quais apresentam os diferentes trabalhos desenvolvidos durante o doutorado. No capítulo III estão relatados os resultados da primeira investigação sobre a adequação de miRNAs como genes normalizadores em plantas. A estabilidade da expressão dos miRNAs foi investigada em diferentes tecidos e genótipos de soja, bem como entre estresses biótico e abiótico. Ao final deste trabalho, foram mostradas evidências de que a estabilidade da expressão de miRNAs pode ser maior que a de genes codificadores de proteínas em análises de RT-qPCR. No capítulo IV está descrita a descoberta de novos miRNAs a partir de bibliotecas de deficiência hídrica e FAS em soja, pelo emprego de sequenciamento de alto desempenho (Solexa). Neste estudo, foram detectados 256 miRNAs. Análises de RT-qPCR foram realizadas para alguns dos novos miRNAs, observando-se alguns miRNAs diferencialmente expressos, indicando evidência molecular para um possível envolvimento de miRNAs em processos responsivos à deficiência hídrica e FAS. Para um dos miRNAs detectado, um novo alvo foi validado, correspondendo a um gene codificador de ascorbato oxidase e possivelmente relacionado com a infecção pelo fungo da ferrugem asiática em genótipo suscetível, dados estes apresentados no capítulo V. Esta tese contribui para o aumento de informações sobre miRNAs de plantas, bem como para o entendimento da regulação gênica em soja sob estresses de deficiência hídrica e FAS.

---

[1] Tese de Doutorado, Programa de Pós-graduação em Biologia Celular e Molecular, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil. (184 p.) Fevereiro, 2013.

Identification and expression analysis of microRNAs in soybean under biotic and abiotic stresses[1]

Author: Franceli Rodrigues Kulcheski

Adviser: Rogério Margis

Abstract

Drought and Asian Soybean Rust (ASR) are the major abiotic and biotic stresses that negatively affect soybean (*Glycine max* L. Merrill) productivity around the world. The genetic basis of drought tolerance and ASR resistance are not well understood, and clarification on how the response to these stresses occur in soybean is still a challenge. Currently, it is known that adaptation is achieved through the regulation of gene expression at the transcriptional and post-transcriptional levels. In the way of post-transcriptional regulation, microRNAs (miRNAs) have been found to act as key regulator factors in many other plants under biotic and abiotic stresses. However, in soybean there was no report of miRNAs responsive to these conditions. In this context, our goal was to identify new miRNAs in soybean, characterize some of the miRNA expression patterns during both stresses, and try to detect target genes for the miRNAs. In this way, this thesis was divided in chapters which present the different works that were developed during the PhD period. In chapter three, the suitability of miRNAs as housekeeping genes in plants was investigated. MiRNA expression stability was analysed in different soybean tissues and genotypes as well as after abiotic or biotic stress treatments. It was shown that miRNA expression stability can be higher than the expression stability of protein-coding genes by RT-qPCR analysis. In chapter four, new miRNAs were discovered from Solexa deep sequencing of soybeans submitted to water deficit and rust infection. From these analyses 256 miRNAs were detected. RT-qPCRs were performed for some of the new miRNAs and the identification of differentially expressed miRNAs was observed, provinding molecular evidence for the possible involvement of miRNAs in the process of water deficit- and rust-stress responses. For one of the new miRNAs detected by Solexa sequencing, a new target was validated, which is a gene encoding ascorbate oxidase and that seems to be related with soybean rust infection in the genotype susceptible to the fungus (results were presented in chapter five). The present thesis contributes to improve the information about miRNAs in plantas, as well as to the understanding of soybean gene regulation under water deficit and ASR stresses.

---

[1] Ph.D. thesis, Graduating Program in Celular and Molecular Biology, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil. (184 p.) February, 2013.

Sumário

Página

Lista de Abreviaturas *

µg – micrograma

µL – microlitro

ºC – graus Celsius

5'RACE – Rapid Amplification of cDNA 5' (rápida amplificação da região terminal 5' de cDNA)

ASR – Asian soyben rust

ATP – Adenosine triphosphate (trifosfato de adenosina)

BLAST – Basic Local Alignment Sequence Tool (ferramenta básica de alinhamento local de sequências)

bp – base pair (pares de bases)

cDNA – complementary DNA (DNA complementar)

CDS – Coding sequence (sequência codificadora)

DNA – Desoxirribonucleic acid (ácido desoxirribonucleico)

FAS – Ferrugem asiática da soja

mRNA – messenger RNA (RNA mensageiro)

miRNA – microRNA

nt - nucleotídeo

pre-miRNA – precursor of microRNA (precursor de microRNA)

pri-miRNA – primary microRNA (microNA primário)

PCR – Polymerase Chain Reaction (Reação da DNA polimerase em cadeia)

qPCR – quantitative PCR (PCR quantitativo)

RNA – Ribonucleic acid (ácido ribonucleico)

RT-qPCR – Reverse transcription qPCR (transcrição reversa seguida de qPCR)

siRNA- small interfering RNA (pequeno RNA de interferência)

smRNA – small RNA (pequeno RNA)

* Nesta lista estão citadas apenas as abreviaturas mais frequentes ao longo do trabalho. As demais abreviaturas empregadas possuem seu significado no próprio corpo do texto.

CAPÍTULO I

Introdução

1.1 Soja: caracterização e importância da cultura

A soja (*Glycine max* (L.) Merrill) é uma planta da família das Fabáceas (leguminosas) de grande importância econômica no mundo. Segundo dados da *Food and Agriculture Organization of United Nations* (FAO), no ano de 2010 foram produzidos aproximadamente 264 milhões de toneladas de soja em todo o mundo (FAO, 2012). Neste contexto, o Brasil aparece como o segundo maior produtor mundial, com produção de 66,3 milhões de toneladas na safra 2011/12 (CONAB, 2012). Além disto, a soja lidera os produtos agropecuários, na pauta de exportações do Brasil.

Originária da Ásia, sobretudo da China, a cultura que hoje se planta resultou da evolução de sucessivos processos de melhoramento de genótipos ancestrais. O cultivo da soja é muito antigo. Alguns relatos revelam que os plantios datam de 2.838 anos a.C., sendo que, naquele período, era considerada uma planta sagrada (Hymowitz, 1970). Por séculos, a cultura permaneceu restrita ao oriente, sendo apenas introduzida no ocidente, pela Europa, por volta do século XV, não com finalidade de alimentação, como acontecia na China e Japão, mas de ornamentação. Mais de 500 anos passaram-se até que a civilização ocidental percebesse o valor do grão de soja na alimentação, principalmente o seu valor proteico. As primeiras tentativas de produção de soja na Europa fracassaram, provavelmente devido a fatores climáticos, a ausência de conhecimento sobre a cultura e suas exigências. Os norte-americanos foram os que, entre o fim do século XIX e início do século XX, conseguiram desenvolver o cultivo

comercial da soja, criando novas variedades com teor de óleo mais elevado. A partir de então, ocorreu a expansão do seu cultivo (CISoja, 2012).

A introdução da soja no Brasil ocorreu por volta de 1882, mas apenas no início do século XX a cultura começou a se estabelecer no território brasileiro. Relatos indicam que foi nesse período que na região sul do país, mais especificamente no estado do Rio Grande do Sul, começou a ser cultivada a soja, e foi nessa região que a cultura encontrou condições ideais para seu desenvolvimento. A boa adaptação ao clima da região sul do país é creditada à similaridade com o clima do sul dos Estados Unidos, local de origem dos primeiros genótipos da soja brasileira. A região sul foi responsável, até 1960 e 1970, por ser a produtora majoritária do país, sobretudo no Rio Grande do Sul e Paraná, ainda hoje grandes produtores. Porém, atualmente, já perderam em volume para o Mato Grosso, que é agora o maior produtor nacional (CISoja, 2012).

A soja tem se destacado por ser uma valiosa fonte de proteínas e óleo vegetal. As porcentagens de proteína e óleo em soja, considerando que estes valores são influenciados pelo genótipo e ambiente, têm uma média aproximada de 40% e 20%, respectivamente (Clemente e Cahoon, 2009). O grão é o componente essencial na fabricação de rações animais, sendo que o uso crescente na alimentação humana encontra-se em franco crescimento. A indústria nacional transforma, por ano, cerca de 30,7 milhões de toneladas de soja, produzindo 5,8 milhões de toneladas de óleo comestível e 23,5 milhões de toneladas de farelo proteico, contribuindo para a competitividade nacional na produção de carnes, ovos e leite. Além disto, a soja e o farelo de soja brasileiros possuem alto teor de proteína e padrão de qualidade *Premium*, o que permite sua entrada em mercados extremamente exigentes como os da União Européia e do Japão. Além disto, a soja é uma alternativa para a fabricação do biodiesel, combustível capaz de reduzir em 78% a emissão dos gases causadores do efeito estufa

na atmosfera (Ministério da Agricultura, 2012). Outro papel fundamental da soja está relacionado ao aumento da fertilidade do solo devido à fixação do nitrogênio atmosférico resultante do processo simbiótico entre raízes de leguminosas e bactérias nitrificantes (Crespi e Frugier, 2008; Markmann e Parniske, 2009).

A soja é uma planta anual com caule ereto, com grande diversidade quanto ao ciclo de vida, variando de 70 dias para os genótipos de ciclo mais precoce até 200 dias para os mais tardios. O ciclo da planta pode ser dividido em duas fases: vegetativa e reprodutiva (Fehr e Caviness, 1977; Neumaier et al., 2000). A fase vegetativa corresponde do período da emergência da plântula até a abertura das primeiras flores e pode ser classificada em VE, VC, V1, V2, V3, Vn, e a fase reprodutiva compreende o período do início da floração até a maturação sendo classificada em R1, R2, R3, R4, R5, R6, R7 e R8 (Fehr e Caviness, 1977; Neumaier et al., 2000).

A soja adapta-se melhor a temperaturas do ar entre 20ºC e 30ºC, podendo variar segundo a cultivar. Sempre que possível, a semeadura da soja não deve ser realizada quando a temperatura do solo estiver abaixo de 20ºC pois prejudica a germinação e a emergência. Como a maioria das plantas, o florescimento da soja é influenciado pela temperatura e pela duração do período luminoso, o fotoperíodo. A soja é considerada uma planta de dia curto, precisa de uma duração da noite maior que o dia para iniciar o processo de floração e frutificação no momento certo, isto é, após ter atingido o crescimento vegetativo adequado. A grande produção de soja no Brasil hoje deve-se ao desenvolvimento de novas cultivares com fotoperíodo mais longo, tornando-as aptas às diversas regiões (EMBRAPA, 2011).

A soja apresenta um genoma moderamente complexo e tamanho aproximado de 1.150 milhões de pares de bases (Mpb) (Cannon e Shoemaker, 2012; Schmutz et al., 2010), distribuídos em 20 pares de cromossomos. Comparada com outras espécies de

plantas economicamente importantes, a soja tem um genoma com aproximadamente o triplo do genoma do arroz, mas com a metade do genoma do milho. O número de genes codificadores preditos em soja também é relativamente alto, aproximadamente 46.400 genes, cerca de 70% a mais do que em arabidopsis (Cannon e Shoemaker, 2012; Schmutz et al., 2010). Duplicações do genoma ocorreram há aproximadamente 59 e 13 milhões de anos, resultando em um genoma altamente duplicado com cerca de 75% dos genes presentes em múltiplas cópias. Os dois eventos de duplicação foram seguidos por diversificação, perda e numerosos rearranjos cromossômicos (Schmutz et al., 2010), o que adiciona um grau de dificuldade em projetos genômicos para essa cultura.

1.2 Principais vilões da cultura da soja: deficiência hídrica e ferrugem asiática

Plantas estão expostas a ambientes dinâmicos que podem frequentemente impor dificuldades ao seu crescimento e desenvolvimento, resultando em significantes perdas no rendimento de espécies com importância econômica, como é o caso da soja. Diversos fatores bióticos e abióticos prejudicam a cultura da soja. A seca é o principal estresse abiótico que afeta negativamente a produtividade da soja em todo o mundo, sendo este o principal motivo sinistrante na cultura (71% dos casos) (Casagrande et al., 2001). Já a ferrugem asiática da soja (FAS) é o estresse biótico mais danoso que acomete a cultura nos países produtores de soja. Relatos indicam que os danos na lavoura causados por esta doença podem variar entre 10% e 90% nas diferentes regiões onde tem sido identificada (Sinclair e Hartman, 1999; Yorinori et al., 2005).

1.2.1 Deficiência hídrica

A água constitui aproximadamente 90% do peso da planta, atuando em praticamente todos os processos fisiológicos e bioquímicos. Desempenha a função de solvente, através do qual gases, minerais e outros solutos entram nas células e movem-se pela planta. Apresenta também um papel importante na regulação térmica da planta, agindo tanto no resfriamento como na manutenção e distribuição do calor (Norman e Nepomuceno, 1994).

A disponibilidade de água é importante, principalmente, em dois períodos de desenvolvimento da soja: germinação-emergência (Bewley e Black, 1994) e floração-enchimento de grãos (Desclaux et al., 2000; Desclaux e Roumet, 1996). Durante o primeiro período, tanto o excesso quanto a deficiência de água são prejudiciais à obtenção de uma boa uniformidade na população de plantas.

A semente de soja necessita absorver, no mínimo, 50% de seu peso em água para assegurar boa germinação. Nessa fase, o conteúdo de água no solo não deve exceder a 85% do total máximo de água disponível e nem ser inferior a 50%. A necessidade de água na cultura da soja vai aumentando com o desenvolvimento da planta, atingindo o máximo durante a floração-enchimento de grãos (7 a 8 mm/dia) e decrescendo após esse período. Uma deficiência hídrica expressiva durante a floração e o enchimento de grãos, provoca alterações fisiológicas na planta, como o fechamento estomático e o enrolamento de folhas, provocando queda prematura de folhas e flores e abortamento de vagens, resultando, por fim, em redução do rendimento de grãos (EMBRAPA, 2008; Karam et al., 2005; Liu et al., 2003). Além disso, foi observado que plantas de soja submetidas à limitação de água apresentam o metabolismo de nitrogênio desregulado, devido a uma forte inibição na fixação biológica de nitrogênio (Clement et al., 2008).

Uma resposta fisiológica específica à deficiência hídrica é o resultado da combinação de eventos moleculares que são ativados ou desativados pela percepção do estresse. Por exemplo, Casagrande et al. (2001) observaram diferenças de expressão da enzima NADH desidrogenase entre cultivares de soja sensíveis e tolerantes à seca. As enzimas NADH desidrogenases são componentes da membrana mitocondrial, e estão diretamente envolvidas no transporte de elétrons da cadeia respiratória, e consequentemente, no fornecimento de energia para a célula (Casagrande et al., 2001; Dey e Harbone, 1997).

Outro estudo abordando a expressão diferencial de genes em uma cultivar de soja tolerante à seca demonstrou que alguns genes são induzidos ou reprimidos dependendo do tempo de estresse ao qual a planta é submetida (Martins et al., 2008). O fator transcricional bHLH (*basic Helix-loop-Helix*), envolvido na regulação de genes responsáveis pela diferenciação celular, foi altamente induzido após 100 minutos de desidratação. Enquanto que os genes codificadores da proteína transportadora de fosfatidilinositol (*PITP*) e do regulador de crescimento independente de auxina (*AXI1*) foram reprimidos após o ínicio do estresse (a partir de 50 minutos de exposição à falta de água) em comparação à situação controle (não estressada). Proteínas PITP têm a sua atividade relacionada a processos de desenvolvimento normal da célula, como a percepção do ciclo circadiano, divisão celular e controle estomático. Já a proteína AXI1 está diretamente relacionada com processos de divisão celular (Martins et al., 2008).

No trabalho recentemente desenvolvido por Guimaraes-Dias et al. (2012), as análises envolvendo dois genótipos diferentes quanto à tolerância ou sensibilidade ao estresse hídrico demonstraram que os genes da galactinol sintase, *GmGOLS2-like2* e *GmGOLS2-like3*, tiveram sua expressão aumentada no genótipo tolerante sob estresse. Estes genes fazem parte da biossíntese dos oligossacarídeos rafinose e estaquiose.

GOLS2 já foi estudado previamente em arabidopsis, e a superexpressão deste gene aumentou a tolerância à desidratação nestas plantas (Taji et al., 2002).

Uma classe de genes de bastante interesse nas pesquisas com seca são os fatores transcricionais *DREB* (*Dehydration Responsive Element Binding Protein*) (Yamaguchi-Shinozaki e Shinozaki, 1994). Este gene codifica uma proteína regulatória, proteína DREB, um fator de transcrição que está envolvido na ativação de outros genes relacionados à tolerância ao estresse hídrico. O gene *DREB* foi identificado em *Arabidopsis thaliana* e patenteado pelo instituto japonês *Japan International Research Center for Agricultural Sciences* (JIRCAS). Este gene vem sendo utilizado há anos em pesquisas de soja para aumentar a tolerância à seca. O JIRCAS liberou o gene DREB para uso da EMBRAPA em soja em 2003. Para comprovar a eficácia do uso do gene, ele foi introduzido em uma cultivar de soja brasileira que é sensível à seca, e resultados positivos foram observados em laboratório e em estufas (EMBRAPA, 2012).

Outros estudos de genômica funcional utilizando estratégias combinadas de transcriptômica, proteômica e metabolômica em reposta à seca em outras espécies além da soja também têm reportado uma ampla gama de fatores importantes envolvidos neste processo. Entre eles pode-se salientar as proteínas LEA (*late-embryogenesis-abundant proteins*); aquaporinas; proteínas de transferência de lipídeos; proteínas envolvidas em reparo; proteínas de membrana envolvidas no aparato fotossintético, inibidores de proteinases; enzimas antioxidantes; osmólitos; fatores transcricionais; receptores de membranas envolvidos em sinalização (Ramanjulu e Bartels, 2002).

1.2.2 A ferrugem asiática

A FAS é causada pelo fungo patogênico *Phakopsora pachyrhizi* Sydow & Sydow, originário da região Australasiana. No Brasil, as primeiras epidemias ocorreram

a partir de 2001. Em 2002, a doença já estava disseminada em 60% da área de cultivo do país e 90% no ano seguinte (Furtado et al., 2009). Este patógeno apresenta rápida dispersão aérea e alta capacidade de colonizar tecido foliar, e em menor escala, caule e vagens (van de Mortel et al., 2007). O processo de infecção dos urediniósporos de *P. pachyrhizi* compreende diversos passos distintos: adesão da superfície do hospedeiro, germinação, formação do apressório, penetração através da cutícula, invasão e crescimento das hifas no tecido hospedeiro (Furtado et al., 2009).

Os primeiros sintomas da ferrugem da soja iniciam com diminutas manchas de cor marrom ou vermelho-tijolo nas folhas. No campo, estas manchas inicialmente aparecem nas folhas mais baixas do dossel durante ou após a floração, embora plântulas podem ser infectadas sob certas circunstâncias. Frequentemente, as primeiras lesões aparecem na base do folíolo perto do pecíolo e das nervuras da folha. Esta região do folíolo provavelmente retém a umidade por um tempo mais longo, proporcionando condições mais favoráveis à infecção. As lesões permanecem pequenas (2-5 mm de diâmetro), porém crescem em número com o progresso da doença. Pústulas, que são as urédias, são formadas nas lesões principalmente na face inferior das folhas e podem produzir uma grande massa de uredósporos. Pústulas maduras podem ser vistas a olho nu, especialmente durante a esporulação. A germinação dos urediniósporos de *P. pachyrhizi* ocorre através de um poro central, com a produção de um tubo germinativo que termina em um apressório, o qual o fungo utiliza para penetrar diretamente o hospedeiro ultrapassando a cutícula, ou com menor frequência pelos estômatos. A disseminação da ferrugem é feita principalmente pela dispersão dos urediniósporos pelo vento.

A infecção por *P. pachyrhizi* causa rápido amarelecimento e queda prematura das folhas. Quanto mais cedo ocorrer a desfolha, menor será o tamanho dos grãos e,

consequentemente, maior a perda do rendimento e da qualidade dos mesmos (Furtado et al., 2009). Em casos severos, quando a doença atinge a soja na fase de formação das vagens ou no início da granação, pode causar o aborto e a queda das vagens, resultando em até perda total do rendimento. Elevadas perdas de rendimento têm sido registradas na Austrália (80%), na Índia (90%) e em Taiwan (70%-80%). No Brasil, reduções de produtividade de até 80% têm sido observadas, quando se comparam áreas tratadas e não tratadas com fungicidas. As regiões onde a doença tem sido mais agressiva variam de safra para safra, em função das condições climáticas e do inóculo inicial (EMBRAPA, 2008).

Até o momento, não há cultivares comerciais resistentes a essa doença. Contudo, já foram relatados seis genes de resistência vertical (ou também chamados de "genes maiores") denominados *Rpp1*, *Rpp2*, *Rpp3*, *Rpp4*, *Rpp5* e *Rpp6* (Bromfield e Hartwig, 1980; Cheng e Chan, 1968; Garcia et al., 2008; Hartwig, 1986; Hidayat e Somaatmadja, 1977; Li et al., 2012a). Embora todos os genes *Rpp* tenham sido mapeados em grupos de ligação em particular, nenhum destes genes foi isolado e caracterizado até a presente data (Goellner et al., 2010; Morceli et al., 2008). Algum progresso foi obtido na caracterização da resistência mediada por *Rpp4*, sendo que o locus gênico, bem como o sequenciamento do mesmo, demonstraram que este gene apresenta similaridade à família de genes *RGC2* em alface, o quais apresentam sítios de ligação de nucleotídeos e de repetição ricas em leucina (NBS-LRR), sendo estes domínios comumente encontrados em genes de resistência (Goellner et al., 2010; Meyer et al., 2009).

Ainda com relação ao aspecto molecular da resistência à FAS, van de Mortel et al. (2007) demonstraram uma expressão diferencial de genes dependentes do tempo de infecção em plantas contendo genes *Rpp2*. Neste trabalho observaram que genes que

alcançavam picos de expressão 12 horas após a infecção (hai) retornavam a uma expressão basal nas 24 hai. Em outro estudo, com plantas de soja contendo o gene *Rpp1*, foi realizada uma ampla análise do padrão de transcritos e observaram que os genes de lipoxigenase e peroxidase tiveram expressão aumentada durante interações incompatíveis, isto é, de resistência, sugerindo a participação destes genes na resistência mediada por *Rpp1* (Choi et al., 2008).

1.3 Os microRNAs em plantas

Os microRNAs, ou miRNAs, constituem a principal classe de pequenos RNAs envolvidos na regulação da expressão gênica, atuando em uma série de processos biológicos, como crescimento, desenvolvimento e adaptação a estresses diversos (Chen, 2005; Lu et al., 2008b; Mallory e Vaucheret, 2006; Shukla et al., 2008). O primeiro miRNA, *lin-4* (do inglês, *lineage-deficient-4*), foi descoberto em 1993 e identificado como regulador pós-transcricional do gene *lin-14* que está envolvido no controle do desenvolvimento larval de *Caenorhabditis elegans* (Lee e Ambros, 2001). Em plantas, a identificação de miRNAs ocorreu pela primeira vez em arabidopsis e foi descrita em 2002 por dois grupos distintos (Park et al., 2002; Reinhart e Bartel, 2002). Atualmente, os miRNAs já foram descritos em 67 espécies de plantas e todas as suas sequências estão disponíveis em um banco de dados público denominado miRBase (http://www.sanger.ac.uk/cgi-bin/Rfam/mirna/browse.pl) (Griffiths-Jones, 2004, 2006; Griffiths-Jones et al., 2006; Griffiths-Jones et al., 2008; Kozomara e Griffiths-Jones, 2010). O aumento na identificação de miRNAs tem demonstrado a relação dos mesmos com diversas rotas celulares. Inúmeros miRNAs identificados em plantas foram implicados com processos biológicos tais como desenvolvimento (Lelandais-Briere et al., 2010; Yang et al., 2007), sinalização hormonal (Liu e Chen, 2009), floração e

determinação sexual (Chuck et al., 2009), bem como respostas a estresses bióticos e abióticos (Katiyar-Agarwal e Jin, ; Lu e Huang, 2008; Lu et al., 2008b; Shukla et al., 2008; Sunkar e Zhu, 2004; Yang et al., 2007).

Análises experimentais e computacionais têm demonstrado que muitos miRNAs, bem como seus alvos, são conservados entre espécies de monocotiledôneas (arroz) e dicotiledôneas (arabidopsis) (Bonnet et al., 2004; Wang et al., 2004a; Wang et al., 2004b). Estes estudos têm evidenciado que a origem dos miRNAs ocorreu em uma fase inicial da evolução dos eucariotos, antecedendo à divergência entre plantas monocotiledôneas e dicotiledôneas.

Em plantas, genes de miRNAs são transcritos por uma enzima RNA polimerase do tipo II (RNA Pol II) (Jones-Rhoades et al., 2006; Lee et al., 2004); entretanto, alguns miRNAs podem ser transcritos pela RNA Pol III (Faller e Guo, 2008). Os transcritos iniciais de miRNAs são chamados miRNAs primários (pri-miRNAs). RNA Pol II gera pri-miRNAs capeados e poliadenilados tanto em plantas quanto em animais (Lee et al., 2004). Após a transcriçao dos pri-miRNAs, a proteína Dawdle (DDL) liga-se a eles protegendo-os contra degradação. Este pri-miRNA forma uma estrutura em formato de grampo ("*hairpin structure*") imperfeita, a qual será processada em um pre-miRNA (precusor do miRNA) também conhecido como estrutura "*stem-loop*". Em plantas, ambos pri e pré-miRNAs são processados no núcleo por uma enzima do tipo RNaseIII denominada Dicer-Like 1 (DCL1) (Kurihara et al., 2006). O processamento do pri-miRNA em pre-miRNA ocorre em um centro de processamento nuclear conhecido como corpo-D (do inglês, "*D-body"* ou "SmD3/SmB-*body*"). Embora não esteja claro o comprimento efetivo dos pri-miRNAs, sabe-se que todos podem dobrar-se e originar uma estrutura secundária em formato de grampo (Meyers et al., 2008). Possuir estruturas *hairpin* ou "grampo" é uma importante característica de miRNAs (Zhang et

al., 2006). Pre-miRNAs, ainda no núcleo, são clivados pela DCL1 e pela proteína HYL1 (*Hyponastic Leaves* 1) originando uma dupla miRNA:miRNA* (Kurihara et al., 2006; Song et al., 2007; Song et al., 2011). As duplas miRNA:miRNA* recentemente processadas possuem dois nucleotídeos não pareados na região terminal 3' de cada fita e são facilmente degradados por uma classe de exonucleases conhecidas por SDN (*Small RNA Degradading Nuclease*) (Ramachandran e Chen, 2008). Para estabilizar a dupla miRNA:miRNA*, uma metil-transferase de pequenos RNAS, HEN1 (Hua *Enhancer*), imediatamente metila os nucleotídeos no terminal 3' de cada fita prevenindo assim sua uridilação e subsequente degradação (Yu et al., 2005). Em mutantes *hen1*, observou-se falta ou acúmulo mínimo de miRNAs, sugerindo o papel de HEN1 na proteção de miRNAs contra degradação (Li et al., 2005). O duplex miRNA/miRNA* é então transportado do núcleo para o citoplasma pela proteína de membrana HASTY (Bollman et al., 2003).

No citoplasma, a dupla miRNA/miRNA* é separada, a fita de miRNA é incorporada no complexo de silenciamento induzido por RNA ou RISC (*RNA-induced silencing complex*) formando, assim, o complexo miR-RISC (Chen, 2005), no qual uma proteína AGO1 (*Argonaute* 1) cliva o mRNA alvo no meio da dupla mRNA-miRNA. Já a fita miRNA* será degradada. Entretanto, alguns estudos têm demonstrado que miRNA* também podem funcionar como uma sequência efetiva de miRNA e controlar a expressão de genes alvos específicos (Guo e Lu, 2009).

A sequência de um miRNA maduro pode variar entre 19 a 24 nucleotídeos (nt) e atua silenciando genes pós-transcricionalmente, através do pareamento com a sequência do mRNA alvo, levando à clivagem ou à repressão traducional deste (Bartel, 2004, 2009). Além disto, um mesmo miRNA maduro pode apresentar sequências variando de um a dois nucleotídeos no seu comprimento. Estas populações de miRNAs são

chamadas de isomiRNAs, o que significa isoforma de um miRNA. A origem destas variantes de miRNAs ainda não está elucidada. Alguns autores acreditam que elas ocorram devido a um erro de clivagem da DCL1 durante o processamento do pré-miRNA (Guo e Lu, 2009), entretanto existem opiniões divergentes quanto a esta teoria.

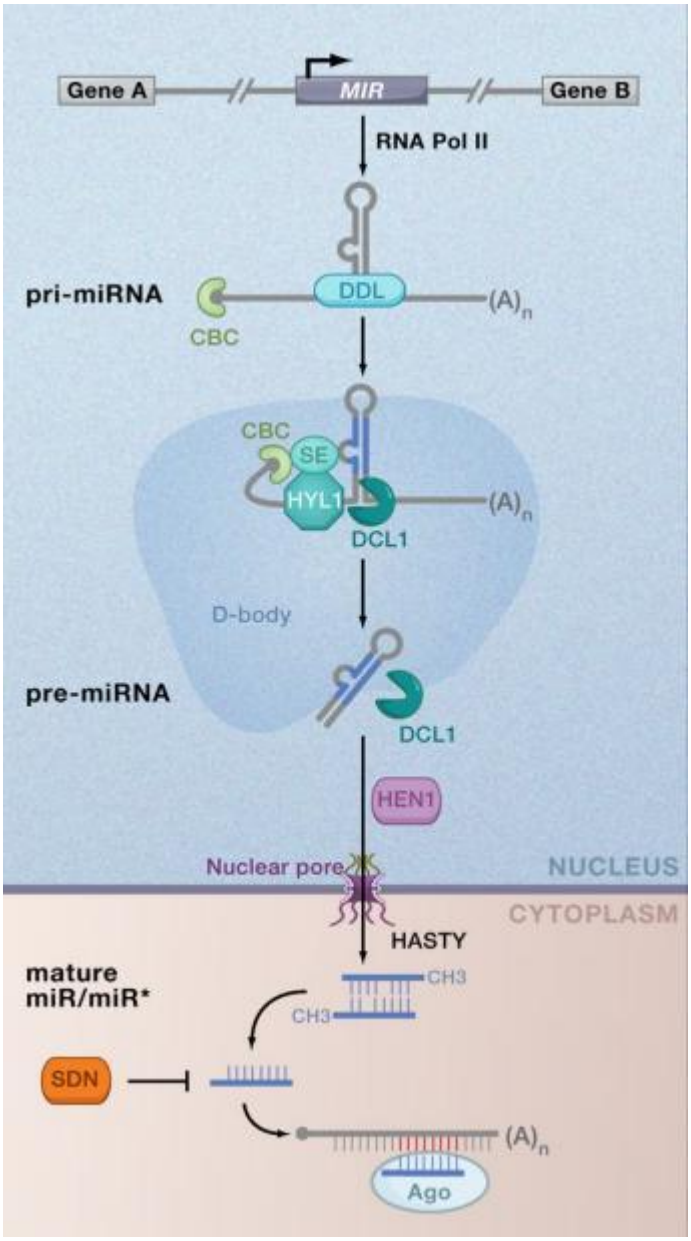


Figura 1. Biogênese dos miRNAs de plantas. Os pri-miRNAs são transcritos por uma RNA Pol II formando uma estrutura em formato de grampo. A esta estrutura liga-se a proteína Dawdle (DDL), responsável por estabilizar o pri-miRNA até a sua conversão em pre-miRNA no centro de processamento nuclear (D-body). Este processo conta com a ação combinada das proteínas: Serrate (SE), Hyponastic Leaves 1 (HYL1), Dicer-like 1 (DCL1), e nuclear cap-binding complex (CBC). O pre-miRNA é então clivado novamente por uma DCL1, gerando um duplex miRNA/miRNA*, que será exportado para o citoplasma pela proteína de membrana HASTY. Ambas sequências miRNA/miRNA* serão metiladas por uma metilase HEN1, protegendo as sequências da degradação via nucleases degradadoras de smRNAs (SDN). A fita guia de miRNA é então incorporada em proteínas AGO no complex RISC, formando o complexo miR-RISC agindo no silenciamento gênico (adaptado de Voinnet, 2009).

Como descrito acima, os processos de biogênese e ação de miRNAs requerem a combinação e interações físicas de várias enzimas e/ou proteínas como as proteínas DCL1, SE (proteínas Serrate a qual contém motivos C2H2-*zinc finger*), CBC (*cap-binding complex*), AGO, RNA polimerases dependente de RNA (RDP), RNA helicase SDE3, HYL1 e CPL1 (*C-terminal domain phosphatase-like* 1) (Dalmay et al., 2001; Fagard et al., 2000; Fang e Spector, 2007; Kurihara et al., 2006; Manavella et al., ; Mourrain et al., 2000).

A perda de função de DCL1 e SE são geralmente letais e causam a morte da planta ainda nos primeiros estádios embrionários uma vez que nenhum acúmulo de miRNA maturo é observado nas células vegetais (Xie et al., 2003). Mutantes *ddl* e *cbc* também têm baixo acúmulo de miRNAs (Kim et al., 2008). Entretanto, outras evidências sugerem que a proteína DDL pode ter outras funções além de participar da biogênese de miRNAs, uma vez que mutantes *ddl* possuem anormalidades mais fortes no desenvolvimento da planta que *dcl1* mutantes (Yu et al., 2008).

Outro estudo demonstrou que a perda de função do gene *HASTY* resulta apenas em um decréscimo de miRNAs no citoplasma, o que sugere que miRNAs podem também ser transportados por outros mecanismos (Voinnet, 2009). Mutação do gene *HYL1*, responsável pela proteína que atua em conjunto com DCL1 na primeira fase de processamento do pri-miRNA, também levou à alteração do desenvolvimento foliar, além de afetar a dominância apical e a sensibilidade hormonal em arabidopsis. Estas plantas apresentaram alta sensibilidade ao ácido abscísico e baixa sensibilidade à citocina e à auxina (Lu e Fedoroff, 2000; Vazquez et al., 2004).

1.4 Os microRNAs e o estresse hídrico em plantas

Para reduzir os danos causados por estresses, as plantas têm desenvolvido, ao longo da evolução, sofisticadas respostas adaptativas envolvendo reprogramação da expressão gênica em nível transcricional, pós-transcricional e pós-traducional (Shukla et al., 2008). Recentemente descobertos, os miRNAs têm sido apontados como importantes fatores envolvidos nas repostas a estresses em plantas.

A deficiência hídrica é um dos estresses abióticos que mais afeta a fisiologia das plantas, e por isto tem sido alvo de vários estudos enfocando as vias de modulação gênica durante esta condição. Neste contexto, tem-se realizado vários estudos abordando a expressão de miRNAs tanto em plantas modelo, como arabidopsis, quanto em espécies de interesse econômico, caso da soja.

Experimentos realizados com arabidopsis, demonstraram que o MIR393, MIR397b e MIR402 foram fortemente induzidos em situação de seca (Sunkar e Zhu, 2004). MIR393 regula TIR1, um regulador positivo de crescimento e desenvolvimento, e desta forma atenua o crescimento e o desenvolvimento da planta durante o estresse (Sunkar e Zhu, 2004). Nesta mesma linha, MIR397b regula um gene de lacase o qual foi demonstrado estar envolvido na redução do crescimento de raiz durante desidratação em um mutante *knockout* (Cai et al., 2006; Martin et al., 2010).

Estudos pioneiros em monocotiledôneas também confirmaram que miRNAs desempenham um papel na adaptação à seca. Zhao et al. (2007), examinando a expressão de miRNAs de arroz durante seca, observou que o MIR169g era induzido em resposta ao estresse, e esta indução era mais proeminente na raiz que na parte aérea. Estes autores analisaram o promotor do *MIR169g* e encontraram dois elementos responsivos à desidratação (DREs), sugerindo o papel do MIR169g no estresse hídrico.

Durante o estresse hídrico, plantas acumulam o osmoprotetor prolina como mecanismo de proteção. A prolina, por sua vez, é degradada por uma prolina desidrogenase (PDH) a qual é reprimida durante a seca e induzida em períodos de reidratação, sugerindo que ela desempenha importante papel na regulação dos níveis de prolina em plantas (Rayapati e Stewart, 1991). Neste cenário, identificou-se que o MIR474, o qual tem por alvo a PDH, foi induzido em milho submetido à seca (Wei et al., 2009). Desta forma, com menos PDH, a prolina acumula-se na planta e ajuda na proteção contra os danos causados pela deficiência hídrica.

Foi observado em tomate que a superexpressão de MIR169 conferiu aumento da tolerância à seca (Zhang et al., 2011). Na mesma condição de desidratação, as plantas não transgênicas demonstraram claros sintomas de desidratação, enquanto aquelas superexpressando MIR169 cresceram normalmente em situação de estresse hídrico. Análises fisiológicas demonstraram que os tomates transgênicos apresentaram redução na abertura estomática e decréscimo na taxa de transpiração, o que ocasionou uma prevenção na perda de água nas folhas destas plantas (Zhang et al., 2011).

Atualmente, o perfil de expressão de miRNAs sob estresse hídrico tem sido avaliado em um amplo número de espécies de plantas como: *A. thaliana* (Liu et al., 2008), *Oryza sativa* (Zhou et al., 2010), *Populus trichocarpa* (Lu et al., 2008a), *Phaseolus vulgaris* (Arenas-Huertero et al., 2009), *Medicago truncatula* (Trindade et al., 2010), *Triticum turgidum* (Kantar et al., 2010), entre muitas outras.

Em soja, o primeiro trabalho abordando a expressão de miRNAs em condições de estresse hídrico foi apresentado pelo nosso grupo (Kulcheski et al., 2011). Neste trabalho, foram construídas quatro bibliotecas de miRNAs provenientes de amostras de tecidos de raiz de cultivares sensíveis e tolerantes à seca. As análises de RT-qPCR de 11 novos miRNAs entre as quatro bibliotecas demonstrou que a maioria do miRNAs foi

induzido no genótipo sensível durante o estresse, enquanto o contrário foi observado no genótipo tolerante. Estes dados serão apresentados e discutidos detalhadamente no capítulo IV desta tese.

Alguns meses depois, Li et al. (2011) realizaram um estudo sobre miRNAs de soja em resposta à seca, salinidade e alcalinidade. Utilizando sequenciamento solexa, eles identificaram 133 miRNAs sendo expressos sob estas três condições, sendo que 71 deles foram unicamente expressos em seca, sugerindo que muitos miRNAs são induzíveis e são diferencialmente expressos em respostas a estresses específicos.

O mais recente trabalho sobre miRNAs relacionados à seca em soja foi desenvolvido por Ni et al. (2012). Neste estudo, MIR394a foi caracterizado quanto ao seu papel na tolerância à seca. Análise de expressão do MIR394a revelou que o mesmo foi diferencialmente expresso em vários tecidos e induzido por seca, além de outros estresses. O transcrito do locus *Glyma08g11030* foi identificado como alvo do MIR394a e validado via 5'RACE. A superexpressão do MIR394a resultou em plantas com redução na perda de água nas folhas e aumento na tolerância à seca. Além do mais, a superexpressão do *MIR394a* de soja em arabidopsis reduziu transcritos do gene *F-BOX* (*At1g27340*) devido o mesmo apresentar um sítio alvo de clivagem complementar ao MIR394a. Os autores sugerem que o gene *gma-MIR394a* funciona na modulação positiva da tolerância à seca (Ni et al., 2012).

1.5 Os microRNAs e a ferrugem asiática

O estudo de miRNAs em soja com ferrugem asiática não foi descrito por outro grupo que não o nosso até o presente momento. A identificação de miRNAs envolvidos na interação *G. max-P. pachyrhizi* foi um dos objetivos desta tese, não havendo qualquer trabalho prévio ou posterior ao experimento que foi baseado no

desenvolvimento de quatro bibliotecas de miRNAs de amostras provenientes de um genótipo suscetível e outro resistente infectados ou não com *P. pachyrhizi* (Kulcheski et al., 2011). Este trabalho está descrito no capítulo IV desta tese.

1.6 Os microRNAs em soja

Atualmente existem 555 miRNAs maduros, originados a partir de 506 sequências precursoras, identificados em *G. max* (miRBase database, release 19, http://www.mirbase.org/). Vale salientar que ao início do trabalho desenvolvido nesta tese, no ano de 2009, havia apenas 69 miRNAs maduros depositados no miRBase, demonstrando o rápido avanço que foi realizado nos últimos quatro anos na identificação e caracterização de miRNAs em soja.

A primeira identificação de miRNAs em soja foi relatada por Subramaniam e et al. em 2008. Neste estudo, foram identificados 35 novas famílias de miRNAs, investigando o papel dos miRNAs durante a simbiose com *Rhizobium*. No mesmo ano, Zhang et al. (2008), baseados em genômica comparativa, realizaram uma busca *in silico* de miRNAs em bancos de ESTs (*Expressed Sequence Tags*) de soja, e utilizaram PCR quantitativo para evidenciar 69 miRNAs pertencentes a 33 famílias. Um segundo estudo envolvendo miRNAs e nódulos em raízes de soja foi desenvolvido por Wang et al. (2009). Eles identificaram 32 miRNAs pertencentes a 11 famílias. A identificação de nove novos miRNAs em soja selvagem (*Glycine soja*) foi descrita por Chen et al. (2009). Outro estudo em quatro diferentes tecidos (raiz, semente, flor e nódulo) identificou 87 novos miRNAs (Joshi et al., 2010). Song et al. (2011) identificaram 26 novos miRNAs e seu genes alvos em análises de desenvolvimento de semente de soja.

Nesta sequência, em 2011 foi publicado o artigo apresentado no capítulo IV desta tese, onde estão descritos experimentos que permitiram identificar miRNAs

provenientes de bibliotecas de estresse hídrico e ferrugem asiática em soja (Kulcheski et al., 2011). Neste mesmo ano, três outros trabalhos abordando miRNAs envolvidos com estresses foram publicados. Guo et al. (2011) revelaram o perfil de expressão de miRNAs em soja envolvidos na resistência à *Phytophora sojae*, e Li et al. (2011) analisaram miRNAs associados a estresses abióticos (seca, salinidade e alcalinidade). Radwan et al. (2011) descreveram análises de hibridização de microarranjos de DNA medindo a abundância de transcritos em raízes de sojas resistentes e suscetíveis a *Fusarium viguliformes*. Neste estudo, além dos genes codificadores de proteínas, 42 microRNAs foram identificados. Destes, 29 miRNAs aumentaram e 10 diminuíram em abundância comparando o tecido infectado com o controle. Um destes miRNAs, pertencente à família conservada miR397, foi induzido durante a doença, e seus alvos, sete membros da família das lacases, tiveram transcritos tanto aumentados quanto reduzidos em abundância.

No ano seguinte, Turner et al. (2012) caracterizaram a organização de miRNAs no genoma da soja. Neste trabalho além da descoberta de cinco novas famílias de miRNAs, eles observaram que os genes de miRNAs no genoma da soja são em sua maioria intergênicos e uma pequena porcentagem são intragênicos, sugerindo uma potencial co-regulação com seus genes parentais. Diferenças no número e na orientação de genes de miRNAs duplicados em tandem entre locus ortólogos foram observados, indicando uma continua evolução e diversificação segundo os autores. Além disto, observaram que as famílias de miRNAs conservadas são frequentemente maiores em tamanho e produzem miRNAs maduros menos diversos que aqueles produzidos por famílias de miRNAs específicas de legumes ou de soja. Em adição, foi detectado que a maioria dos miRNAs conservados ou legumes-específicos produzem miRNAs maduros

de 21 nt com uma distribuição distinta de nucleotídeos e regulam um conjunto de mRNAs alvos mais conservados comparado a famílias de miRNAs específicas de soja.

O primeiro estudo sobre a regulação gênica via miRNAs nos diferentes estágios de desenvolvimento de semente de soja foi apresentado por Shamimuzzaman e Vodkin et al. (2012). Neste trabalho, os autores também utilizaram sequenciamento do degradoma para detectar alvos clivados por miRNAs.

A identificação de miRNAs em soja envolvidos com o nematóide do cisto da soja foi reportada por Li et al. (2012b). Sequenciando cultivares de soja resistentes e suscetíveis, os autores detectaram 364 miRNAs conhecidos de soja e 21 candidatos a novos miRNAs. Deste total 101 miRNAs pertencentes à 40 famílias foram responsivos à doença, e 20 diferencialmente expressos entre duas cultivares testadas.

Um estudo avaliando a população global de pequenos RNAs (smRNAs), incluindo miRNAs, foi realizado para comparar o padrão destes RNAs entre tecidos vegetais e sementes em desenvolvimento de soja (Zabala et al., 2012). As análises revelaram que existe uma prevalência quanto ao tamanho dos pequenos RNAs com relação aos diferentes tecidos analisados. SmRNAs de 24 nt foram prevalentes em tecidos de sementes comparado aos tecidos vegetais, sendo que smRNAs deste tamanho são conhecidos por derivar de elementos repetitivos, incluindo transposons.

O trabalho envolvendo miRNAs de soja relacionado ao estresse com alumínio foi realizado por Zeng el al. (2012). Nas bibliotecas desenvolvidas foram identificados 97 miRNAs conhecidos e 31 novos. Do total de miRNAs identificados, 30 foram responsivos ao estresse por alumínio. As análises de degradoma adicionalmente realizadas neste estudo identificaram mRNAs correspondentes às proteínas ARF (*auxin response factor*), NB-ARC (*domain-containing disease resistance protein*), LRR-TIR (*leucine-rich repeat and toll/interleukin-1 receptor-like protein*), ATPases

transportadoras de cátions, fatores transcricionais MYB e proteínas de meristema apical, sendo clivados por miRNAs em amostras sob estresse por alumínio.

O mais recente trabalho identificando novos miRNAs em soja foi baseado em bibliotecas de degradoma provenientes de diferentes tecidos de soja. Um total de 211 potenciais alvos de miRNAs, includindo 174 alvos de miRNAs conservados e 37 de miRNAs específicos de soja, além de 25 novas famílias de miRNAs foram detectadas. Interessantemente, 23 alvos de miRNAs produziram pequenos RNAs de interferência secundários (siRNA). Estes alvos foram pareados com cinco miRNAs: MIR393, MIR1508, MIR1510, MIR1514, e novel-11. Múltiplos alvos foram identificados para estes siRNAs secundários. Estes 23 genes alvos de miRNAs têm potencial para ser novos genes *TAS* (locus que produzem trans-acting siRNAs ou tasiRNA) em soja (Hu et al., 2012).

1.7 Regulação da expressão gênica por microRNAs

Embora os miRNAs tenham sido descobertos em plantas há pouco mais de uma década, estes pequenos RNAs têm sido apontados como alguns dos mais importantes dentre os reguladores gênicos. Investigações indicam que miRNAs controlam a expressão gênica de aproximadamente 30% dos genes codificadores de proteínas em humanos (Lewis et al., 2005). Em plantas, ainda não foi realizado qualquer estudo predizendo este percentual, mas também se acredita que eles estejam envolvidos em uma parcela significativa da regulação gênica vegetal.

Uma vez que o complexo funcional miR-RISC é formado, o miRNA guia a maquinaria protéica (a qual está associado) até seu alvo cognato. Em plantas, a ação dos miRNAs pode ser via clivagem ou inibição da tradução dos mRNAs alvos. Além disto, é sabido que em plantas ao contrário de animais, uma sequência de miRNA liga-se a um

número limitado de alvos; fato atribuído à alta complementaridade requerida entre eles (Naqvi et al., 2012).

A identificação dos mRNAs alvos de miRNAs é inicialmente realizada através de uma predição via ferramentas de bioinformática. A partir de uma sequência madura de miRNA, o sistema em uso, procura por potenciais sítios alvos de reconhecimento, com um número de erros toleráveis entre o miRNA e o potencial alvo. Por exemplo, o psRNATarget, um servidor para análises de alvos de miRNAs de plantas, realiza duas importantes análises: 1) complementaridade reversa entre miRNA e mRNA alvo através de um sistema de pontuação, e 2) avaliação quanto à acessibilidade do sítio alvo pelo cálculo de energia livre ("*unpaired energy*"-UPE) necessário para "abrir" a estrutura secundária em torno do sítio alvo do miRNA no mRNA (Dai e Zhao, 2011).

Após a predição computacional dos possíveis alvos de um miRNA, o passo seguinte é tentar a validação experimental da regulação. A análise de 5' RACE, que permite detectar a clivagem endonucleotídica no sítio de pareamento entre o miRNA e o suposto alvo, tem sido a abordagem prevalente e melhor sucedida nas validações experimentais (Llave et al., 2002). Outra forma de validação de alvos de miRNAs tem sido através do perfil de transcrição (bastante empregado em estudos de outras espécies que não arabidopsis). Além destes, o sequenciamento de degradoma é a mais recente estratégia para validar alvos de miRNA, esta técnica combina sequenciamento de última geração de RNA com análises de bioinformática. O sequenciamento de degradoma revela os alvos de miRNAs identificando os fragmentos originados por clivagem direcionada por smRNA através do sequenciamento da região 5' terminal desprovidas de *cap* (Addo-Quaye et al., 2008; Addo-Quaye et al., 2009).

Embora na teoria, a regulação via miRNAs aparente ser simples e direta, nos sistemas *in* vivo podemos observar que não é totalmente verdadeiro. Por exemplo, há

certos transcritos aos quais miRNAs se ligam mas que não são degradados, e desta forma permitem que outros alvos escapem à regulação mediada pelo respectivo miRNA, fenômeno conhecido por "*target mimicry*" (Franco-Zorrilla et al., 2007). Este é o caso do transcrito IPS1 (*Induced by Phosphate Starvation1*), que possui um sítio complementar ao MIR399 com uma incompatibilidade (erro) na região central (11–13 nt) fora da região "*seed*" (2–8 nts da extremidade 5' do miRNA maduro). Embora isto permita que o MIR399 se ligue à IPS1, ele é incapaz de clivar este alvo, mas ao mesmo tempo, MIR399 está colapsado à IPS1. Desta forma, PHO2, um alvo genuíno do MIR399, acaba escapando da supressão por MIR399 até mesmo nos casos onde este miRNA está presente em altos níveis. Quando a sequência de IPS1 foi mutada permitindo que MIR399 se ligasse com perfeita complementaridade, observou-se a degradação de IPS1 (Franco-Zorrilla et al., 2007). Estes resultados apoiam a ideia de que quando vários alvos estão presentes, alguns serão favorecidos em detrimento de outros, e tal fenômeno pode ser governado pela força com que o miRNA se liga ao seu alvo.

Outro caso que demonstra a dificuldade em revelar um alvo de miRNA estudando os perfis de expressão miRNA-alvo é o caso do MIR395. Foi demonstrado que o MIR395 tem três alvos: duas ATP sulforilases (APS1 e APS4) e, com uma menor afinidade, um transportador de sulfato (SULTR2;1) (Allen et al., 2005; Jones-Rhoades e Bartel, 2004; Jones-Rhoades et al., 2006). As APSs são importantes para a assimilação do enxofre (S) enquanto SULTR2;1 está envolvida na recuperação e translocação de S. Em um experimento de restrição de S em *A. thaliana*, foi observado que ocorria um aumento na expressão de MIR395 com uma observada redução de ambas APSs, mas simultaneamente era observado aumento de SULTR2;1. Na época, levantou-se a incógnita de como a indução do MIR395 poderia ser paralela à indução de um de seus

alvos. Contudo, posteriormente foi demonstrado que a aparente discrepância poderia ser explicada pela especificidade do tipo celular. Enquanto MIR395 é expresso em células companheiras do floema, o mRNA de SULTR2;1 acumula nas células do parênquima do xilema. A explicação é decorrente de um controle (repressão) de SULTR2;1 no floema a fim de confiná-la no xilema (Kawashima et al., 2009; Liang et al., 2010). Por este motivo, deve se ter atenção ao realizar as análises via parâmetros transcricionais uma vez que para um alvo ser regulado por um miRNA, ambos devem ser expressos em um mesmo tempo e espaço. Na via de regulação gênica por miRNAs é muito importante considerar o fato de que alvos podem escapar à regulação por miRNAs devido tanto à expressão em diferentes momentos como pela expressão em diferentes tecidos ou até células.

CAPÍTULO II

2.1 Objetivo geral

Nos últimos anos, avanços na área genômica têm permitido maior acesso e compreensão da estrutura e funcionalidade dos genomas de plantas. Neste cenário, com a descoberta dos miRNAs, abriu-se um novo universo de regulação gênica que, até então, não se supunha existir. Por este motivo, pelo presente trabalho objetivou-se identificar novos miRNAs de soja e caracterizar a expressão dos mesmos em relação aos processos de deficiência hídrica e FAS.

2.2 Objetivos específicos

- Avaliar o potencial dos miRNAs como genes de referência em análises de RT-qPCR;
- Construir bibliotecas de miRNAs e identificar novos miRNAs de soja;
- Identificar miRNAs diferencialmente expressos envolvidos em resposta a estresse hídrico;
- Identificar os miRNAs diferencialmente expressos envolvidos em resposta à FAS;
- Identificar e validar genes alvos dos miRNAs, cuja alteração de expressão esteja relacionada aos mecanismos moleculares de resposta aos estresses.

2.3 Hipótese científica

Devido ao prévio conhecimento de que os microRNAs são fatores essenciais na modulação de inúmeros processos biológicos em plantas, sendo que alguns já foram descritos em estudos de estresses bióticos e abióticos em outras espécies vegetais, será possível, através de técnicas amplamente utilizadas em estudos de miRNAs, identificar novos miRNAs em soja e avaliar se os mesmos são diferencialmente expressos entre distintos genótipos sob deficiência hídrica e de ferrugem asiática.

CAPÍTULO III

The use of microRNAs as reference genes for quantitative PCR in soybean

Analytical Biochemistry 406 (2010) 185–192

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean

Franceli Rodrigues Kulcheski [a], Francismar Correa Marcelino-Guimaraes [b], Alexandre Lima Nepomuceno [b], Ricardo Vilela Abdelnoor [b], Rogério Margis [a,*]

[a] Centre of Biotechnology, 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
[b] EMBRAPA Soja, Rodovia Carlos João Strass, Distrito de Warta, CEP 86001-970, Londrina, PR, Brazil

ARTICLE INFO

Article history:
Received 23 May 2010
Received in revised form 20 July 2010
Accepted 22 July 2010
Available online 27 July 2010

Keywords:
miRNAs
Reference genes
RT-qPCR
Soybean
Real-time PCR

ABSTRACT

Reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR) is a robust and widely applied technique used to investigate gene expression. However, for correct analysis and interpretation of results, the choice of a suitable gene to use as an internal control is a crucial factor. These genes, such as housekeeping genes, should have a constant expression level in different tissues and across different conditions. The advances in genome sequencing have provided high-throughput gene expression analysis and have contributed to the identification of new genes, including microRNAs (miRNAs). The miRNAs are fundamental regulatory genes of eukaryotic genomes, acting on several biological functions. In this study, miRNA expression stability was investigated in different soybean tissues and genotypes as well as after abiotic or biotic stress treatments. The present study represents the first investigation into the suitability of miRNAs as housekeeping genes in plants. The transcript stability of 10 miRNAs was compared to those of six previously reported housekeeping genes for the soybean. In this study, we provide evidence that the expression stabilities of miR156b and miR1520d were the highest across the soybean experiments. Furthermore, these miRNAs genes were more stable than the most commonly protein-coding genes used in soybean gene expression studies involving RT-qPCR.

Reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR)[1] is a powerful technique that is commonly used to study gene expression due to its high sensitivity, good reproducibility, and wide dynamic quantification range. Many experiments have used RT-qPCR to measure the gene expression variation between two different biological groups, for example, a treated versus a control pool of samples. However, in many cases, the variation is caused by a discrepancy in the initial sample amount, efficiency of nucleic acid recovery, RNA degradation, differences in sample quality, pipetting errors, or variation in cDNA synthesis efficiency [1]. In such cases, the normalization process is fundamental for correcting nonspecific variation that can affect quantification results [2].

Several normalization strategies have been proposed, but the most commonly used method is relative quantification. Using this method, the expression level of a gene of interest is normalized to the expression level of a reference gene [2]. An ideal reference gene should be expressed at a constant level across various conditions, such as developmental stages or tissue types, and its expression should be unaffected by experimental parameters [3–5]. Currently, the reference genes frequently used are protein-coding genes, especially "housekeeping genes," which are involved in basic cellular processes [6], such as cytoskeleton construction (actins), glycolysis (glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*)), protein folding (cyclophilin), synthesis of ribosomal subunits (rRNA), electron transport (succinate dehydrogenase complex, *SDH*), and protein degradation (ubiquitin) [7]. These genes are thought to have constant expression levels among different samples and are frequently used to normalize gene expression levels without proper validation [7]. However, the expression of a number of housekeeping genes, although constant under some experimental conditions, varies considerably under other conditions [6,8–11]. In fact, when these genes are used as reference genes under experimental conditions that differ from those conditions in which their stability of expression was originally tested, the results can be misinterpreted [11].

The advances in genome sequencing have provided high-throughput gene expression analysis and have contributed to the identification of a wide range of new gene products, such as the small noncoding RNAs, especially microRNAs (miRNAs). Discovered some decades ago, miRNAs are fundamental regulatory genes of eukaryotic genomes that regulate several biological functions

* Corresponding author. Fax: +55 51 3308 7309.
E-mail address: rogerio.margis@ufrgs.br (R. Margis).

[1] *Abbreviations used*: ASR, Asian soybean rust; miRNAs, microRNAs; RT-qPCR, reverse transcription quantitative real-time polymerase chain reaction.

0003-2697/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2010.07.020

including hormonal control, immune responses, and adaptation to a variety of biotic and abiotic stresses [12–15]. Although the small size of miRNA (19–24 nucleotides) creates challenges for their detection, recent innovative adaptations to RT-qPCR have resulted in improvements in gene expression profiling. The development of stem–loop primers provided a tool for detecting and characterizing mature miRNAs by PCR [16,17]. In addition, stem–loop primers can be used to specifically convert the mature functional miRNA into its DNA complement and can potentially be used for multiplex reverse transcription (RT) reactions [16]. Due to the advances in the ability to characterize miRNA expression, research involving these small RNAs has increased in recent years. However, experiments involving miRNA expression are still a challenge in many research areas due to the lack of proper control genes for normalizing these transcripts. Davoren and co-workers, while searching for suitable reference genes for the normalization of microRNA expression, identified two miRNAs (*Let-7a* and *miR-16*) constantly expressed in human breast cancer tissues [17]. Peltier and Latham [18] also searched for appropriated reference genes for their miRNA RT-qPCR studies and found miRNA expression levels were the most constant RNA levels in their analysis. They identified three miRNAs that were highly consistent in expression across 13 healthy tissues and 5 tumor tissues in humans. These miRNAs were statistically superior to the most commonly used reference RNAs used in miRNA RT-qPCR cancer experiments [18]. Galiveti et al. [19] also reported the detection of five non-protein-coding RNAs as appropriate housekeeping genes in human tissues. Currently, there are no data reporting the expression stability pattern of microRNAs in plant tissues.

Recognizing the potential of miRNAs as a reference gene in RT-qPCR analysis, we evaluated these non-protein-coding genes for use in normalizing gene expression in the soybean. This study is the first in which miRNAs were evaluated for stability alongside other mRNA genes and also tested as potential reference genes for both miRNA and mRNA gene expression in plant tissues. For this study, we selected the soybean (*Glycine max* Merryll L.), which is the major legume crop worldwide and already has established mRNA housekeeping genes [7,20,21]. We compared the expression level of 10 soybean miRNAs (*miR156a*, *miR156b*, *miR167ab*, *miR167c*, *miR171a*, *miR171b*, *miR172ab*, *miR396a*, *miR1520c*, and *miR1520d*) with six common mRNA housekeeping genes (*ACT*, *CDPK*, *CYP2*, *ELF1B*, *F-BOX*, and *TUA*) and found that many of the miRNAs showed better expression stability than the protein-coding housekeeping genes, indicating that these genes can be used as optimal reference genes for both miRNAs and protein-coding genes in RT-qPCR analysis.

Material and methods

Plant materials and treatments

Drought assay

For drought treatment, we used the soybean (*G. max* Merryll L.) cultivars 'Embrapa 48' as a drought-tolerant standard [22] and 'BR 16' as a sensitive standard. The plants were grown in a greenhouse at Embrapa-soybean in Londrina, Brazil, using a hydroponic system compound for plastic containers (30 liters) and an aerated 6.6 pH-balanced nutrient solution [22,23]. Briefly, seeds were pregerminated on moist filter paper in the dark at 25 ± 1 °C and 65 ± 5% relative humidity. Then, the plantlets were placed in polystyrene supports in such a way that the roots of the seedlings were completely immersed in the solution. Each tray containing the seedlings was maintained in a greenhouse at 25 ± 2 °C and 60 ± 5% relative humidity under natural daylight (photosynthetic photon flux density (PPFD) = 1.5×10^3 μmol m^{-2} s^{-1}, equivalent to 8.93×10^4 lx) and a 12 h day. After 15 days, seedlings with the first trifoliate leaf fully developed (V2 developmental stage) were submitted to different treatments. They were removed from the hydroponic solution and kept in a tray in the dark without nutrient solution or water for 0 min (T0 or control) and 125 min of stress (T125). The roots and leaves of both genotypes were analyzed as biological duplicate (T0) or triplicates (T125). They were immediately frozen in liquid nitrogen and stored at −80 °C until RNA extraction.

Asian soybean rust assay

Asian soybean rust (ASR) reaction was assessed after inoculation with a field population of *Phakopsora pachyrhizi* and performed in the greenhouse at Embrapa-soybean Londrina, Brazil. The soybean plants were grown in a pot-based system. The 'Embrapa 48' genotype was used as a susceptible host plant, which develops a susceptible lesion (TAN) after ASR infection [24], and 'PI561356' was used as the resistant host, which carries the ASR resistance mapped on linkage group G (personal communication). Urediniospores were collected by tapping infected leaves over a plastic tray and were then diluted in distilled water with 0.05% Tween 20 to a final concentration of 3×10^5 spores/mL. This suspension of spores was sprayed onto the plantlets at the V2 developmental stage. The same solution minus the spores was used for the mock inoculations. Following the ASR or mock inoculations, water-misted bags were placed over all plantlets for 1 day to aid the infection process and to prevent the cross-contamination of mock-infected plants. One trifoliolate leaf for each plant was collected at 192 h after inoculation for RNA extraction. Three biological replicates of each genotype were analyzed for both treatments.

Genotype and tissue assay

For the genotype assay, 'Embrapa 48', 'BR 16', and 'PI561356' were analyzed. Leaf samples under nonstress conditions from both the drought and the ARS assays at the same V2 developmental stage were used. The tissue assay was performed by analyzing the roots and leaves from the 'Embrapa 48' genotype under nonstress conditions and during the V2 developmental stage.

RNA extraction and cDNA synthesis

For each treatment, total RNA was isolated by extraction with Trizol, (Invitrogen) and the quality was evaluated by electrophoresis on a 1.0% agarose gel. The cDNA synthesis was carried out by multiplex technique [25,26] from approximately 2 μg of total RNA. Each reaction was primed with a pool of 0.5 μM 10 gene-specific stem–loop primers [16] (IDT) and 1 μM oligonucleotide dT24V (Invitrogen). Before transcription, RNA and primers were mixed with RNase-free water to a total volume of 10 μL and incubated at 70 °C for 5 min followed by ice-cooling. Then, 6 μL 5× RT-Buffer (Promega), 1 μL of 5 mM dNTP (Ludwig), and 1 μL MML-V RT Enzyme 200 U (Promega) were added for a final volume of 30 μL. The synthesis was performed at 42 °C for 30 min on a Veriti Thermal Cycler (Applied Biosystem), and inactivation of the enzyme was completed at 85 °C for 5 min. All cDNA samples were 50-fold diluted with RNase-free water before being used as a template in RT-qPCR analysis.

Selection of candidate housekeeping genes and primer design

The six protein-coding genes, *ELF1B*, *CYP2*, *ACT*, *TUA*, *F-BOX*, and *CDPK* (Table 1), were selected based on previous reports of housekeeping genes for the soybean [7,20,21]. The 10 miRNAs were chosen based on our previous gene expression studies in the soybean (data not shown). During the analyses of 43 miRNAs available on miRBase database at http://www.mirbase.org (release 12.0) for

Table 1
Primer sequences and amplicon characteristics for each of the 16 reference gene candidates.

Acronym	Forward primer sequence (5′–3′)	Reverse primer sequence (5′–3′)	Amplicon size (bp)	Efficiency	R^2	G.max loci[a]	Function	Locus accession number[b]
ACT	CGGTGGTTCTATCTTGGCATC	GTCTTTCGCTTCAATAACCCTA	142	1.89	0.998	Gm18:61274000-61274248	Cytoskeletal structural protein	BW652479
CDPK	TAAAGAGCACCATGCCTATCC	TGGTTATGTGAGCAGATGCAA	97	1.97	0.986	Gm10:46251505-46251601	CDPK-related protein kinase	AW396185
CYP2	CGGGACCAGTGTGCTTCTTCA	CCCCTCCACTACAAAGGCTCG	154	1.80	0.996	Gm12:1802441-1802594	Protein folding	CF806591
ELF1B	GTTGAAAAGCCAGGGGACA	TCTTACCCCTTGAGCGTGG	118	1.86	0.996	Gm02:49033741-49033961	Translational elongation	EV279336
F-BOX	AGATAGGGAAATGGTGCAGGT	CTAATGGCAATTGCAGCTCTC	93	1.88	0.983	Gm12:3676720-3677719	F-Box protein family	CD397253
TUA	AGGTCGGAAACTCCTGCTGG	AAGGTGTTGAAGGCGTCGTG	159	1.95	0.982	Gm05:34705808-34705919	Cytoskeletal structural protein	CA801144
156a	TGACAGAAGAGAGTGAGCAC	GTGCAGGGTCCGAGGT	70	1.94	0.992	Gm17: 6149963-6150084	miRNA	MIMAT0001686
156b	TGACAGAAGAGAGAGCACA	"	71	2.02	0.992	Gm14: 990334-990453	miRNA	MIMAT0001692
167ab	TGAAGCTGCCAGCATGATCTA	"	71	2.00	0.992	Gm19: 41871231-41871349 Gm02: 14635614-14635734	miRNA	MIMAT0001679 MIMAT0001680
167c	TGAAGCTGCCAGCATGATCTG	"	71	1.92	0.998	Gm07: 39778512-39778886	miRNA	MIMAT0007355
171a	TGAGCCGTGCCAATATCACGA	"	71	2.00	0.997	Gm12: 8443106-8443207	miRNA	MIMAT0007358
171b	CGAGCCGAATCAATATCACTC	"	71	1.91	0.999	Gm04: 46988579-46988670	miRNA	MIMAT0007363
172ab	AGAATCTTGATGATGCTGCAT	"	71	2.00	0.979	Gm12: 6110704-6110862 Gm13: 40401673-40401825	miRNA	MIMAT0001682 MIMAT0001683
396a	TTCCACAGCTTTCTTGAACTG	"	71	1.90	0.999	Gm13: 26338134-26338273	miRNA	MIMAT0001687
1520c	TTCAATAAGAACGTGACACGTGA	"	73	1.97	0.992	Gm17: 37893185-37893324	miRNA	MIMAT0007395
1520d	ATCAGAACATGACACGTGACAA	"	72	2.00	0.996	Gm07: 119451-119561	miRNA	MIMAT0007379

[a] Position of CDS or pre-miRNA sequences.
[b] NCBI or miRBase database accession number.

RT-qPCR analysis in a drought-stress assay, we observed that *miR156a*, *miR156b*, *miR167ab*, *miR167c*, *miR171a*, *miR171b*, *miR172-ab*, *miR396a*, *miR1520c*, and *miR1520d* had uniform expression along a range of different conditions (unpublished data).

All primers of the 16 candidate reference genes are listed in Table 1. Primer sequences for the six mRNA housekeeping genes were chosen based on current literature [7,20]. The miRNA primers were designed based on the fully tested miRNA sequence (forward), and the reverse primer was the universal reverse primer for miRNA [16]. The stem–loop primers, used for miRNAs cDNA synthesis, were designed according to Chen et al. [16]. The stem–loop sequence consists of 44 conserved and 6 variable nucleotides that are specific to the 3′ end of the miRNA sequence (5′ GTCGTATCCAGTGCAGGGT CCGAGGTATTCGCACTGGATACGAC NNNNNN 3′).

Reverse transcription quantitative PCR

All quantitative PCR were performed in an ABI 7500 Real-Time PCR System (Applied Biosystem) using SYBR Green I (Invitrogen) to detect double-strand cDNA synthesis. Reactions were done in a volume of 24 μL containing 12 μL of diluted cDNA (1:50), 1× SYBR Green I (Invitrogen), 0.025 mM dNTP, 1× PCR Buffer, 3 mM $MgCl_2$, 0.25 U Platinum Taq DNA Polymerase (Invitrogen), and 200 nM of each reverse and forward primers. In all miRNA RT-qPCR the Universal Reverse primer (5′ GTGCAGGGTCCGAGGT 3′) was used. All samples were analyzed as technical triplicates with a no-template control also included. The conditions were set as the following: an initial polymerase activation step for 5 min at 94 °C, 40 cycles of 15 s at 94 °C for denaturation, 10 s at 60 °C for annealing, and 25 s at 72 °C for elongation. At the end of cycling protocol, a melting-curve analysis was included and fluorescence measured from 65 to 99 °C. Threshold and baselines were manually determined using the ABI 7500 Real-Time PCR System SDS Software v2.0.

Data analysis

Primer efficiency (E) and correlation coefficient (R^2) were calculated using SDS software (ABI 7500 Real-Time PCR System v. 2.0) based on a standard curve generated using a twofold dilution series of one sample over five dilution points that were measured in duplicate. PCR amplification efficiencies were calculated for each candidate endogenous control with the formula $E = 10^{-1/\text{slope}}$, using the slope of the plot, C_t versus log input of cDNA. The stability of each candidate gene expression was analyzed using geNorm [27] and NormFinder [28] software. For these analyses, the threshold cycle (C_t) values were converted to quantities via the comparative C_t method. Both geNorm and NormFinder tools were used to determine the most stable reference genes. GeNorm software was also used to determine the optimal number of genes required for normalization. It starts calculating the pairwise variation (V) $V2/3$ between the NF2 (including the two most stable reference genes) and the normalization factor NF3 (including the three most stable reference genes). Then, it performs a stepwise calculation of the $Vn/n+1$ between the NFn and the NF$n+1$.

Reference gene validation

To determine how the use of different reference genes can affect the normalization of the expression data for a gene of interest, we calculated the significant difference in the mean expression of two target genes between drought-stressed and control subgroups. We selected the *miR1513* (5′ CGCCTGAGAGAAAGCCATGACTTAC 3′) as a miRNA target gene and the *CDPK* as a protein-coding or mRNA target gene. The expressions of the two target genes were normalized using three different strategies: (1) with all candidate reference genes individually; (2) with the two most stable reference genes and also with the two most stable mRNA genes selected by NormFinder; (3) with the two and three most stable reference genes, and also with the two and three most stable mRNA reference genes according to geNorm software. To calculate the relative expression of the two target genes, we used the $2^{-\Delta\Delta Ct}$ method [29]. Student's t test was performed to compare pairwise differences in expression. The parameters of two-tailed distribution and two samples assuming unequal variances were established. The means were considered significantly different when $P < 0.05$.

Results and discussion

PCR efficiency and amplification specificity

The amplification efficiency for each primer pair was determined in an RT-qPCR assay using duplicates of a twofold dilution series from a cDNA template. Primer efficiency indicates the amplicon doubling rate of a specific primer pair during a PCR. When the efficiency is 100% or 2 (1 + efficiency = 2), it indicates that the cDNA target is duplicated at every PCR cycle during the exponential phase. The efficiency values of the 16 candidate reference genes ranged from 1.80 to 2.02, as listed in Table 1. About 75% of all primer pairs were in the range of 1.9–2.02, reflecting the high quality of the PCR. Amplification of the specific transcript was confirmed by the appearance of a single peak in the melting-curve analysis following completion of the amplification reaction (in Supplementary Data). Furthermore, the correlation coefficient (R^2) was calculated. The R^2 value indicates the quality of the fit of the standard curve to the plotted data points. All primer pairs showed an $R^2 > 0.98$ (Table 1), indicating a strong relationship between the detected C_t values and the corresponding relative amount of template in all the amplification reactions.

Expression stability of candidate reference genes

To investigate the suitability of 10 miRNAs as reference genes, we analyzed their expression stability along with 6 housekeeping genes commonly employed in soybean studies involving RT-qPCR analysis. The expression stability of these 16 genes was analyzed based on five different data sets: (1) all 31 samples including different tissues, cultivars, abiotic treatments, and biotic treatments; (2) different tissues; (3) different genotypes; (4) abiotic stress consisting of drought-stress; and (5) biotic stress relative to samples from plants infected with ARS.

GeNorm analysis

The program geNorm uses an algorithm to calculate an *M* value, which is a gene expression stability measurement defined as the mean pairwise variation for a given gene compared to the remaining tested genes. Genes with the lowest *M* value indicate the most stable expression, whereas the highest *M* values indicate the least stable expression [27]. When we analyzed all 31 samples (Fig. 1A), the average expression stability values (*M*) of *miR156b* and *miR1520d* were lower than those of the other miRNAs, indicating that they are the most stable candidate genes. Interestingly, we observed that, in this set of analysis, all miRNA genes were more stable than mRNA housekeeping genes. Tubulin-A (*TUA*) was the most unstable gene with an *M* value over 1.2. Similar results were found when the sample sets of different tissues (Fig. 1B), different genotypes (Fig. 1C), and different abiotic conditions (Fig. 1D) were analyzed. The mean expressions of *miR156b* and *miR1520d* were the most stable out of all tested reference genes. In most cases, the miRNAs were the most stable genes, except for *miR396a*, which was less stable than mRNA housekeeping genes in the different tissues and genotypes data sets. In the biotic stress data set, the lowest *M* value was the *miR156a* and *miR156b* pair (Fig. 1E). The *miR1520d* was the third most stable candidate. Although in the first four analyses, *TUA* (Fig. 1A–C) and *CDPK* (Fig. 1D) mRNA genes were the most variable, the *miR167c* and *miR171b* were the most unstable genes for the biotic stress panel (Fig. 1E). The results also indicate that, generally, the miRNAs are more stable than the mRNAs tested (Fig. 1), but not always. The *miR171b*, which was the fourth most stable gene in tissue, genotypic, and abiotic approaches, showed low stability in the biotic assay. Considering the performance of the mRNA housekeeping genes, we compared our results with previous studies. Jian et al. [7] reported that *ELF1B* and *CYP2* were the most stable genes among different soybean tissues. Although *CYP2* had the lowest stability value of all mRNA candidates, *ELF1B* was the second most variable gene in our tissue assay. Furthermore, nine of the miRNA genes were more stable than the *CYP2* gene after geNorm analysis (Fig. 1A). In another study, *F-BOX* was described as the most stable gene among different tissues. *F-BOX* and *CDPK* were the most stable genes in samples infected with ASR and control [20]. However, in our results, for tissue and ASR data sets, both mRNA genes exhibited higher *M* values compared with several miRNA candidates (Fig. 1A and F). However, in all five analyses, the miRNAs were among the most stable genes indicated by the geNorm.

Although the stability analysis based on *M*-value points determines the most stable genes, Vandesompele et al. [27] recommended that in order to measure expression levels accurately, a normalization method using multiple housekeeping genes instead of one is required. To account for this recommendation, geNorm also computes a normalization factor (NF) and estimates the optimal number of reference genes according to the pairwise variation (*V*). A pairwise variation of 0.15 is accepted as a cutoff, below which the inclusion of an additional control gene is not required for reliable normalization [27]. To determine the optimal number of internal control genes for normalization, we calculated the pairwise variation for all five data sets (Fig. 1F). The results indicate that, in all approaches, the combination of the two most stable genes would be sufficient for normalization purposes because the *V*2/3 value was lower than 0.15 (Fig. 1F). Based on this graph, we can conclude that the addition of a third, fourth, or fifth reference gene does not cause a remarkable decrease in the *V* value. Additionally, using a combination of the seven most stable genes, the *V* value actually increases slightly for the tissue, genotypic, abiotic, and biotic assays but still remains below the cutoff value (0.15). In our analyses, the combination of *miR156b* and *miR1520d* was appropriate for comparisons of all sample sets ($V < 0.11$), different tissues ($V < 0.08$), contrasting genotypes ($V < 0.09$), and also abiotic stress treatment ($V < 0.06$), whereas the *miR156a* and *miR156b* pair was suitable for biotic treatment ($V < 0.04$).

NormFinder analysis

Stability of expression was also analyzed using the program NormFinder. Its strategy is based on a mathematical model of gene expression that enables an estimation of the intra- and intergroup variation, which are then combined into a stability value [28]. Candidate control genes with the minimal intra- and intergroup variation will have the lowest stability value and will be ranked at the top. Using this program, we analyzed five sets in a similar manner as in the geNorm analysis: (1) all 31 sample sets with two different analyses; all samples with no subgroups and another where groups were divided into stressed and nonstressed subgroups; (2) tissue sets were subgrouped into root and leaf; (3) genotype sets were divided into three different cultivar subgroups; (4) abiotic sets had drought and control subgroups, and (5) biotic stress sets were grouped into ARS infection and noninfection subgroups (Table 2; Supplementary File 1). Interestingly, the *miR156a* was the most stable gene in four out of the five data sets, including both subgroups investigated in each sample data set, and it was the second most stable gene in the biotic stress group. *TUA* was among the three genes with the worst stability in all the data sets. *CDPK* was the least stable in the abiotic set, and *miR167c* and *miR171b*, such as *miR1520c*, were the most variable genes in the biotic and abiotic stress sets, respectively. Interestingly, these results are similar to those obtained by geNorm. Despite a visible variation in the rankings of the other genes generated by geNorm and NormFinder algorithms, in general, all analyzed sets showed a marked separation

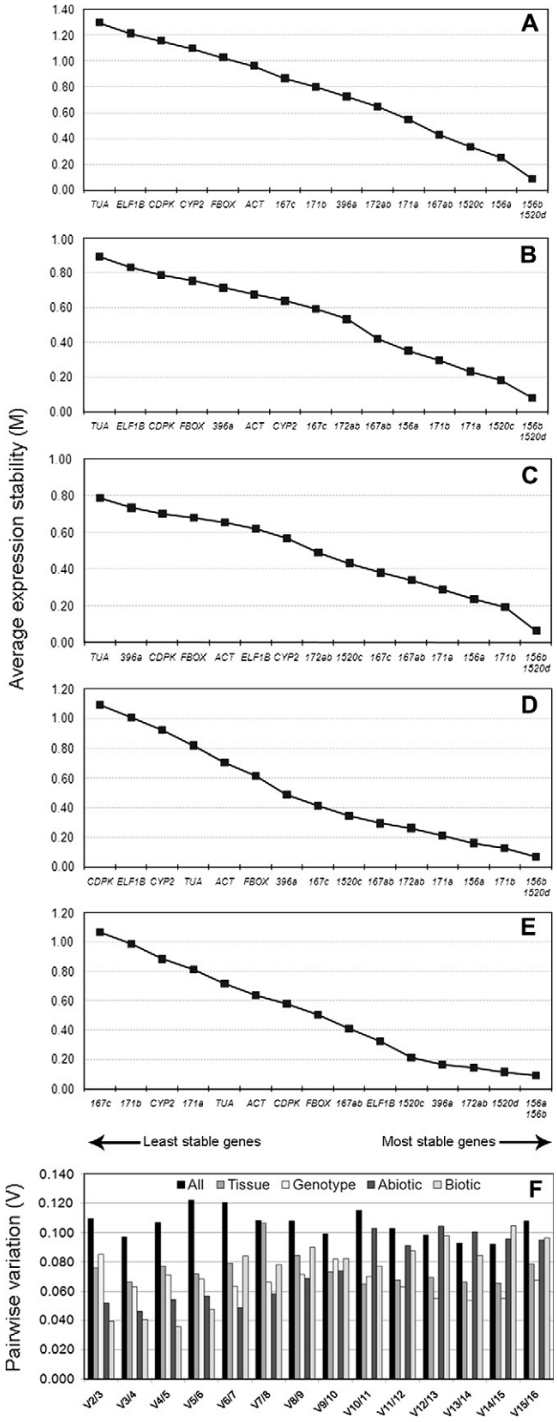


Fig. 1. Average expression stability values (*M*) of candidate reference genes after geNorm analysis: (A) all 31 samples including different tissues, cultivars, abiotic and biotic treatments; (B) different tissues; (C) different genotypes; (D) abiotic stress, consisting of drought stress and a control; and (E) biotic stress, relative to samples infected with ARS and noninfected (control); (F) determination of the number of housekeeping genes for normalization in each of the situations by geNorm analysis.

between the miRNA and the mRNA genes. As observed in the geNorm output, the miRNA genes were more stable compared to the mRNA candidates. Again, the best combination for the normalization of gene expression was two miRNA genes (Table 2).

Validation of putative reference genes

In order to demonstrate the suitability of the putative reference genes under investigation, we monitored the difference in the mean expression of two target genes from the drought-stress subgroup on normalization with different control genes. We decided to analyze two distinct classes of genes: one miRNA and one mRNA. Because there is currently no published data about miRNA expression in soybean, we decided to select the *miR1513* based on our previous RT-qPCR screening of the 43 soybean miRNAs from miRBase release 12.0. We observed that *miR1513* expression was clearly downregulated in soybean plants under drought conditions (unpublished data). The other target gene was *CDPK* (calcium-dependent protein kinases). This gene was one of the candidate reference genes; however, it was ranked as the least stable gene by geNorm and NormFinder analysis during drought-stress. Also, *CDPK* has been associated with drought-stress in plants. For example, studies with Arabidopsis [30,31] and rice [32] have demonstrated that this gene is overexpressed in drought situations.

Case 1: miRNA as target gene

In the first round of analyses, the target *miR1513* was individually normalized by each of the 16 candidate reference genes. A significant difference in gene expression under drought conditions compared to control samples was only determined with normalization based on *miR172ab*, *miR1520d*, *miR156b*, *miR171a*, and *miR156a* (Table 3). In the six cases where the target gene was normalized with an mRNA candidate reference gene, no significant difference was found between the compared situations (Table 3). In the next analysis, we normalized the target gene with *miR156a* and *miR171a*, the best pair for normalization comparison as indicated by NormFinder, and the differences between the subgroups, using miRNAs as reference genes, remained significant. We also selected the two most stable mRNA genes, *F-BOX* and *ACT*, following the NormFinder ranking and combined both with *miR1513* for normalization, and even so, no significant difference was observed between drought and control samples (Table 3).

Similar results were achieved when *miR1513* was normalized using either the two (*miR156b* and *miR1520d*) or three (*miR156b*, *miR1520d*, *miR171b*) best reference genes determined by geNorm. Although the top two reference genes elected by geNorm and NormFinder were not the same, both combinations were suitable for the normalization of the target gene. The mean-fold expression of *miR1513* under drought-stress conditions was significantly reduced to similar levels (0.4) with either *miR156a* and *miR171a* (NormFinder) or *miR156b* and *miR1520d* (geNorm) standardization (Table 3) compared to control samples.

We also selected the two most stable mRNA genes according to the ranking of NormFinder and geNorm for the drought-stress group. The *F-BOX* and *ACT* were determined to be the two best mRNA reference genes by both programs; however, when we normalized the target *miR1513* with these two reference genes, the difference in *miR1513* expression between the drought-stressed samples and the control samples was not significant. The same result was obtained when we added the *TUA* reference gene to these analyses. *TUA* was the third best mRNA reference gene by geNorm analysis. Again, no significant alteration in gene expression was determined when we normalized the miRNA target gene with three mRNA reference genes (Table 3).

In our investigations, we identified that the top ranked housekeeping genes by NormFinder (*miR156a* and *miR171a*) and geN-

Table 2
Ranking of candidate reference genes in order of their expression stability calculated by NormFinder.

Ranking order	All samples set		Tissues set (leaf::root)	Genotypes set (EM::BR::PI)	Abiotic stress set (drought::control)	Biotic stress set (ASR::control)
	Non-subgroups	(stress::control)				
1	*156a*	*156a*	*156a*	*156a*	*156b*	*1520c*
2	*167ab*	*172ab*	*167ab*	*167c*	*1520d*	*156a*
3	*396a*	*1520d*	*171b*	*1520d*	*156a*	*167ab*
4	*172ab*	*156b*	*CYP2*	*156b*	*171b*	*172ab*
5	*156b*	*167ab*	*396a*	*171b*	*171a*	*156b*
6	*1520d*	*396a*	*156b*	*ELF1B*	*172ab*	*1520d*
7	*1520c*	*1520c*	*ACT*	*167ab*	*167c*	*396a*
8	*ACT*	*ACT*	*1520d*	*171a*	*167ab*	*ELF1B*
9	*F-BOX*	*F-BOX*	*F-BOX*	*F-BOX*	*396a*	*CYP2*
10	*171a*	*167c*	*1520c*	*CYP2*	*1520c*	*F-BOX*
11	*167c*	*171b*	*CDPK*	*CDPK*	*F-BOX*	*CDPK*
12	*171b*	*CYP2*	*172ab*	*172ab*	*ACT*	*ACT*
13	*CYP2*	*171a*	*171a*	*ACT*	*CYP2*	*171a*
14	*ELF1B*	*ELF1B*	*167c*	*1520c*	*ELF1B*	*TUA*
15	*CDPK*	*TUA*	*ELF1B*	*TUA*	*TUA*	*167c*
16	*TUA*	*CDPK*	*TUA*	*396a*	*CDPK*	*171b*
Best combination of 2 genes		*156a, 1520d*	*167ab, 171b*	*156a, 1520d*	*156a, 171a*	*172ab, 1520c*

Note: The subgroups are indicated inside the parentheses; mRNA genes are in bold.

Table 3
Differential relative expression of the *miR1513* target gene when normalized individually with 16 different candidate housekeeping genes and when normalized by combining the most stable genes according to the NormFinder and geNorm analyses in abiotic stress treatment.

Normalizer	Mean fold change in gene expression				*P* value[b]
	Ctrl	*SE*[a]	Stress	*SE*[a]	
172ab	1.00	0.01	0.38	0.01	1.45E−04
1520d	1.00	0.00	0.39	0.02	0.002
156b	1.00	0.03	0.38	0.02	0.013
171a	1.00	0.05	0.45	0.09	0.026
156a	1.00	0.04	0.35	0.01	0.037
167c	1.00	0.07	0.33	0.03	0.068
171b	1.01	0.08	0.37	0.03	0.076
396a	1.01	0.10	0.37	0.18	0.096
1520c	1.01	0.08	0.37	0.01	0.102
CYP2	1.02	0.14	0.12	0.04	0.115
167ab	1.02	0.13	0.25	0.02	0.137
CDPK	1.24	0.52	0.04	0.01	0.349
TUA	1.03	0.18	2.82	1.94	0.530
ACT	1.05	0.23	0.84	0.63	0.817
ELF1B	1.03	0.17	1.14	0.86	0.926
F-BOX	1.04	0.21	0.99	0.74	0.958
NormFinder					
156a, 171a	1.00	0.03	0.40	0.05	1.8E−05
F-BOX, ACT	1.05	0.16	0.91	0.49	0.819
GeNorm					
156b, 1520d	1.0	0.02	0.4	0.02	5.9E−08
156b,1520d, 171b	1.0	0.03	0.4	0.01	1.3E−07
F-BOX, ACT	1.0	0.16	0.9	0.49	0.819
F-BOX, ACT, TUA	1.0	0.12	1.6	0.78	0.562

[a] *SE* (Standard Error).
[b] *P* values were calculated using the *t*-test, significance level $P < 0.05$.

orm (*miR156b*, *miR1520d*, *miR171b*) provided an accurate standardization for *miR1513* expression. When the target was normalized with either the two or three best reference genes based on our analysis, a difference in expression could be detected between stressed and control groups, in accordance with previously published data that has shown that *miR1513* is down-regulated under drought-stress conditions. However, when the miRNA target was normalized using *F-BOX* and *ACT* as well as *F-BOX*, *ACT* and *TUA*, which were the most stable among the mRNA reference genes, no difference in expression was detected for *miR1513* between the drought-stressed and the control groups. If these mRNA housekeeping genes were used as reference genes for *miR1513*, without proper investigation, the results could be misinterpreted.

Case 2: mRNA as target gene

In the first set of analyses, the target gene *CDPK* was normalized individually with the remaining 15 housekeeping genes. We observed that with all individual normalizations, except with *ELF1B*, *TUA*, and *CYP2*, *CDPK* expression was significantly different between the drought and the control situation. Although the *CDPK* expression was very distinct between drought and control samples when normalized with *ELF1*, *TUA*, and *CYP2* genes, the values of the standard error (SE) were high, which explains why these marked differences are not statistically supported. In the second set, the *CDPK* expression was normalized with the two most stable genes (*miR156a* and *miR171a*) following the NormFinder ranking (Table 4). The expression of *CDPK* was also investigated using the two most stable mRNA genes identified in our study (*F-BOX* and *ACT*). In analyses, using miRNAs or mRNA as reference genes, the differences in target expression were significant between the stressed and the control groups. The third set of analyses was done based on the most stable genes selected by geNorm. *CDPK* expression was significantly different between drought and control when the target was normalized with the two (*miR156b* and *miR1520d*) and the three (*miR156b*, *miR1520d*, and *miR171b*) most stable genes.

Similar result was obtained when the two (*F-BOX* and *ACT*) and the three (*F-BOX*, *ACT*, and *TUA*) top-ranked mRNA genes were used for normalization (Table 4). These results show that normalizing with one protein-coding gene along with miRNAs or mRNA genes produces consistent results. It has been reported that the *CDPK* gene has increased expression during drought-stress [30–32]. In our research, we observed that this gene showed a fivefold increase in expression in plants under drought conditions even when expression was normalized using the two best miRNA or the two best mRNA housekeeping genes. From these findings, we conclude that miRNAs can be used as optimal reference genes not only for other miRNAs but also for protein-coding genes.

Table 4
Differential relative expression of the *CDPK* target gene when normalized individually with 15 different candidates housekeeping genes and when normalized by combining the most stable genes according to the NormFinder and geNorm analyses in abiotic stress treatment.

Normalizer	Mean fold change in gene expression				*P value*[b]
	Ctrl	*SE*[a]	Stress	*SE*[a]	
1520d	1.03	0.23	5.62	0.27	0.001
F-BOX	1.00	0.10	4.57	0.22	0.001
156b	1.03	0.26	6.14	0.46	0.003
171b	1.03	0.25	6.00	0.45	0.003
171a	1.03	0.23	5.95	0.48	0.004
156a	1.04	0.28	5.10	0.42	0.004
ACT	1.00	0.06	5.62	0.29	0.004
167c	1.03	0.23	3.84	0.23	0.005
1520c	1.00	0.06	5.41	0.43	0.008
396a	1.03	0.26	4.39	0.16	0.012
167ab	1.08	0.41	4.32	0.34	0.019
172ab	1.03	0.24	5.04	0.70	0.021
ELF1B	1.00	0.01	18.79	5.31	0.079
TUA	1.00	0.04	16.74	6.43	0.134
CYP2	1.00	0.01	10.08	6.22	0.282
NormFinder					
156a, 171a	1.03	0.13	5.52	0.31	9.4E−06
F-BOX, act	1.00	0.05	5.09	0.28	2.8E−05
GeNorm					
156b, 1520d	1.03	0.12	5.88	0.24	5.4E−07
156b,1520d, 171b	1.03	0.10	5.92	0.20	2.5E−10
F-BOX, ACT	1.00	0.05	5.09	0.28	2.8E−05
F-BOX, ACT, TUA	1.00	0.03	8.97	2.54	0.018

[a] *SE* (Standard Error).
[b] *P* value were calculated using the *f*-test, significance level $P < 0.05$).

Final considerations

In this study, we observed a marked difference between the expression stability of miRNA and mRNA candidate housekeeping genes. In general, miRNAs were the most stable genes across the five different sets analyzed by geNorm (Fig. 1A–E) and NormFinder (Table 2). Although the optimal combinations of genes selected by geNorm and NormFinder were not the same, both analyses did select the same class of genes as the most stably expressed genes. These differences in the ranking were previously reported [28] and thus affirm that the discrepancies observed in the NormFinder versus the geNorm results were caused by the differences between these two approaches. In conclusion, we provide evidence that miRNAs can have better expression stability than protein-coding genes. In addition, we demonstrate that microRNAs are optimal reference genes not only for other miRNAs but also for protein-coding genes in RT-qPCR analysis.

Ideally, a reference gene for quantitative gene expression studies should not be influenced by the experimental conditions, type of tissues, or developmental stages. In our work, we have found that miRNAs genes are, in general, more stable than the protein-coding genes. Analyses of the best reference gene among different classes of RNAs (mRNA, snRNA, and miRNA) were previously done using different human tissues [19] and comparing normal and malignant breast tissues [17]. As in our results, miRNAs have been shown to have a more stable expression than the other classes of RNAs. The reason why miRNAs expression is more stable than protein-coding genes remains unanswered.

Despite the use of miRNAs as reference genes in our and other studies, it must be considered that not all miRNAs have a constitutive expression among tissues and under stressing conditions. Indeed, several of them are modulated by environmental and hormonal conditions and are implicated in cellular and tissue differentiation. At the same time, if we consider a miRNA showing a biological function during a very narrow developmental stage [33], its expression may remain stable in different organs and under several stressing conditions, presenting considerable variation just in a specific situation associated to its biological role.

It is important to stress that the use of miRNAs as reference genes in RT-qPCR-derived expression analysis of other miRNAs or even protein-coding genes does not impose technical difficulties or costs, since miRNA-specific primers can be mixed with the standard poli-T primer in the reverse transcription reaction.

Analyses in other plant and animal systems need to be undertaken to demonstrate the universality of the present results. Nonetheless, the evolutionary conservation of diverse miRNA families among distinct plant taxa suggests that the genes that encode miRNAs can be used as reference genes in place of the traditionally used protein-coding genes.

Acknowledgments

F.R.K. and R.M. were sponsored by Ph.D. and research grants (140578/2009-9 and 303967/2008-0) from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil. This work was financially supported by Biotecsur, GenoSoja consortium (CNPq 552735/2007-8) and GenoProt (CNPq 559636/2009-1).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ab.2010.07.020.

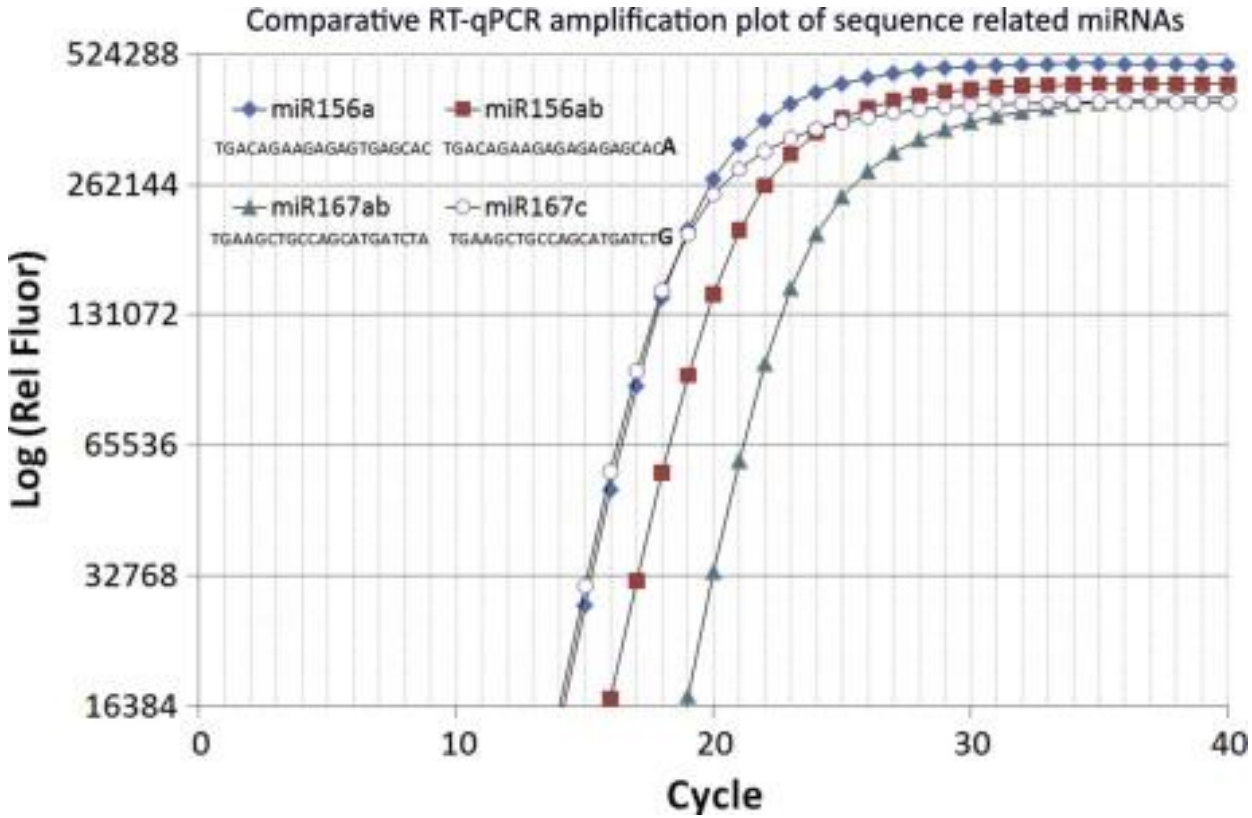
References

[1] S.A. Bustin, Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems, J. Mol. Endocrinol. 29 (2002) 23–39.
[2] A.M. Brunner, I.A. Yakovlev, S.H. Strauss, Validating internal controls for quantitative plant gene expression studies, BMC Plant Biol. 4 (2004) 14.
[3] O. Thellin, W. Zorzi, B. Lakaye, B. De Borman, B. Coumans, G. Hennen, T. Grisar, A. Igout, E. Heinen, Housekeeping genes as internal standards: use and limits, J. Biotechnol. 75 (1999) 291–295.
[4] T.D. Schmittgen, B.A. Zakrajsek, Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR, J. Biochem. Biophys. Methods 46 (2000) 69–81.
[5] Z. Tong, Z. Gao, F. Wang, J. Zhou, Z. Zhang, Selection of reliable reference genes for gene expression studies in peach using real-time PCR, BMC Mol. Biol. 10 (2009) 71.
[6] T. Remans, K. Smeets, K. Opdenakker, D. Mathijsen, J. Vangronsveld, A. Cuypers, Normalisation of real-time RT-PCR gene expression measurements in *Arabidopsis thaliana* exposed to increased metal concentrations, Planta 227 (2008) 1343–1349.
[7] B. Jian, B. Liu, Y. Bi, W. Hou, C. Wu, T. Han, Validation of internal control for gene expression study in soybean by quantitative real-time PCR, BMC Mol. Biol. 9 (2008) 59.
[8] N. Nicot, J.F. Hausman, L. Hoffmann, D. Evers, Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress, J. Exp. Bot. 56 (2005) 2907–2914.
[9] R.A. Volkov, I.I. Panchuk, F. Schoffl, Heat-stress-dependency and developmental modulation of gene expression: the potential of housekeeping genes as internal standards in mRNA expression profiling using real-time RT-PCR, J. Exp. Bot. 54 (2003) 2343–2349.
[10] T. Czechowski, M. Stitt, T. Altmann, M.K. Udvardi, W.R. Scheible, Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis, Plant Physiol. 139 (2005) 5–17.
[11] L. Gutierrez, M. Mauriat, J. Pelloux, C. Bellini, O. Van Wuytswinkel, Towards a systematic validation of references in real-time RT-PCR, Plant Cell 20 (2008) 1734–1735.
[12] I. Pedersen, M. David, MicroRNAs in the immune response, Cytokine 43 (2008) 391–394.
[13] R. Sunkar, V. Chinnusamy, J. Zhu, J.K. Zhu, Small RNAs as big players in plant abiotic stress responses and nutrient deprivation, Trends Plant Sci. 12 (2007) 301–309.
[14] O. Voinnet, Origin, biogenesis, and activity of plant microRNAs, Cell 136 (2009) 669–687.
[15] O. Voinnet, Post-transcriptional RNA silencing in plant-microbe interactions: a touch of robustness and versatility, Curr. Opin. Plant Biol. 11 (2008) 464–470.
[16] C. Chen, D.A. Ridzon, A.J. Broomer, Z. Zhou, D.H. Lee, J.T. Nguyen, M. Barbisin, N.L. Xu, V.R. Mahuvakar, M.R. Andersen, K.Q. Lao, K.J. Livak, K.J. Guegler, Real-time quantification of microRNAs by stem-loop RT-PCR, Nucleic Acids Res. 33 (2005) e179.

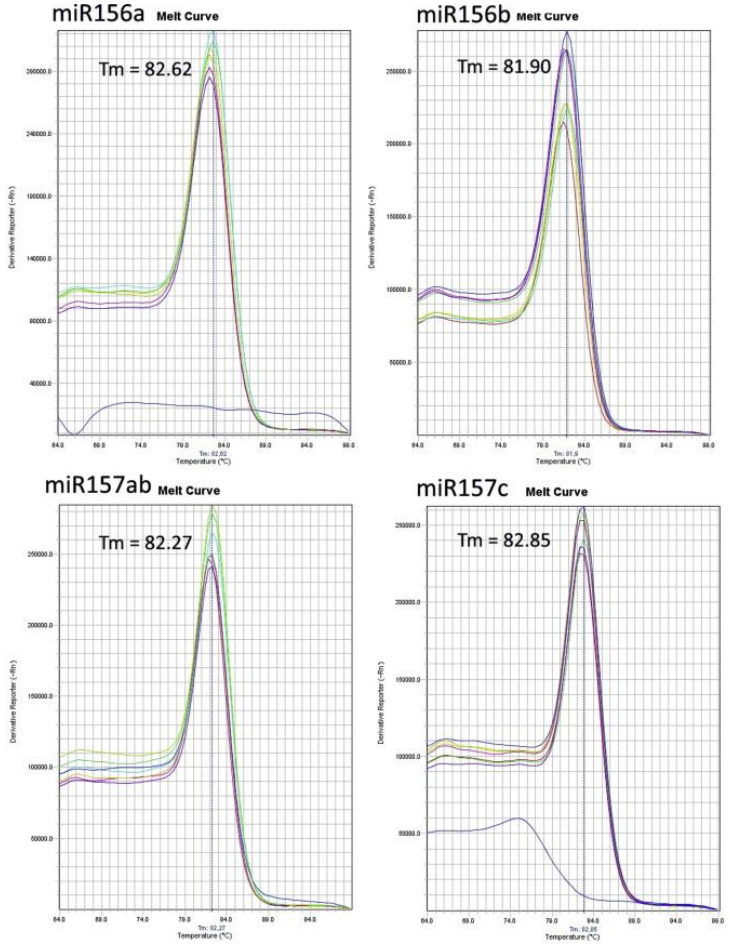
[17] P.A. Davoren, R.E. McNeill, A.J. Lowery, M.J. Kerin, N. Miller, Identification of suitable endogenous control genes for microRNA gene expression analysis in human breast cancer, BMC Mol. Biol. 9 (2008) 76.
[18] H.J. Peltier, G.J. Latham, Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues, RNA 14 (2008) 844–852.
[19] C.R. Galiveti, T.S. Rozhdestvensky, J. Brosius, H. Lehrach, Z. Konthur, Application of housekeeping npcRNAs for quantitative expression analysis of human transcriptome by real-time PCR, RNA 16 (2010) 450–461.
[20] M. Libault, S. Thibivilliers, D.D. Bilgin, O. Radwan, M. Benitez, S.J. Clough, G. Stacey, Identification of four soybean reference genes for gene expression normalization, Plant Genome 1 (2008) 11.
[21] R. Hu, C. Fan, H. Li, Q. Zhang, Y.F. Fu, Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR, BMC Mol. Biol. 10 (2009) 93.
[22] P.K. Martins, B.Q. Jordão, N. Yamanaka, J.R.B. Farias, M.A. Beneventi, E. Binneck, R. Fuganti, R. Stolf, A.L. Nepomuceno, Differential gene expression and mitotic cell analysis of the drought tolerant soybean (*Glycine max* L. Merrill Fabales, Fabaceae) cultivar MG/BR46 (Conquista) under two water deficit induction systems, Genet. Mol. Biol. 31 (2008) 512–521.
[23] E.J. Hewitt, Nutrition of plants in culture media, in: F.C. Steward (Ed.), Inorganic Nutrition of Plants, Academic Press, London, 1963, pp. 97–360.
[24] M. van de Mortel, J.C. Recknor, M.A. Graham, D. Nettleton, J.D. Dittman, R.T. Nelson, C.V. Godoy, R.V. Abdelnoor, A.M. Almeida, T.J. Baum, S.A. Whitham, Distinct biphasic mRNA changes in response to Asian soybean rust infection, Mol. Plant Microbe Interact. 20 (2007) 887–899.
[25] K. Lao, N.L. Xu, V. Yeung, C. Chen, K.J. Livak, N.A. Straus, Multiplexing RT-PCR for the detection of multiple miRNA species in small samples, Biochem. Biophys. Res. Commun. 343 (2006) 85–89.
[26] F. Tang, P. Hajkova, S.C. Barton, K. Lao, M.A. Surani, MicroRNA expression profiling of single whole embryonic stem cells, Nucleic Acids Res. 34 (2006) e9.
[27] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, F. Speleman, Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes, Genome Biol. 3 (2002). RESEARCH0034.
[28] C.L. Andersen, J.L. Jensen, T.F. Orntoft, Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets, Cancer Res. 64 (2004) 5245–5250.
[29] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method, Methods 25 (2001) 402–408.
[30] T. Urao, T. Katagiri, T. Mizoguchi, K. Yamaguchi-Shinozaki, N. Hayashida, K. Shinozaki, Two genes that encode Ca(2+)-dependent protein kinases are induced by drought and high-salt stresses in *Arabidopsis thaliana*, Mol. Gen. Genet. 244 (1994) 331–340.
[31] A.A. Ludwig, T. Romeis, J.D. Jones, CDPK-mediated signalling pathways: specificity and cross-talk, J. Exp. Bot. 55 (2004) 181–188.
[32] Y. Saijo, S. Hata, J. Kyozuka, K. Shimamoto, K. Izui, Over-expression of a single Ca^{2+}-dependent protein kinase confers both cold and salt/drought tolerance on rice plants, Plant J. 23 (2000) 319–327.
[33] J. Wang, B. Czech, D. Weigel, miR156-regulated SPL transcription factors define and endogenous flowering pathway in *Arabidopsis thaliana*, Cell 138 (2009) 738–749.

Supplementary file 1: NormFinder analysis of candidate reference genes and respectively stability values within each set analized.

	All samples set				Tissues set (leaf::root)		Genotypes set (EM::BR::PI)		Abiotic stress set (drought::control)		Biotic stress set (ASR::control)	
	non-subgroups		(stress::control)									
	Gene	Stability value	Gene	Stability value	Gene	Stability value	Gene	Stability value	Gene	Stability value	Gene	Stability value
	156a	0.297	156a	0.112	156a	0.210	156a	0.146	156b	0.133	1520c	0.160
	167ab	0.298	172ab	0.150	167ab	0.212	167c	0.156	1520d	0.146	156a	0.183
	396a	0.411	1520d	0.150	171b	0.224	1520d	0.184	156a	0.148	167ab	0.185
	172ab	0.436	156b	0.150	CYP2	0.245	156b	0.208	171b	0.163	172ab	0.187
	156b	0.441	167ab	0.153	396a	0.265	171b	0.219	171a	0.169	156b	0.203
	1520d	0.453	396a	0.156	156b	0.294	ELF1B	0.231	172ab	0.186	1520d	0.211
	1520c	0.501	1520c	0.186	ACT	0.297	167ab	0.252	167c	0.235	396a	0.235
	ACT	0.637	ACT	0.209	1520d	0.306	171a	0.253	167ab	0.236	ELF1B	0.242
	FBOX	0.697	FBOX	0.223	FBOX	0.347	FBOX	0.307	396a	0.240	CYP2	0.324
	171a	0.711	167c	0.227	1520c	0.348	CYP2	0.313	1520c	0.251	FBOX	0.360
	167c	0.721	171b	0.235	CDPK	0.357	CDPK	0.321	FBOX	0.258	CDPK	0.367
	171b	0.785	CYP2	0.239	172ab	0.366	172ab	0.345	ACT	0.319	ACT	0.396
	CYP2	0.814	171a	0.240	171a	0.367	ACT	0.367	CYP2	0.355	171a	0.399
	ELF1B	0.872	ELF1B	0.280	167c	0.429	1520c	0.374	ELF1B	0.478	TUA	0.442
	CDPK	0.894	TUA	0.350	ELF1B	0.459	TUA	0.399	TUA	0.594	167c	0.495
	TUA	1.184	CDPK	0.380	TUA	0.469	396a	0.441	CDPK	0.626	171b	0.504
Best gene	156a		156a		156a		156a		156b		1520c	
Stability value			0.112		0.210		0.146		0.133		0.160	
Best combination of two genes			156a and 1520d		167ab and 171b		156a and 1520d		156a and 171a		172ab and 1520c	
Stability v for best comb of 2 genes			0.097		0.129		0.102		0.097		0.125	



Supplementary File 2. An amplification plot of four sequence-related microRNAs after RT-qPCR using SYBR-green.



Supplementary File 3. A melting curve of four sequence-related microRNAs after RT-qPCR using SYBR-green.

CAPÍTULO IV

Identification of novel soybean miRNAs involved in abiotic and biotic stresses

Kulcheski et al. BMC Genomics 2011, 12:307
http://www.biomedcentral.com/1471-2164/12/307

BMC Genomics

RESEARCH ARTICLE Open Access

Identification of novel soybean microRNAs involved in abiotic and biotic stresses

Franceli R Kulcheski[1], Luiz FV de Oliveira[1,2], Lorrayne G Molina[1,2], Maurício P Almerão[1], Fabiana A Rodrigues[3], Juliana Marcolino[3], Joice F Barbosa[3], Renata Stolf-Moreira[3], Alexandre L Nepomuceno[3], Francismar C Marcelino-Guimarães[3], Ricardo V Abdelnoor[3], Leandro C Nascimento[4], Marcelo F Carazzolle[4,5], Gonçalo AG Pereira[4] and Rogério Margis[1,2*]

Abstract

Background: Small RNAs (19-24 nt) are key regulators of gene expression that guide both transcriptional and post-transcriptional silencing mechanisms in eukaryotes. Current studies have demonstrated that microRNAs (miRNAs) act in several plant pathways associated with tissue proliferation, differentiation, and development and in response to abiotic and biotic stresses. In order to identify new miRNAs in soybean and to verify those that are possibly water deficit and rust-stress regulated, eight libraries of small RNAs were constructed and submitted to Solexa sequencing.

Results: The libraries were developed from drought-sensitive and tolerant seedlings and rust-susceptible and resistant soybeans with or without stressors. Sequencing the library and subsequent analyses detected 256 miRNAs. From this total, we identified 24 families of novel miRNAs that had not been reported before, six families of conserved miRNAs that exist in other plants species, and 22 families previously reported in soybean. We also observed the presence of several isomiRNAs during our analyses. To validate novel miRNAs, we performed RT-qPCR across the eight different libraries. Among the 11 miRNAs analyzed, all showed different expression profiles during biotic and abiotic stresses to soybean. The majority of miRNAs were up-regulated during water deficit stress in the sensitive plants. However, for the tolerant genotype, most of the miRNAs were down regulated. The pattern of miRNAs expression was also different for the distinct genotypes submitted to the pathogen stress. Most miRNAs were down regulated during the fungus infection in the susceptible genotype; however, in the resistant genotype, most miRNAs did not vary during rust attack. A prediction of the putative targets was carried out for conserved and novel miRNAs families.

Conclusions: Validation of our results with quantitative RT-qPCR revealed that Solexa sequencing is a powerful tool for miRNA discovery. The identification of differentially expressed plant miRNAs provides molecular evidence for the possible involvement of miRNAs in the process of water deficit- and rust-stress responses.

Background

Small, non-coding RNAs have been characterized in plants as important factors involved in gene expression regulation in developmental processes [1,2], as well as adaption to biotic and abiotic stress conditions [3,4]. In general, small RNAs are grouped into two major classes: microRNAs (miRNAs) and short-interfering RNAs (siRNAs). These two classes of small RNAs cannot be discriminated by either their chemical composition or mechanism of action [5,6]. However, siRNAs and miRNAs can be distinguished by their origin, evolutionary conservation and the types of genes that they silence [5,6]. In this way, miRNAs are well differentiated due to some particular characteristics. These characteristics include the following: derived from genomic loci distinct from other recognized genes, processed from transcripts that can form local RNA hairpin structures, and usually, miRNAs sequences are nearly always conserved in related organisms [6,7].

* Correspondence: rogerio.margis@ufrgs.br
[1]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
Full list of author information is available at the end of the article

In plants, *MIRNA* genes are transcribed by RNA polymerase II enzymes (Pol II) generating primary miRNA (pri-miRNA). The pri-miRNA forms an imperfect fold-back structure, which is processed into a stem-loop precursor (pre-miRNA) by nuclear RNaseIII-like enzymes called DICER-LIKE proteins (e.g., DCL1) [8]. The resulting pre-miRNA contains a miRNA:miRNA* intermediate duplex, formed by a self-complementary fold-back structure. A mature miRNA sequence can range from 19 to 24 nucleotides (nt) in length and act as a regulatory molecule in post-transcriptional gene silencing by base pairing with target mRNAs. This leads to mRNA cleavage or translational repression, depending on the degree of complementarity between the miRNA and its target transcript [6,9]. The same mature miRNA can also present several variants of their sequence in length. These populations of miRNA variants are called isomiRNAs, which are isoforms of microRNAs [10]. They are caused by an imprecise or alternative cleavage of Dicer during pre-miRNA processing [10]. IsomiRNAs have been recently identified in both plants and animals [10-12].

The first plant miRNAs were described in *Arabidopsis thaliana* [13,14] and later in other species. Currently, miRNAs have been reported in 41 plants species, and all of their sequences have been deposited in a publicly-available miRNA database, miRBase (http://www.sanger.ac.uk/cgi-bin/Rfam/mirna/browse.pl) [15-18]. Several miRNAs have been identified in plants, and they are characterized in a wide variety of metabolic and biological processes in plants with important functions in development [19,20], phytohormone signaling [21], flowering and sex determination [22] and responses to biotic and abiotic stresses [3,4,19,23-25].

In soybean (*Glycine max* (L.) Merrill), the major legume crop worldwide, Subramanian et al. in 2008 [26] identified 35 novel miRNA families for the first time. In this study, the role of miRNAs in soybean-rhizobial symbiosis was investigated [26]. During that same year, Zhang et al. [27] used a comparative genome-based *in silico* screening of soybean EST databases and quantitative PCR to provide evidence for 69 miRNAs belonging to 33 families. A second study involving miRNAs and soybean root nodules was performed by Wang and colleagues [28]. They identified 32 miRNAs belonging to 11 miRNA families. The identification of nine novel miRNAs in wild soybean (*Glycine soja*) was also reported by Chen et al. [29]. Another study looked at four different soybean tissues (root, seed, flower and nodule) and identified 87 novel soybean miRNAs [30]. Recently, Song and coworkers [31] identified 26 new miRNAs and their related target genes from developing soybean seeds. Although these studies resulted in a large number of miRNAs identified in soybean, none of them looked at microRNAs with respect to biotic and abiotic stresses.

Drought is the major abiotic stress factor to negatively affect soybean productivity around the world. The impact of limited water during the flower formation can cause shorter flowering periods [32,33], and water stress during the later phases of soybean reproductive development has been reported to accelerate senescence, which decreases the duration of the seed-filling period [32,33]. With regards to biotic stress, Asian soybean rust (ASR) is a foliar disease caused by the fungus *Phakopsora pachyrhizi* Sydow & Sydow. This pathogen presents a rapid aerial spread and a high capacity to colonize leaf tissue and, to a lesser extent, stem and pods [34]. ASR is one of the most severe diseases on the soybean culture, which causes damage between 10% and 90% in the different regions where it has been identified [35,36]. This disease is the main threat in soybean-producing countries.

Currently, there are 203 miRNAs identified in *Glycine max* (miRBase database, release 16, http://www.mirbase.org/); however, none of these miRNAs were associated with water deficit or ASR stress conditions. We consider that the identification of these miRNAs is important to understanding small RNA-mediated gene regulation in soybean roots under water deficit stress and in leaves during rust infection. In this context, our goal was to identify new miRNAs and to discover those that may be regulated by water deficit and soybean rust stress. Using high-throughput sequencing, we constructed four libraries of small RNAs from the roots of drought-sensitive and tolerant seedlings in response to control or water deficit conditions. We also constructed four libraries from leaves of rust-susceptible and resistant seedlings with mock and infected conditions. A set of eight small RNAs libraries was analyzed from soybean plants. A total of 256 miRNAs were detected in Solexa sequencing. We discovered 24 novel miRNAs families and also detected several isomiRNAs in soybean. In our RT-qPCR analysis, we verified that the expression profile of several miRNAs varied during abiotic and biotic stresses. This study has important implications for gene regulation under water deficit and pathogen-infection conditions and also contributes significantly to increase the number of identified miRNAs in soybean.

Methods

Plant materials and treatments

Water deficit assay

For water deficit treatment, we used the soybean (*Glycine max* (L.) Merrill) cultivars 'Embrapa 48' as a drought-tolerant standard and 'BR 16' as a sensitive standard [37]. Plants were grown in a greenhouse at Embrapa-Soybean (Londrina, Brazil) using a hydroponic

system compound for plastic containers (30 liters) and an aerated pH 6.6-balanced nutrient solution. Seeds were pre-germinated on moist filter paper in the dark at 25°C ± 1°C and in 65% ± 5% relative humidity. Plantlets were then placed in polystyrene supports so the roots of the seedlings were completely immersed in the nutrient solution. Each seedling tray was maintained in a greenhouse at 25°C ± 2°C and in 60% ± 5% relative humidity under natural daylight (photosynthetic photon flux density (PPFD) = 1.5 × 10^3 μmoles m^{-2} s^{-1}, equivalent to 8.93 × 10^4 lux) for a 12 h day. After 15 days, seedlings with the first trifoliate leaf fully developed (V2 developmental stage) [38] were submitted to different water-deficit treatments according to Martins et al. [39]. The nutrient solution was removed from each plastic container where the roots were kept in the tray in the dark without nutrient solution or water for 0 minutes (T0 or control), 125 minutes (T125) and 150 minutes (T150). At the end of each water-deficit period, the roots of the seedlings were immediately frozen in liquid nitrogen and stored at -80°C until RNA extraction. The experimental design was a factorial (cultivars × duration of water deficit) with three replicates. Each replicate was composed of five plantlets that were sampled in bulk. Four libraries of small RNAs were constructed for the water deficit-stress assays from the following root tissues: 1) roots of drought-sensitive seedlings submitted to 0 minutes of stress (Drought-Sensitive Root Control (DSRC)); 2) roots of drought-sensitive seedlings submitted to 125 minutes and 150 minutes of stress (Drought-Sensitive Root Treated (DSRT)); 3) roots from drought-tolerant seedlings submitted to 0 minutes of stress (Drought-Tolerant Root Control (DTRC)); and 4) roots of drought-tolerant seedlings submitted to 125 minutes and 150 minutes of stress (Drought-Tolerant Root Treated (DTRT)).

Asian Soybean Rust assay

The ASR reaction was evaluated in soybean plants in a greenhouse at Embrapa-Soybean (Londrina, Brazil) using a field population of *Phakopsora pachyrhizi* collected from soybean fields in the state of Mato Grosso, which were maintained for over 10 generations on the susceptible cv. BRSMS-Bacuri. ASR identification was confirmed by ITS-sequencing analysis as described by Silva et al. [40], and it revealed a similarity to the MUT Zimbabwe isolate. The soybean plants were grown in a pot-based system. The 'Embrapa 48' genotype was used as a susceptible host plant, which develops a susceptible lesion (TAN) after *Phakopsora pachyrhizi* infection. The 'PI561356' genotype was used as the resistant host, which carries an ASR resistance gene mapped onto linkage group G (Ricardo V. Abdelnoor, personal communication) and develops a reddish-brown (RB) lesion with few or no spores.

Urediniospores were collected from infected BRSMS-Bacuri plants in a separate greenhouse by tapping infected leaves over a plastic tray. The urediniospores were then diluted in distilled water with 0.05% Tween-20 to a final concentration of 3 × 10^5 spores/mL. This spore suspension was sprayed onto three plants per pot at the V2 to V3 growth stages [38]. A solution without the spores was used for the mock inoculations. Following the ASR or mock inoculations, water-misting bags were placed over all plantlets for one day to aid the infection process and to prevent cross-contamination of the mock-infected plants. The third trifoliolate leaves of six plants were collected 12 hours after inoculation (hai) for RNA extraction. The experiment followed a completely randomized design with the three replicates as blocks and a full factorial treatment structure consisting of three treatment factors: hai (12 hours), genotype (resistant or susceptible), and inoculation type (ASR or mock).

For the rust-stress assay, we constructed the other four libraries of small RNAs from leaves which were compounded by: 1) leaves of rust-susceptible seedlings with mock inoculation (Rust-Susceptible Leaf Control (RSLC)); 2) leaves of rust-susceptible seedlings with rust-spore inoculation (Rust-Susceptible Leaf Treated (RSLT)); 3) leaves of rust-resistant seedlings with mock inoculation (Rust-Resistant Leaf Control (RRLC)); and 4) leaves of rust-resistant seedlings with rust-spore inoculation (Rust-Resistant Leaf Treated (RRLT)).

RNA extraction and sequencing

Total RNA was isolated from fresh leaves and root materials using Trizol (Invitrogen, CA, USA), and the RNA quality was evaluated by electrophoresis on a 1% agarose gel. The amount of the RNA was verified using a Quibit fluorometer and Quant-iT RNA assay kit according to the manufacturer's instructions (Invitrogen, CA, USA). Total RNA (> 10 μg) was sent to Fasteris Life Sciences SA (Plan-les-Ouates, Switzerland) for processing and sequencing using Solexa technology on the Illumina Genome Analyzer GAII. The libraries were constructed from the eight bar-coded samples (DSRC, DSRT, DTRC, DTRT, RSLC, RSLI, RRLC and RRLI) sequenced in a total of two channels. Quality scores were generated from Illumina's data analysis pipeline, which are similar to SAGE Phred scores with a maximum value of 40. Quality scores are based on the relative confidence of base calls using elements of cluster generation and image quality. Briefly, the processing by Illumina for the miRNA analyses consisted of the following successive steps: acrylamide gel purification of the RNA bands corresponding to the size range 20-30 nt, ligation of the 3' and 5' adapters to the RNA in two separate subsequent steps each followed by acrylamide

gel purification, (3) cDNA synthesis followed by acrylamide gel purification, and a final step of PCR amplification to generate cDNA colonies template library for Illumina sequencing. After removing the adapter sequences, the sequences were trimmed into different read lengths from 19 to 24 nt for further analysis.

Prediction of miRNAs

The reads were grouped into unique sequences, and the read counts were calculated for each library. The sequences that presented low read counts (read count < = 2) were discarded from the final list of unique sequences, which are referred to as a tag. The sequences were mapped into the soybean genome (http://www.phytozome.net) assembly using the SOAP program [41], which returns information concerning the alignment position, chromosome number and strand. No mismatches were allowed in the alignments. The tag alignment position's upstream and downstream genomic sequences (200 bp each) were extracted from the genome assembly using homemade Perl scripts. These genomic regions were then aligned against the reverse complement of its respective tag (rc-tag) using the Smith-Waterman algorithm [42]. To ensure that these pre-miRNA sequences could be precisely processed into mature miRNA, the candidates were examined according the following criteria [43]: i) the miRNA and anti-sense miRNA should derive from the opposite stem-arms and must be entirely within the arm of the hairpin; ii) base-pairing between the miRNA and anti-sense miRNA were restricted to four or fewer mismatches; and iii) the frequency of asymmetric bulges was restricted to less than one and the size should be less than two bases. The genomic regions that were not possible to align the tag and rc-tag were discarded. Finally, the genomic regions that were limited between the alignment positions of the tag and rc-tag were considered as pre-microRNA candidates. From all the pre-microRNA candidate sequences that we selected, only the ones with no more than five matches to the soybean genome were selected for analyzing the secondary structure using the RNA-folding program Mfold [44]. If a perfect stem-loop structure was formed, the small RNA sequence was at one arm of the stem, and the respective anti-sense sequence was at the opposite arm; then, the small RNA was consisted as a novel soybean miRNA.

miRNA validation and expression analysis by RT-qPCR

To validate predicted new miRNAs, RT-qPCR in respect to eleven miRNAs was performed to examine their expression across the eight different libraries. From those, six were new miRNAs belonging to conserved soybean miRNAs families (MIR166a-5p, MIR166f, MIR169f-3p, MIR482bd-3p, MIR1513c, MIR4415b); one new miRNA pertencing to conserved miRNAs families in other plants species (MIR397ab); and four were miRNAs belonging to novel miRNAs families (MIR-Seq07, MIR-Seq11, MIR-Seq13, MIR-Seq15ab). The forward miRNAs primers were designed based on the full miRNAs sequence, and the reverse primer was the universal reverse primer for miRNA [45]. The stem-loop primer, used for miRNA cDNA synthesis, was designed according to Cheng et al. [45]. The stem-loop sequence consisted of 44 conserved and six variable nucleotides that were specific to the 3' end of the miRNA sequence (5' GTCGTATCCAGTGCAGGGTCCGAGGTATTCG-CACTGGATACGACNNNNNN 3'). The RT-qPCR was performed in an ABI 7500 Real-Time PCR System (Applied Biosystems) using SYBR Green I (Invitrogen) to detect double-stranded cDNA synthesis. Reactions were completed in a volume of 24 μL containing 12 μL of diluted cDNA (1:50), 1X SYBR Green I (Invitrogen), 0.025 mM dNTP, 1X PCR Buffer, 3 mM $MgCl_2$, 0.25 U Platinum Taq DNA Polymerase (Invitrogen) and 200 nM of each reverse and forward primer. The universal reverse primer (5' GTGCAGGGTCCGAGGT 3') was used in all RT-qPCR reactions. Samples were analyzed in biological triplicate in a 96-well plate, and a no-template control was included. We used MIR156b (5'-TGACAGAAGAGAGAGAGCACA - 3'), MIR172ab (5'-AGAATCTTGATGATGCTGCAT - 3') and MIR1520d (5'- ATCAGAACATGACACGTGACAA - 3') as reference genes, which has been demonstrated as optimal normalizers for water deficit and rust-stress analysis in *Glycine max* [46]. The conditions were set as the following: an initial polymerase activation step for 5 minutes at 94°C, 40 cycles for 15 seconds at 94°C for denaturation, 10 seconds at 60°C for annealing and 25 seconds at 72°C for elongation. A melting curve analysis was programmed at the end of the PCR run over the range 65-99, increasing the temperature stepwise by 0.4°C. Threshold and baselines were manually determined using the ABI 7500 Real-Time PCR System SDS Software v2.0. To calculate the relative expression of the miRNAs, we used the $2^{-\Delta\Delta Ct}$ method. Student's *t*-test was performed to compare pair-wise differences in expression. The parameters of two-tailed distribution and two samples assuming unequal variances were established. The means were considered significantly different when $P < 0.05$.

Prediction of miRNA targets

Target prediction for miRNAs is straightforward because it is assumed that most of them match their targets with almost perfect complementarity [8,9]. The putative target genes for all miRNAs identified were searched for by using the web-based computer psRNA Target Server (http://biocomp5.noble.org/psRNATarget/) [47] which

can identify putative targets that may be regulated at post-transcriptional or at translational levels. Mature miRNA sequences were used as queries to search for potential target mRNAs in the *Glycine max* database (DFCI gene index release 15). The total scoring for an alignment was calculated based on the miRNA length, and the sequences were considered to be miRNA targets if the total score were less than 3.0 points (mismatch = 1 and G:U = 0.5). Results from these analyses were individually inspected on the Phytozome, where the loci and protein annotation were obtained. In order to look for evidences of the predicted targets of the novel identified miRNA, we searched for the miRNA targets sites in the soybean degradome libraries published by Song et al. [31] available under NCBI-GEO accession nμ. GSE25260. Finally, all putative targets regulated by soybean new miRNAs which were analyzed by RT-qPCR were subjected to AgriGO database to investigate the gene ontology [48].

Results

To identify miRNAs from soybean under water deficit and rust stresses, we generated eight libraries of small RNAs species. From these libraries, a total of 256 miRNAs ranged from 19 to 24 nt-long sequence sizes were identified (Table 1). All pre-miRNA sequence candidates that were selected by the parameters stipulated during the miRNA prediction and those that had no more than five matches on the soybean genome were folded using the Mfold program. All miRNA sequences with the respective precursor sequence originating at a hairpin structure were submitted to the miRBase to determine if they were a new or known miRNA. We separated the results of these miRNAs according the following classes: novel miRNAs belonging to miRNAs families never detected before (29 miRNAs); new miRNAs belonging to conserved miRNA families in other plants species detected for the first time in soybean (15 miRNAs); miRNAs belonging to conserved miRNA soybean families (71 miRNAs); different isoforms of new and known miRNAs (121 isoforms); and known miRNAs already deposited into the miRBase database (20 miRNAs) (Table 1).

Identification of novel miRNAs from soybean

A total of 29 new miRNAs belonging to 24 novel families (Table 2) were identified by Solexa sequencing in libraries from water deficit and rust infections of *Glycine max*. These families were provisory nominated Seq01 to Seq25 (Table 2). The precursor miRNA sequences varied from 55 to 239 nt in length. Precursors of these novel miRNAs were identified, and they formed proper secondary hairpin structures, with MFEs ranging from -16.50 to -153.80 kcal/mol (Additional file 1). The most abundant mature miRNAs were 21 nt in length. We also evaluated the genomic location of the new miRNAs (Table 2). Of the 29 new miRNAs genes identified in soybean, around 86% (25) were located in intergenic regions and the rest were situated inside genes. The mature miRNAs sequences were localized inside the stem-loop sequence with almost half in each arm: 17 miRNAs were localized in the 3' arm and 12 miRNAs were in the 5' arm. More than 63% of the pre-miRNA sequences were in the same sense direction (+) as the soybean genome annotation. For all 24 novel families identified, four were compounded by miRNAs provided from two loci, and we detected only one miRNA member for the rest. Sense and anti-sense miRNAs were detected only in one family, the Seq10, and both were nominated according the arm localization (3p or 5p). Most of the new mature miRNA sequences presented a uracil (U) as their first nucleotide, which is in agreement with previous results for soybean root sequences [26].

Identification of homologues miRNAs of other plant species

To determine whether any of the miRNAs identified in our libraries were conserved among other plant species, we searched miRBase for homologues. Besides the novel families identified, we also detected 15 miRNAs belonging to six conserved families in other plants species (Table 3). The families MIR170, MIR395, MIR397, MIR408, MIR2118 and MIR3522 were detected for the first time in soybean. For families MIR170 and MIR3522, only a single locus was identified, and for MIR408, three genes were found. In two families, MIR408 and MIR2118, we detected sense and antisense

Table 1 The amount of different miRNA classes detected by high-throughput sequencing.

Class	Size (nt)						Total
	19	20	21	22	23	24	
Novel miRNAs	4	3	12	5	1	4	29
New miRNAs pertencing to conserved miRNAs families in other plants species	1	2	9	3	-	-	15
New miRNAs identified in conserved soybean miRNAs families	1	7	52	9	2	-	71
Isoforms of new and known miRNAs	24	50	26	16	4	1	121
Known miRNAs	-	1	16	3	-	-	20
miRNAs detected	30	63	115	36	7	5	256

Table 2 The novel soybean microRNA families determined from Solexa sequencing.

Sequence Code[a]	Mature miRNA			Pre-miRNA					Region[b]
	Sequence	Size (nt)	Ch	Start	End	Length (nt)	Sense	Arm	
gma-MIR-Seq01	GGACAGUCUCAGGUAGACA	19	Gm04	30764003	30764171	169	-	3p	intergenic
gma-MIR-Seq03	UGAGAAAAGGAGGAUGUCA	19	Gm11	29821812	29821926	115	+	3p	intergenic
gma-MIR-Seq04a	GCUGGAUGUCUUUGAAGGA	19	Gm08	46853906	46853991	86	+	3p	intergenic
gma-MIR-Seq04b	GCUGGAUGUCUUUGAAGGA	19	Gm18	61624611	61624690	80	-	3p	intergenic
gma-MIR-Seq05	AACCCUCAAAGGCUUCCUAG	20	Gm18	61626669	61626771	103	+	5p	intergenic
gma-MIR-Seq06	AGUGGAACUUUGAGGCCUGC	20	Gm08	46848259	46848354	96	+	3p	intergenic
gma-MIR-Seq07	AAAUGACUUGAGAGGUGUAG	20	Gm01	44787899	44787988	90	+	5p	intergenic
gma-MIR-Seq08	CUAAAGAUUGUCCAAAAGGAA	21	Gm14	6763304	6763456	153	+	5p	intergenic
gma-MIR-Seq09	GUAGUGGAUGCCUAGAGGUCC	21	Gm18	61655979	61656075	97	-	3p	intergenic
gma-MIR-Seq10-5p	UAGGAAUUAGUCACUCAGAUC	21	Gm15	31542836	31543058	223	+	5p	intergenic
gma-MIR-Seq10-3p	AUCUCAGUGACUAAUUUCUAG	21	Gm15	31542836	31543058	223	+	3p	intergenic
gma-MIR-Seq11	UUGUUCGAUAAAACUGUUGUG	21	Gm16	5744795	5744863	69	-	5p	intergenic
gma-MIR-Seq12	UCUCUUGAUUCUAGAUGAUGU	21	Gm16	27653048	27653102	55	+	3p	CDS
gma-MIR-Seq13	UGUUGCGGGUAUCUUUGCCUC	21	Gm04	28578972	28579075	104	-	5p	intergenic
gma-MIR-Seq14a	UGAGAAUUUGGCCUCUGUCCA	21	Gm09	28264427	28264515	89	+	5p	intergenic
gma-MIR-Seq14b	UGAGAAUUUGGCCUCUGUCCA	21	Gm09	28272488	28272562	75	+	5p	intergenic
gma-MIR-Seq15a	UUAGAUUCACGCACAAACUUG	21	Gm02	1041996	1042084	89	+	3p	intergenic
gma-MIR-Seq15b	UUAGAUUCACGCACAAACUUG	21	Gm10	1085223	1085322	100	+	3p	intergenic
gma-MIR-Seq16	UUAUAGUCUGACAUCUGGAAU	21	Gm05	9279518	9279737	220	+	5p	intergenic
gma-MIR-Seq17	ACUAUAGAAGUACUUGUGGAGC	22	Gm16	2916844	2917034	191	+	5p	CDS/intronic
gma-MIR-Seq18	CCUCAUUCCAAACAUCAUCUAA	22	Gm09	16565935	16566025	91	-	3p	intergenic
gma-MIR-Seq19	UGAAGAUUUGAAGAAUUUGGGA	22	Gm15	16900161	16900327	167	+	5p	intronic
gma-MIR-Seq20	CAUCGUUGACGCUGACUGUACG	22	Gm04	35428794	35428950	157	-	5p	5'UTR/intronic
gma-MIR-Seq21	CUGAAGGAUCGAUGUAGAAUGCU	23	Gm02	39825520	39825641	122	+	3p	intergenic
gma-MIR-Seq22	CAUCUGAAGGAUAGAACACAUA	22	Gm09	29816467	29816705	239	+	3p	intergenic
gma-MIR-Seq23	AGUUUCGUGACUACAACUUCUGAA	24	Gm15	16900193	16900294	102	-	3p	intergenic
gma-MIR-Seq24	AUGAAAAUCAUUCAUUAUGAUAUC	24	Gm16	28536014	28536181	168	-	3p	intergenic
gma-MIR-Seq25a	GAAAAUGAAUGAUGAGGAUGGGGA	24	Gm11	7787358	7787494	137	-	3p	intergenic

[a] The number refers to a new family and the letter refers to the new gene in that family. [b] CDS: codon sequence.

Table 3 New *Glycine max* miRNA families conserved in other plants species.

Family	Acronym	miRNA Sequence	Size (nt)	Species
MIR170	gma-MIR170	UAUUGGCCUGGUUCACUCAGA	21	*ath, aly*
MIR395	gma-MIR395a	CUGAAGUGUUUGGGGGAACUC	21	*ath, ptc, vvi, sly, rco, aly, csi, osa,*
	gma-MIR395b	CUGAAGUGUUUGGGGGAACUC	21	*sbi, mtr, zma, tae, pab*
	gma-MIR395c	CUGAAGUGUUUGGGGGAACUC	21	
MIR397	gma-MIR397a	UCAUUGAGUGCAGCGUUGAUG	21	*ath, osa, ptc, bna, vvi, sbi, bdi, rco,*
	gma-MIR397b	UCAUUGAGUGCAGCGUUGAUG	21	*aly, csi, zma, pab, sly, hvu*
MIR408	gma-MIR408a	AUGCACUGCCUCUUCCCUGGC	21	*ath, ptc, pta, vvi, ahy, aly, csi, osa,*
	gma-MIR408b-5p	CUGGGAACAGGCAGGGCACG	20	*sof, zma, ppt, smo,*
	gma-MIR408b-3p	AUGCACUGCCUCUUCCCUGGC	21	*tae, sbi, bdi, rco, aqc*
	gma-MIR408c	AUGCACUGCCUCUUCCCUGGC	21	
MIR2118	gma-MIR2118a-5p	GGAGAUGGGAGGGUCGGUAAAG	22	
	gma-MIR2118a-3p	UUGCCGAUUCCACCCAUUCCUA	22	*pvc, gso, mtr, osa, zma*
	gma-MIR2118b-5p	GGAGAUGGGAGGGUCGGUAA	20	
	gma-MIR2118b-3p	UUGCCGAUUCCACCCAUUCCUA	22	
MIR3522	gma-MIR3522a	UGAGACCAAAUGAGCAGCUGA	21	*gso*

Arabidopsis lyrata (aly), Arabidopsis thaliana (ath), Brassica napus (bna), Ricinus communis (rco), Medicago truncatula (mtr), Phaseolus vulgaris (pvu), Arachis hypogaea (ahy), Glycine soja (gso), Aquilegia coerulea (aqc), Seleginella moellendorffii (smo), Physcomitrella patens (ppt), Pinus taeda (pta), Picea abies (pab), Populus trichocarpa (ptc), Citrus sinensis (csi), Vitis vinifera (vvi), Solanum lycopersicum (sly), Brachypodium distachyon (bdi), Hordeum vulgare (hvu), Oryza sativa (osa), Saccharum officinarum (sof), Selaginella moellendorffii (smo), Sorghum bicolor (sbi), Triticum aestivum (tae), and *Zea mays (zma).*

miRNAs (Table 3). MIR170 was only conserved in *Arabidopsis lyrata* and *Arabidopsis thaliana*. MIR408 was found in more different plants species than the other families. It was found in 17 species: *Arabidopsis thaliana, Populus trichocarpa, Pinus taeda, Vitis vinifera, Arachis hypogaea, Arabidopsis lyrata, Citrus sinensis, Oryza sativa, Saccharum officinarum, Zea mays, Physcomitrella patens, Selaginella moellendorffii, Triticum aestivum, Sorghum bicolor, Brachypodium distachyon, Ricinus communis and Aquilegia coerulea* (Table 3). We observed two families (MIR2118 and MIR3522) to be conserved between *Glycine max* and *Glycine soja*; however, we expect that more miRNA families could be conserved between these species considering that they are closely related. This low number is probably due to *Glycine soja* showing fewer miRNAs identified to date.

Identification of conserved soybean miRNAs

To identify conserved soybean miRNAs, all 256 sequences were searched using BLASTn against the soybean miRNAs in miRBase. We identified 22 families of conserved soybean miRNAs in our libraries. Only 20 miRNA soybean genes that were registered in the miRBase were observed (indicated by the number five in Table 1). From the remaining 71 miRNA genes, 12 were miRNAs antisense (in the opposite arm) to the miRNAs presents in miRBase (indicated as group four in Table 4), and 59 were new members detected from new loci of known families (indicated by number three in Table 4). Of the 12 miRNAs identified from the opposite strand of previously known miRNAs, six were in the 5' arm and six in the 3'arm. For the 59 new members of conserved soybean families, 45 miRNAs were 21 nt in length. The family with the largest number of new miRNA genes (nine genes) was MIR319 (Table 4). Interestingly, in family MIR166, we found three new members with sense and antisense miRNAs. Also, in MIR159, two new genes with sequences originated from both the 3'and 5'arms were identified. One new gene was detected in MIR169, MIR172, MIR396 and MIR482 with mature sequences originated from both the 3'and 5'arms (Table 4). Similar to the observation for the novel soybean miRNAs (Table 2), the new genes in these conserved soybean families were compounded for a majority of mature miRNAs with a uracil as the first nucleotide in the 5' end.

Identification of miRNAs isoforms

Isoforms of microRNAs (isomiRNAs) are a population of known miRNA variants. They are caused by an imprecise or alternative cleavage of Dicer during pre-miRNA processing [10]. We detected numerous miRNAs with additional nucleotides in the 5'or 3' terminus compared to the recorded mature miRNAs. As isomiRNAs were previously reported in soybean high-throughput sequencing [31], we found 121 isomiRNAs in our libraries (Table 5). These isoforms were observed in 22 conserved miRNA families and in four novel families. These miRNA isoforms occurred in both strands from the 5' or 3' arm. The conserved MIR1507a and MIR1507b were found with the most isomiRNAs detected (eight isoforms each). MIR1507a showed a variation of three nucleotides in the 5'end and six nucleotides in the 3'end, and MIR1507b showed a variation of three and five nucleotides in the 5'and 3' terminal region respectively (Table 5). From the novel miRNAs identified, the MIR-Seq07 was the read with the most isoforms detected in our sequencing. This miRNA presented a total of 14 different sequences with 14 varying nucleotides in both the 5'and 3' ends from six fixed nucleotides (Table 5). All isoforms and their respective nominated mature miRNAs can be found in Additional File 1.

Validation of miRNAs validation and expression profile by RT-qPCR

The stem-loop RT-qPCR was used to validate and measure the expression of the respective miRNAs: MIR166a-5p, MIR166f, MIR169f-3p, MIR397ab, MIR482bd-3p, MIR1513c, MIR4415b, MIR-Seq07, MIR-Seq11, MIR-Seq13 and MIR-Seq15ab, detected by Solexa sequencing. These miRNAs were validated in all genotypes analyzed during dehydration and rust stress. The relative expressions of these miRNAs in the same eight conditions are shown in Figure 1.

Expression patterns of miRNAs during water deficit

To identify water deficit-responsive miRNAs, we compared the expression profiles of the 11 miRNAs in both genotypes before and after stress (Figure 1A). A set of five different miRNAs (*MIR166-5p*, *MIR169f-3p*, *MIR1513c*, *MIR397ab* and *MIR-Seq13*) presented the same behavior during the water deficit stress. These miRNAs were commonly up-regulated during the stress condition in the sensitive genotype, and the opposite occur in the tolerant genotype, where they were down-regulated during the water deficit. *MIR-Seq11* and *MIR-Seq15* demonstrated a similar expression across the four conditions. Water deficit significantly increased *MIR-Seq11* and *MIR-Seq15* expression in the roots compared to the control condition in the sensitive genotype, but both miRNAs did not vary in the tolerant plants. *MIR166f* had its level increased in the sensitive genotype and decreased in the tolerant during the stress compared to the control situation. Interestingly, both genotypes presented the same level during the control condition. In the sensitive plants, *MIR-482bd-3p* showed a strong decrease when submitted to water deficit, being this low level equally observed in the tolerant genotype

Table 4 Families of conserved soybean miRNAs.

Group[a]	miRNA ID	miRNA ID sequence	Size(nt)	Ch	Start	End	Arm	Members registered in miRBase[b]
5	gma-MIR156d	UUGACAGAAGAUAGAGAGCAC	21	Gm08	3891365	3891489	5'	a*,b*,c*,d,e*,f*,g*
3	gma-MIR156h	UUGACAGAAGAUAGAGAGCAC	21	Gm02	7812526	7812628	5'	
3	gma-MIR156i	UUGACAGAAGAUAGAGAGCAC	21	Gm05	38621690	38621813	5'	
3	gma-MIR156j	UUGACAGAAGAGAGUGAGCAC	21	Gm06	4699149	4699240	5'	
3	gma-MIR156k	UUGACAGAAGAUAGAGAGCAC	21	Gm07	9347139	9347259	5'	
3	gma-MIR156l	UUGACAGAAGAUAGAGAGCAC	21	Gm09	37843750	37843864	5'	
3	gma-MIR156m	UUGACAGAAGAGAGUGAGCAC	21	Gm14	10664512	10664600	5'	
3	gma-MIR156n	UUGACAGAAGAGAGUGAGCAC	21	Gm17	37759446	37759535	5'	
5	gma-MIR159a-3p	UUUGGAUUGAAGGGAGCUCUA	21	Gm09	37672410	37672586	3'	a(3'),b(3'),c*,d*
4	gma-MIR159a-5p	GAGCUCCUUGAAGUCCAAUUG	21	Gm09	37672410	37672586	5'	
5	gma-MIR159b-3p	AUUGGAGUGAAGGGAGCUCCA	21	Gm07	5386107	5386292	3'	
4	gma-MIR159b-5p	GAGUUCCCUGCACUCCAAGUC	21	Gm07	5386107	5386292	5'	
3	gma-MIR159e-3p	UUUGGAUUGAAGGGAGCUCUA	21	Gm07	9524917	9525127	3'	
3	gma-MIR159e-5p	GAGCUCCUUGAAGUCCAAUU	20	Gm07	9524917	9525127	5'	
3	gma-MIR159f-3p	AUUGGAGUGAAGGGAGCUCCA	21	Gm16	2794128	2794307	3'	
3	gma-MIR159f-5p	GAGUUCCCUGCACUCCAAGUC	21	Gm16	2794128	2794307	5'	
5	gma-MIR162a	UCGAUAAACCUCUGCAUCCAG	21	Gm06	20176238	20176339	3'	a
3	gma-MIR162b	UCGAUAAACCUCUGCAUCCAG	21	Gm05	7692594	7692698	3'	
3	gma-MIR162c	UCGAUAAACCUCUGCAUCCAG	21	Gm17	10181489	10181607	3'	
5	gma-MIR166a-3p	UCGGACCAGGCUUCAUUCCCC	21	Gm16	1912570	1912715	3'	a(3'),b*
4	gma-MIR166a-5p	GGAAUGUUGUCUGGCUCGAGG	21	Gm16	1912570	1912715	5'	
3	gma-MIR166c-3p	UCGGACCAGGCUUCAUUCCCC	21	Gm02	14340767	14340863	3'	
3	gma-MIR166c-5p	GGAAUGUCGUCUGGUUCGAG	20	Gm02	14340767	14340863	5'	
3	gma-MIR166d-3p	UCGGACCAGGCUUCAUUCCCG	21	Gm08	14990547	14990731	3'	
3	gma-MIR166d-5p	GGAAUGUUGUUUGGCUCGAGG	21	Gm08	14990547	14990731	5'	
3	gma-MIR166e-3p	UCGGACCAGGCUUCAUUCCCG	21	Gm15	3688764	3688931	3'	
3	gma-MIR166e-5p	GGAAUGUUGUUUGGCUCGAGG	21	Gm15	3688764	3688931	5'	
3	gma-MIR166f	UCUCGGACCAGGCUUCAUUCC	21	Gm20	43105394	43105500	3'	
5	gma-MIR167c	UGAAGCUGCCAGCAUGAUCUG	21	Gm07	39778512	39778886	5'	a*,b*,c,d*,e*,f*,g*
3	gma-MIR167h	UGAAGCUGCCAGCAUGAUCUG	21	Gm20	44765096	44765173	5'	
5	gma-MIR168a	UCGCUUGGUGCAGGUCGGGAA	21	Gm09	41353226	42353350	5'	a
3	gma-MIR168b	UCGCUUGGUGCAGGUCGGGAA	21	Gm01	48070311	48070420	5'	
5	gma-MIR169a	CAGCCAAGGAUGACUUGCCGG	21	Gm09	35771804	35771924	5'	a,b*,c*,d*,e*
3	gma-MIR169f-3p	UUUCGACGAGUUGUUCUUGGC	21	Gm02	46876643	46876727	3'	
3	gma-MIR169f-5p	UAGCCAAGAAUGACUUGCCGG	21	Gm02	46876643	46876727	5'	
3	gma-MIR169g	CAGCCAAGAAUGACUUGCCGG	21	Gm09	5263992	5264096	5'	
3	gma-MIR169h	CAGCCAAGAAUGACUUGCCGG	21	Gm14	5324798	5324911	5'	
3	gma-MIR169i	CAGCCAAGGAUGACUUGCCGG	21	Gm10	40332790	40332926	5'	
3	gma-MIR169j	CAGCCAAGGAUGACUUGCCGG	21	Gm13	368563	368441	5'	
3	gma-MIR169k	CAGCCAAGGGUGAUUUGCCGG	21	Gm15	14150069	14150183	5'	
3	gma-MIR169l	CAGCCAAGGAUGACUUGCCGG	21	Gm17	4861963	4861816	5'	
3	gma-MIR171d	UUGAGCCGUGCCAAUAUCACG	21	Gm06	48920631	48920715	3'	a*,b*,c*
3	gma-MIR171e	CGAUGUUGGUGAGGUUCAAUC	21	Gm13	26271135	26271232	5'	
3	gma-MIR171f	CGAUGUUGGUGAGGUUCAAUC	21	Gm17	9101701	9101798	3'	
4	gma-MIR172b-5p	GUAGCAUCAUCAAGAUUCAC	20	Gm13	40401688	40401809	5'	a*,b(3')*,c,d*,e*,f*
5	gma-MIR172c	GGAAUCUUGAUGAUGCUGCAG	21	Gm18	2968986	2969138	3'	
3	gma-MIR172g	GCAGCACCAUCAAGAUUCAC	20	Gm10	31592576	31592689	5'	
3	gma-MIR172h-3p	AGAAUCUUGAUGAUGCUGCAU	21	Gm10	43474725	43474831	3'	
3	gma-MIR172h-5p	GCAGCAGCAUCAAGAUUCACA	21	Gm10	43474725	43474831	5'	
3	gma-MIR172i	GCAGCAGCAUCAAGAUUCACA	21	Gm15	2892962	2893122	5'	

Table 4 Families of conserved soybean miRNAs. *(Continued)*

3	gma-MIR172j	GCAGCAGCAUCAAGAUUCACA	21	Gm20	40895747	40895836	5'	
3	gma-MIR319d	UUGGACUGAAGGGAGCUCCUUC	22	Gm02	43885398	43885595	3'	a*,b*,c*
3	gma-MIR319e	UUGGACUGAAGGGAGCUCCCU	21	Gm02	45704227	45704412	3'	
3	gma-MIR319f	UUGGACUGAAGGGGAGCUCCUUC	23	Gm04	46348798	46348991	3'	
3	gma-MIR319g	UUGGACUGAAGGGAGCUCCCU	21	Gm11	1374020	1374198	3'	
3	gma-MIR319h	UUGGACUGAAGGGAGCUCCCU	21	Gm11	32902062	32902247	3'	
3	gma-MIR319i	UUGGACUGAAGGGAGCUCCCU	21	Gm14	47959350	47959535	3'	
3	gma-MIR319j	UUGGACUGAAGGGAGCUCCUUC	22	Gm14	45953433	45953649	3'	
3	gma-MIR319k	UUGGACUGAAGGGAGCUCCUUC	22	Gm17	9436178	9436279	3'	
3	gma-MIR319l	UUGGACUGAAGGGAGCUCCCU	21	Gm18	4278883	4279072	3'	
4	gma-MIR396a-3p	UUCAAUAAAGCUGUGGGAAG	20	Gm13	26338134	26338273	3'	a,b(5'),c,d(3')*,e*
5	gma-MIR396a-5p	UUCCACAGCUUUCUUGAACUG	21	Gm13	26338134	26338273	5'	
4	gma-MIR396b-3p	GCUCAAGAAAGCUGUGGGAGA	21	Gm13	26329931	26330056	3'	
5	gma-MIR396b-5p	UUCCACAGCUUUCUUGAACUU	21	Gm13	26329931	26330056	5'	
5	gma-MIR396c	UUCCACAGCUUUCUUGAACUU	21	Gm13	43804777	43804893	5'	
4	gma-MIR396d-5p	UUCCACAGCUUUCUUGAACUU	21	Gm17	9053051	9053155	5'	
3	gma-MIR396f	UCCACAGCUUUCUUGAACUG	20	Gm14	13971419	13971566	5'	
3	gma-MIR396g	UUCCACAGCUUUCUUGAACUU	21	Gm15	556707	556796	5'	
3	gma-MIR396h-3p	GUUCAAUAAAGCUGUGGGAAG	21	Gm17	9044850	9044984	3'	
3	gma-MIR396h-5p	UUCCACAGCUUUCUUGAACUG	21	Gm17	9044850	9044984	5'	
4	gma-MIR482b-3p	UCUUCCCUACACCUCCCAUACC	22	Gm20	35360312	35360406	3'	a*,b(5')
5	gma-MIR482b-5p	UAUGGGGGGAUUGGGAAGGAAU	22	Gm20	35360312	35360406	5'	
3	gma-MIR482c	AUUUGUGGGAAUGGGCUGAUUGG	23	Gm18	61452904	61453003	5'	
3	gma-MIR482d-3p	UCUUCCCUACACCUCCCAUACC	22	Gm10	48569629	48569723	3'	
3	gma-MIR482d-5p	UAUGGGGGGAUUGGGAAGGAAU	22	Gm10	48569629	48569723	5'	
5	gma-MIR1507a	UCUCAUUCCAUACAUCGUCUGA	22	Gm13	25849777	25849883	3'	a,b
5	gma-MIR1507b	UCUCAUUCCAUACAUCGUCUG	21	Gm17	6190604	6190701	3'	
5	gma-MIR1508b	UAGAAAGGGAAAUAGCAGUUG	21	Gm09	28530168	28530271	3'	a*,b
5	gma-MIR1509a	UUAAUCAAGGAAAUCACGGUCG	22	Gm17	10099759	10099869	5'	a, b*
4	gma-MIR1510b	AGGGAUAGGUAAAACAACUACU	22	Gm02	6599299	6599392	5'	a*,b(3')
5	gma-MIR1510b	UGUUGUUUUACCUAUUCCACC	21	Gm02	6599299	6599392	3'	
3	gma-MIR1512b	UAACUGGAAAUUCUUAAAGCAU	22	Gm02	8618692	8618781	5'	a*
5	gma-MIR1513a	UGAGAGAAAGCCAUGACUUAC	21	Gm07	43245809	43245901	5'	a
3	gma-MIR1513b	UAUGAGAGAAAGCCAUGAC	19	Gm17	1401433	1401523	5'	
3	gma-MIR1513c	AAAGCCAUGACUUACACACGC	21	Gm20	223679	223766	3'	
4	gma-MIR2109a	GGAGGCGUAGAUACUCACACCU	22	Gm04	28532441	28532537	3'	a(5')*
4	gma-MIR4376a-3p	AGCAUCAUAUCUCCUGCAUAG	21	Gm13	40845925	40846034	3'	a(5')*
5	gma-MIR4413a	AAGAGAAUUGUAAGUCACUG	20	Gm19	1788518	1788617	5'	a
3	gma-MIR4413b	UAAGAGAAUUGUAAGUCACU	20	Gm13	5170460	5170527	5'	
4	gma-MIR4415a-3p	UUGAUUCUCAUCACAACAUGG	21	Gm18	60474198	60474369	3'	a(5')*
3	gma-MIR4415b	UUGAUUCUCAUCACAACAUGG	21	Gm08	23142767	23142922	3'	

[a] The group number refers to: (3) the new miRNAs identified in the conserved soybean miRNAs families; (4) miRNAs originated from the opposite arm of miRNAs previously identified; and (5) miRNAs registered in the miRBase that were detected in our libraries. [b] * miRNAs registered in the miRBase database that were not detected in our libraries.

during the control condition and decreasing when subjected to stress. *MIR4415b* presented an effective rise in its expression level during the water deficit in the sensitive plants, and its high level was also observed in the tolerant genotype independent of the condition. Both sensitive and tolerant genotype exhibited the same expression pattern for *MIR-Seq07* and its level was increased during the stress compared to the control situation.

Expression patterns of miRNAs during soybean rust stress

The RT-qPCR analyses of four libraries from the rust assays are shown in Figure 1B. The differential

Table 5 miRNA isoforms identified in the soybean.

Group[a]	Acronym	Sequence[b]	N isos[c]	Pre-miRNA		
				Ch	Start	End
5	gma-MIR156g	**+2**/ACAGAAGATAGAGAGCAC/**+2**	2	Gm19	8895390	8895493
5	gma-MIR159a-3p	**+2**/TGGATTGAAGGGAGCTCT/**+1**	4	Gm09	37672410	37672586
4	gma-MIR159a-5p	GAGCTCCTTGAAGTCCAATT/**+1**	2	Gm09	37672410	37672586
3	gma-MIR159e-3p	**+2**/TGGATTGAAGGGAGCTC/**+2**	5	Gm07	9524917	9525127
3	gma-MIR166f	TCTCGGACCAGGCTTCATTC/**+1**	2	Gm20	43105394	43105500
5	gma-MIR167g	TGAAGCTGCCAGCATGATCTG/**+1**	2	Gm10	39044877	39044954
3	gma-MIR169g	**+1**/AGCCAAGAATGACTTGCCGG	2	Gm09	5263992	5264096
3	gma-MIR169h	**+1**/AGCCAAGAATGACTTGCCGG	2	Gm14	5324798	5324911
5	gma-MIR172c	**+1**/GAATCTTGATGATGCTGCAG	2	Gm18	2968986	2969138
5	gma-MIR172d	**+1**/GAATCTTGATGATGCTGCAG/**+3**	3	Gm14	5548752	5548901
5	gma-MIR172e	**+1**/GAATCTTGATGATGCTGCAG/**+3**	3	Gm11	35957808	35957960
3	gma-MIR172h-5p	GCAGCAGCATCAAGATTCAC/**+1**	2	Gm10	43474725	43474831
3	gma-MIR172i	GCAGCAGCATCAAGATTCAC/**+1**	2	Gm15	2892962	2893122
3	gma-MIR172j	GCAGCAGCATCAAGATTCAC/**+1**	2	Gm20	40895747	40895836
5	gma-MIR319a	TTGGACTGAAGGGAGCTCCC/**+1**	2	Gm05	40832097	40832279
5	gma-MIR319b	TTGGACTGAAGGGAGCTCCC/**+1**	2	Gm08	1647815	1647987
3	gma-MIR319d	**+2**/GGACTGAAGGGAGCTCCTTC	2	Gm02	43885398	43885595
3	gma-MIR319f	**+1**/TGGACTGAAGGGGAGCTCCTTC	2	Gm04	46348798	46348991
3	gma-MIR319j	**+2**/GGACTGAAGGGAGCTCCTTC	2	Gm14	45953433	45953649
3	gma-miR319k	**+2**/GGACTGAAGGGAGCTCCTTC	2	Gm17	9436178	9436279
4	gma-MIR396a-3p	**+1**/TTCAATAAAGCTGTGGGA/**+2**	3	Gm13	26338134	26338273
5	gma-MIR396a-5p	**+1**/TCCACAGCTTTCTTGAACTG	2	Gm13	26338134	26338273
4	gma-MIR396b-3p	**+1**/CTCAAGAAAGCTGTGGGAGA	2	Gm13	26329931	26330056
5	gma-MIR396d-3p	**+4**/AAGAAAGCTGTGGGAGA/**+7**	3	Gm17	9053051	9053155
4	gma-MIR396d-5p	TTCCACAGCTTTCTTGAACT/**+1**	2	Gm17	9053051	9053155
5	gma-MIR396e	**+1**/TCCACAGCTTTCTTGAACT/**+2**	4	Gm17	35366535	35366668
3	gma-MIR396g	TTCCACAGCTTTCTTGAACT/**+1**	2	Gm15	556707	556796
3	gma-MIR396h-3p	**+1**/TTCAATAAAGCTGTGGGA/**+2**	3	Gm17	9044861	9044973
3	gma-MIR396h-5p	**+1**/TCCACAGCTTTCTTGAACT/**+1**	3	Gm17	9044850	9044984
5	gma-MIR482a-5p	**+12**/AATGGGCTGATTGG/**+5**	5	Gm01	7783819	7783913
5	gma-MIR482b-5p	**+1**/ATGGGGGGATTGGGAAGGA/**+2**	4	Gm20	35360312	35360406
3	gma-MIR482d-5p	TATGGGGGGATTGGGAAGGA/**+2**	3	Gm10	48569629	48569723
5	gma-MIR1507a	**+3**/CATTCCATACATCGTC/**+6**	8	Gm13	25849777	25849883
5	gma-MIR1507b	**+3**/CATTCCATACATCGTC/**+5**	8	Gm17	6190604	6190701
5	gma-MIR1508a	**+4**/GAAAGGGAAATAGCAGT/**+2**	6	Gm16	32903737	32903831
5	gma-MIR1508b	**+2**/GAAAGGGAAATAGCAGTTG	3	Gm09	28530168	28530271
5	gma-MIR1509b	TTAATCAAGGAAATCACGGTT/**+1**	2	Gm05	7774098	7774206
5	gma-MIR1510a	**+3**/TTGTTTTACCTATTCCA/**+6**	7	Gm16	31518908	31519000
5	gma-MIR1510b-3p	TGTTGTTTTACCTATTCCA/**+3**	4	Gm02	6599299	6599392
4	gma-MIR1510b-5p	**+3**/GATAGGTAAAACAACTA/**+2**	5	Gm02	6599299	6599392
5	gma-MIR1511	AACCAGGCTCTGATACCATG/**+1**	2	Gm18	21161236	21161334
5	gma-MIR1514a	TTCATTTTTAAAATAGGCATT/**+1**	2	Gm07	43175810	43175908
5	gma-MIR1523	**+1**/ATGGGATAAATGTGAGCTC/**+1**	2	Gm02	12253303	12253397
5	gma-MIR2109a-5p	TGCGAGTGTCTTCGCCTCTG/**+1**	2	Gm04	28532441	28532537
4	gma-MIR2109a-3p	**+2**/AGGCGTAGATACTCACAC/**+2**	4	Gm04	28532441	28532537
5	gma-MIR4345	**+9**/ACTTACAAAGAT/**+12**	3	Gm14	49069099	49069193
5	gma-MIR4413a	**+1**/AAGAGAATTGTAAGTCACT/**+1**	3	Gm19	1788518	1788617
3	gma-MIRSeq07	**+14**/GACTTG/**+14**	14	Gm01	44787899	44787988
3	gma-MIRSeq14b	**+2**/AGAATTTGGCCTCTGTCCA	2	Gm09	28272488	28272562
3	gma-MIRSeq10-3p	**+20**/G/**+20**	4	Gm15	31542836	31543058

Table 5 miRNA isoforms identified in the soybean. *(Continued)*

3	gma-MIRSeq20	CATCGTTGACGCTGACTGT/**+3**	2	Gm04	35428794	35428950
2	gma-MIR408a	**+1**/TGCACTGCCTCTTCCCTGGC	2	Gm02	837416	837548
2	gma-MIR408c	**+1**/TGCACTGCCTCTTCCCTGGC	2	Gm10	36557005	36557130
2	gma-MIR2218a-5p	GGAGATGGGAGGGTCGGTAA/**+2**	2	Gm10	48574017	48574137
2	gma-MIR3522a	**+8**/AGACCAAATGAGC/**+6**	4	Gm15	4318787	4318873

[a] The group number refers to: (2) the miRNAs previously identified in other plant species as described in Table 2; (3) the new miRNAs identified in the families of conserved miRNAs in soybean; (4) miRNAs originated from the opposite arm of miRNAs previously identified; and (5) miRNAs registered in the miRBase database that were detected in our libraries. [b] Sequence conserved between all isoforms and the number of nucleotide variations in each end. [c] Total number of isoforms (isos) including the typical member for that gene.

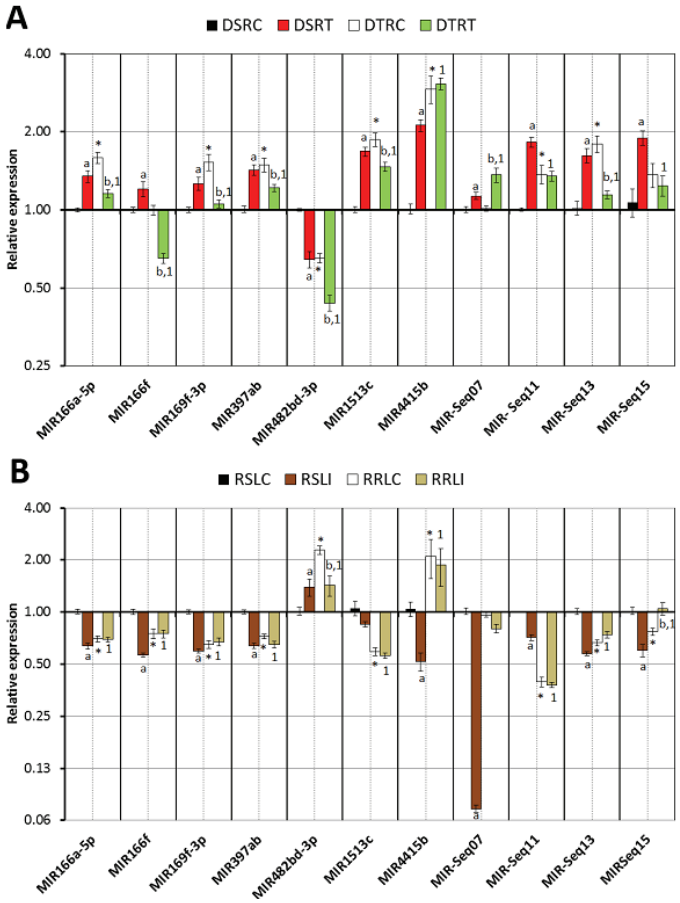


Figure 1 Effects of biotic and abiotic stresses on miRNA relative expression evaluated by RT-qPCR. A) Comparative analyses of four libraries from the water deficit experiment. For the water deficit-stress assay, the four libraries were named as: DSRC (drought-sensitive seedlings root submitted to 0 minutes of stress); DSRT (drought-sensitive seedlings root submitted to 125 minutes and 150 minutes of stress); DTRC (drought-tolerant seedlings root submitted to 0 minutes of stress) and DTRT (drought-tolerant seedlings root submitted to 125 minutes and 150 minutes of stress). B) Comparative analyses of four libraries from the rust infection experiment. For the rust-stress assay, the four libraries were named as: RSLC (rust-susceptible seedlings leaves mock inoculation); RSLI (rust-susceptible seedlings leaves with rust-spore inoculation); RRLC (rust-resistant seedlings leaves with mock inoculation) and RRLT (rust-resistant seedlings leaves with rust-spore inoculation). Samples that significantly differs ($P < 0.05$) according to a Students t-test statistical analysis, were label as: "*" effective differences between cultivars in control conditions; "a" effective differences between control and stressed conditions for sensitive or susceptible plants; "b" effective differences between control and stressed conditions for tolerant or resistant plants and "1" when an effective difference was also observed between sensitive or susceptible and tolerant or resistant under stress conditions.

expression analyses revealed that *MIR166a-5p*, *MIR166f*, *MIR169-3p*, *MIR397ab* and *MIR-Seq13* were dow-regulated in the susceptible genotype during pathogen infection, and equally expressed in the resistant plants. The level of *MIR482bd-3p* did not vary significantly between the two different conditions in the susceptible. However in the resistant genotype, its level is higher during the control condition and decrease with the pathogen attack. *MIR1513c* presented unchangeable expression in the control and stressed condition for both genotypes, but when we compared the two genotypes; the resistant was down-regulated compared to the susceptible. A strong decrease was observed for *MIR4415b* in the rust infection when compared with the control in the susceptible plants, and its level is higher in the resistant genotype showing no expression alteration between the conditions. *MIR-Seq07* was down-regulated with respect to the soybean rust infection in both genotypes. Significant difference was observed in *MIR-Seq11* expression between the mock and infected plants from the susceptible genotypes. This miRNA presented a low expression level after rust inoculation, and its level decreased in the resistant genotypes remaining similar in the both conditions. *MIR-Seq15ab* expression level was significantly decreased in the rust compared to the mock treatment in the susceptible genotype, the opposite occurs in the resistant genotype, when the control showed a lower level of expression compared to the stressed condition.

Target prediction of the soybean miRNAs

MiRNAs suppress gene expression by inhibiting translation, promoting mRNA decay or both [9]. Target gene identification is challenging due to many factors including the following: binding to their target mRNAs by partial complementarity over a short sequence, suppression of an individual target genes is often small, and targeting rules are not completely understood. We predicted the potential miRNAs targets in the psRNA database using all identified miRNAs as queries. The results of the analysis were divided into two tables, showing the targets predicted for the novel (Table 6) and for the conserved miRNAs families (Additional file 2).

Among the 24 novel identified miRNAs families, only 14 families had targets predicted (Table 6). The miRNAs families MIR-Seq01, MIR-Seq03, MIR-Seq06, MIR-Seq07, MIR-Seq08, MIR-Seq12 and MIR-Seq13 had multiple distinct targets. MIR-Seq10, MIR-Seq15 and MIR-Seq18 targeted only one locus. Although, MIR-Seq05, MIR-Seq11, MIR-Seq16 and MIR-Seq19 presented several loci as targets, all of them are coding for the same proteins. Fructose-bisphosphate aldolase, LRR (leucine-rich-repetitions)-containing proteins, translation elongation factor were predicted to be potential targets of the novel MIR-Seq07 which was investigated by RT-qPCR. The search for a target of the novel MIR-Seq11, also analyzed by RT-qPCR, showed a match to *Glycine max* peroxidase precursors mRNAs as potential targets. The oxidoreductase and a transcription regulator factor were predicted to be targeted by MIR-Seq13; and for the MIR-Seq15 only a translation initiator factor was predicted as a target.

After a comparative analysis of our novel identified miRNAs and the degradome libraries of developing soybean seeds it was possible to identify specific sequences in the degradome that corresponds to the downstream sequence of the predicted miRNA recognition site. We identified target sequences to six among the 24 novel soybean miRNAs (MIR-Seq01, MIR-Seq 06, MIR-Seq07, MIR-Seq11, MIR-Seq12 and MIR-Seq16). The list of the 10 identified genes is composed by a glucosyl transferase, serine carboxypeptidase, fructose biphosphate aldolase, three leucine-rich repeat protein, two peroxidases and two ATP dependent RNA helicases (Additional file 3).

Although many soybean conserved miRNAs targets have been predicted and validated by previous studies [26,27,30,31], we also investigated the possible targets for the 28 known families of miRNAs detected in our sequencing. Of these, only 21 families had predicted targets and they are listed in the Additional file 2. The conserved miRNA families showed multiples targets, however families MIR156, MIR172, MIR396, MIR397, MIR1510 and MIR1513 were highly conserved about their targets. For example, all members from the MIR156 family (which had a predicted target) targeted SBP (squamosa promoter binding)-domain protein. AP (2) APETALA 2 transcription factors were targeted by MIR172 family. The same occur with MIR396, MIR397, MIR1510 and MIR1513 families that targeted various genes families as GRF (growth regulating factor) transcription factor, multicopper oxidases, LRR (leucine-rich-repetitions)-containing proteins and F-BOX domain proteins respectively. These results were already observed across several plant species (except for MIR1510 and MIR1513) [25,49-53].

Gene Ontology analysis

The targets of those miRNAs which the expression was analyzed by RT-qPCR were also investigated in respect to their gene ontology (GO) [48]. Among the 11 miRNA genes, six presented target predictions, which were: MIR397ab, MIR1513c, MIR-Seq07, MIR-Seq11, MIR-Seq13 and MIR-Seq15ab. The putative soybean miRNAs targets presented diverse functions, however the most representative group was the proteins involved in oxidoreductase activity followed by the proteins involved in the catabolic process (Figure 2). The result demonstrates that more than 76% of the target proteins

Table 6 Predicted *Glycine max* mRNA targets for the novel miRNAs.

miRNA ID	Locus target[a]	Target description[a]	miRNA/mRNA pairing[b]
gma-MIR-Seq01	Glyma13g01690	glucuronosyl/glucosyl transferase	- | | | | | | | | | | | | | - : |
	Glyma14g35220	glucuronosyl/glucosyl transferase	- | | | | | | | | | | | | | - : |
	Glyma15g00330	GTPase-activating protein	| | | - | | : | | | | | | | | | - |
gma-MIR-Seq03	Glyma08g22900	LRR-containing proteins	- | | | | | | | | | : | | | | | | : |
	Glyma07g03200	LRR-containing proteins	- | | | | | | | | | : | | | | | | : |
	Glyma05g33790	methyltransferase	| - | : | | | | | | | | | | : | : |
	Glyma04g00810	EF-hand-containing proteins	| | : | | | | | - : : | | | | | | | |
	Glyma11g34320	EF-hand-containing proteins	| | : | | | : | | | : - | | | | | | |
	Glyma10g06740	triosephosphate isomerase	– | | | | : | | | | | | | | | | |
gma-MIR-Seq05	Glyma07g18570	pyruvate decarboxylase	| - | | - : | | | | | | | | | | | |
	Glyma01g29190	pyruvate decarboxylase	| - | | - : | | | | | | | | | | | |
	Glyma18g43460	pyruvate decarboxylase	| - | | - : | | | | | | | | | | | |
gma-MIR-Seq06	Glyma08g37480	mt transcription factor	| : : - | | | | | | | : | | | | | | : |
	Glyma16g26070	serine carboxypeptidase	| | | | - | - | : | : | | | | | | | |
gma-MIR-Seq07	Glyma04g01020	fructose-bisphosphate aldolase	| | | | - | : | | | | | | | | | | - |
	Glyma16g05500	LRR-containing proteins	| | : | | : : | : | : | : | | | | | | |
	Glyma19g27280	LRR-containing proteins	| | : | | : : | : | : | : | | | | | | |
	Glyma19g07240	translation elongation factor	| | - | | | - | | | | | | | | | | | -
gma-MIR-Seq08	Glyma14g23860	oxidoreductase activity	| | - | | | | | | | | | | | | | | - |
	Glyma13g03430	oxidoreductase activity	| | - | | | | | | | | | | | | | | - |
	Glyma01g20670	nucleotide excision repair factor	| | | | | | | | - | | | : | | | | | : -
gma-MIR-Seq10	Glyma04g09770	mt oxoglutarate/malate carrier	| : | | | : : | | | | | | : | : | | | | :
gma-MIR-Seq11	Glyma15g13500	peroxidase activity	: : | | - | | | | | | | : | : | | | | | |
	Glyma09g02600	peroxidase activity	| : | | - | | | | | | | : | : | | | | | |
gma-MIR-Seq12	Glyma08g20670	ATP-dependent RNA helicase	: | | | | | | | | | | : | | : | | | | | -
	Glyma07g01260	ATP-dependent RNA helicase	: | | | | | | | | | | : | | : | | | | | -
	Glyma20g16950	predicted alpha/beta hydrolase	| | - | | | | | | | | : | | | | | | : | :
	Glyma10g23470	predicted alpha/beta hydrolase	| | - | | | | | | | | : | | | | | | : | :
	Glyma19g35390	serine/threonine protein kinase	| | | | | | | | – | | - | | | | | | |
	Glyma03g32640	serine/threonine protein kinase	| | | | | | | | – | | - | | | | | | |
gma-MIR-Seq13	Glyma02g26160	oxidoreductase activity	| - | | - | | | | | | | | : | | | | | : |
	Glyma10g31690	transcription regulator activity	| | | | | | | | - | - | | | | | : : | | |
gma-MIR-Seq15	Glyma20g02820	translation initiation factor	– | | | | | | | | | | | | | | | | | |
gma-MIR-Seq16	Glyma17g20860	LRR-containing proteins	| | | | | | | | | | | - | | : | | | | |
	Glyma05g09440	LRR-containing proteins	| | | | | | | | | | | - | | : | | | | |
gma-MIR-Seq18	Glyma11g21200	LRR-containing proteins	| | : | | | | | | | | | | | | | | | –
gma-MIR-Seq19	Glyma15g37290	LRR-containing proteins	- | | | | - | | | | | | | | | | | | -
	Glyma09g34200	LRR-containing proteins	| : | | | : | | | : | - | | | | | : | | | |

[a] The Data from Phytozome version 6.0. [b] Pairing obtained in psRNATarget Server: "|" indicates a Watson-Crick base pairing; ":" is a G:U base pairing, and "-" indicates a mismatch.

are involved in oxidoreductase activity is consistent with the fact that some of the miRNAs libraries are originated from stressed plants. A consequence of many environmental stresses - including water deficit and pathogen attack - is a oxidative stress, i. e. the accumulation of reactive oxygen species (ROS), which damage cellular structures [49,54]. As miRNAs MIR397, MIR-Seq11 and MIR-Seq13 were predicted to match proteins with oxidative activity, they may act in some level of regulation during water deficit or ASR stress.

Discussion

The use of deep-sequencing technology was efficient to identify 256 miRNAs of *Glycine max*. These miRNAs were identified from eight different libraries from precursors with stem-loop secondary structures that also map to the soybean genome (Additional file 1). They were detected from water deficit and rust libraries and were characterized as following: detected for the first time, already detected in some plant species, conserved in soybean, or a variant of a known miRNA (isoform).

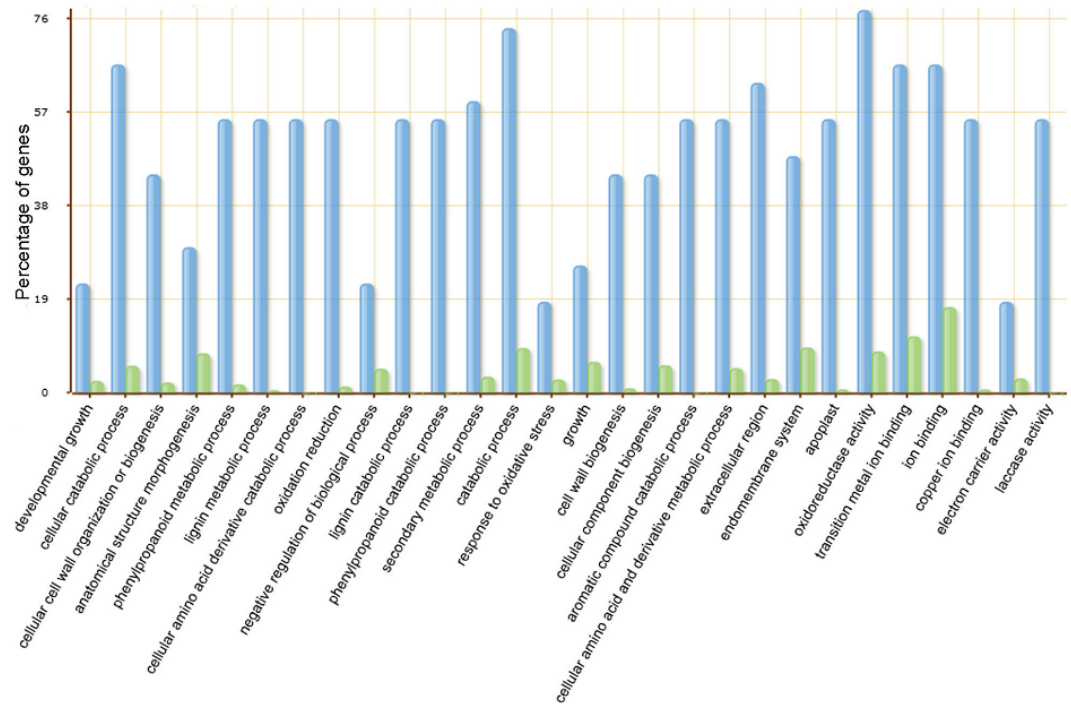


Figure 2 Go analysis of miRNAs target genes. Blue bars indicate the enrichment of miRNA targets in GO terms. Green bars indicate the percentage of total annotated soybean genes mapping to GO terms. Only the predicted target genes for miRNAs analyzed by RT-qPCR were considered.

From these analyses, we found 24 novel families that had not been detected before, six families that had already been detected in Coniferophytes, Embryophytes and Magnoliophytes (dicotyledons and monocotyledons), and 22 conserved soybean families. In terms of conserved soybean miRNAs, we only detected 20 known miRNAs in our sequencing. This small number of known miRNA genes detected in our libraries could be due to the two filters used in our processing. These filters may have missed some known, conserved soybean miRNAs because they discarded reads with low frequency and those with more than five matches in the genome.

We detected 121 miRNAs with additional nucleotides in the 3' or 5' terminus compared to the recorded mature miRNA. These miRNA variants (isomiRNAs) were very common in our population of small, detected RNAs. Out of the isomiRNAs, we observed 21 pairs of sense and antisense miRNAs. The duplex presents the antisense strand paired to the corresponding miRNA with two nucleotides 3' overhangs (Additional file 1). This shows that the sense and antisense miRNAs originated from DCL1 processing and supports their validation as true miRNAs [26,55,56].

In addition, we validated the conserved miRNAs in our libraries based on homology to known miRNAs in miRBase. The phylogenetic conservation of miRNA sequences is one rule proposed by Ambros et al. [7] to characterize miRNAs. In this study, we established new miRNAs in soybean that were already detected in other plants species. However, as opposed to some studies that only blast the candidate to the known miRNA mature sequence, our identifications were determined by precursor sequence folding and verification of the genuine hairpin structures.

The complexity of the plant response to biotic and abiotic stresses involves many genes and biochemical and molecular mechanisms, and adaptation to these stresses is achieved through regulating gene expression at the transcriptional and post-transcriptional levels. With regard to post-transcriptional regulation, miRNAs are associated with water deficit response in others plants, but this was the first time that differential expression of these small RNAs were observed in

soybean during water deficit. In order to validate 11 of the novel miRNAs detected in sequencing by the RT-qPCR method, we constructed primers stem-loop and analyzed their expression during abiotic and biotic stresses (Figure 1). We observed that several miRNAs were up-regulated during the water deficit in the sensitive genotype (Figure 1A). However, during the same stress, these miRNAs had a different expression in the tolerant genotype. This distinct miRNAs behavior between the two contrasting genotypes under the same conditions could be involved with the drought-tolerance that is observed in the tolerant genotype. One of these miRNAs with this expression pattern is the new MIR-Seq11. Interestingly, MIR-Seq11 was predicted to target peroxidase protein. As known, stress conditions can produce excess concentrations of reactive oxygen species (ROS), resulting in oxidative damage at the cellular level [57]. The increase of this miRNA in the sensitive genotype, when subjected to water deficit, could be one of the factors associated with vulnerability of these sensitive plants. Whereas in tolerant genotype during the two conditions, the expression levels of MIR-Seq11 are lower than in the sensitive cultivar during stress. This situation could indicate that the unchangeable MIR-Seq11 levels in the tolerant genotype may be related to its drought-tolerance capacity.

Another interesting point is the expression of a novel miRNA MIR-Seq07 that showed increased expression levels during the water deficit stress for both genotypes. This result allows us to associate this miRNA with water deficit stress mechanism independently of the genotype background. Our computational approach showed that one of the loci targeted by MIR-Seq07 corresponds to a fructose-bipfosphato-aldolase enzyme which is a constituent of both the glycolytic/gluconeogenic pathway and the pentose phosphate cycle in plants [58]. Therefore increase and/or activation of aldolase appear to be implicated in the plant growth mainly through promotion of the glycolytic pathway function to synthesize ATP [58]. Since, MIR-Seq07 expression was increased during the stress condition in both genotypes and assuming that it can inhibit or degrade aldolases, it could be associated to metabolism decreasing during water deficit in roots.

Plants possess several adaptive traits to support pathogen attacks. In *Glycine max*, ASR is responsible for significant losses in soybean growth areas. Nevertheless, no study investigating miRNAs and ASR disease had been preformed to date. To determine if miRNAs act as key factors during rust infection or for resistance maintenance, we performed expression analyses with the same 11 miRNAs during mock and infected conditions in two different genotypes (Figure 1B). In general the miRNAs under the fungus infection were down-regulated in the susceptible genotype (except MIR482bd-3p). For example, MIR-Seq11, MIR-Seq13 and MIR-Seq15 which had predicted peroxidases, oxidoreductases and translational initiation factor respectively as targets proteins, were down regulated when infected with ASR. The peroxidases enzymes help to metabolize H_2O_2 in higher plants, and these proteins, as also others proteins with oxidoreductase activity, have already been reported to be up-regulated after pathogen infection and especially after ASR [57], indicating a possible involvement of MIR-Seq11 and MIR-Seq13 with the responses to ASR infection. Considering, that a translational initiator factor was predicted to be targeted by MIR-Seq15, we could speculate about the participation of this miRNA in the protein synthesis machinery.

In the resistant plants, most of the miRNAs analyzed by RT-qPCR (except MIR482bd-3p, MIR-Seq07, MIR-Seq15ab) did not vary across the mock and rust infection. Surprisingly, MIR-Seq07 was the unique miRNA that was down-regulated during the fungi infection for both genotypes analyzed in our study. We already mentioned that the MIR-Seq07 had predicted protein target related to metabolism and thus its possible association with water stress. However MIR-Seq07 also had predicted LRRs (leucine-rich repeats)-domain target which are known to be present in disease resistance proteins [59,60]. This suggested a good candidate for the investigation of the miRNAs' regulatory mechanisms during ASR stress. Although we investigated the expression patterns of some miRNAs detected in our sequencing and predicted the target genes that it regulates, additional experimental approaches must be addressed to confirm these hypotheses.

Conclusions

The present study detected a large number of small RNA sequences that were characterized as novel and as already known soybean miRNAs. We grouped some of these unique sequences into 24 novel soybean miRNAs and further classified several of new members in known families or as new loci in the soybean genome. Validation of new miRNA with quantitative RT-qPCR revealed that Solexa sequencing is a powerful tool for miRNA discovery. Many miRNA expression patterns were up- or down-regulated by water deficit and rust stresses, which is an important discovery. Future investigations should use supplementary experimental approaches to verify the targets and to understand the complex gene regulatory network of these miRNAs. This work will contribute to improve systems to support soybean crop production and to mitigating crop losses during biotic or abiotic stresses.

Additional material

Additional file 1: Predicted precursor structures of all miRNAs identified. The mature miRNAs (red) and pre-miRNA sequences with chromosome and locus information. The pre-miRNA length (nt) and its directional information (sense (+) or anti-sense (-) compared to the soybean genome sequence) is provided. The fold-back structure with respect to the free energy value (dG) was predicted using the Mfold program.

Additional file 2: Identified targets of known conserved plant miRNAs families. [a] The Data from Phytozome version 6.0. [b] Pairing obtained in psRNATarget Server: "|" indicates a Watson-Crick base pairing; ":" is a G:U base pairing, and "-"indicates a mismatch.

Additional file 3: The soybean transcript loci which were identified as new-miRNA families target by degradome sequencing. The miRNA target site is indicated in red and underlined while the degradome sequence is highlighted.

Acknowledgements

FRK was sponsored by a PhD grant (140578/2009-9), MA by a PDJ (509828/2010-8) and RM by a Productivity and Research grant (303967/2008-0) from the National Council for Scientific and Technological Development (CNPq, Brazil). This work was financially supported by the Biotecsur, GenoSoja consortium (CNPq 5527/2007-8) and GenoProt (CNPq 559636/2009-1).

Author details

[1]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. [2]PPGGBM at Federal University of Rio Grande do Sul - UFRGS, Porto Alegre, RS, Brazil. [3]EMBRAPA Soja, Rodovia Carlos João Strass, Distrito de Warta, CEP 86001-970, Londrina, PR, Brazil. [4]Institute of Biology, Laboratory of Genomic and Expression, State University of Campinas, CEP 13083-970, Campinas, SP, Brazil. [5]National Center for High Performance Processing (CENAPAD-SP), State University of Campinas, CEP 13083-970, Campinas, SP, Brazil.

Authors' contributions

FRK and RM conceived and designed the study. FRK performed the sequence analyses to identify the miRNAs and secondary structures and to predict the target genes conceived, executed the RT-qPCR, performed the data management and processing, and wrote the draft manuscript. RM was the supervisor of this study, provided critical revision, obtained financial support and performed data interpretation. LFVO contributed to the data assembly, prediction and identification de new miRNAs. LM and MA contributed to the analysis of the miRNA secondary structures and processing of the data. FR, JM, JFB and RSM performed the plant experiments and RNA extractions. ALN, FCMG and RVA provided the studied material, critically revised the article for important intellectual content and obtained funding. MFC, GAGP and LCN created the Perl scripts to identify the microRNAs. MFC participated in writing the methods section. All authors read and approved the final version of manuscript.

Received: 13 January 2011 Accepted: 10 June 2011
Published: 10 June 2011

References

1. Mallory AC, Vaucheret H: **Functions of microRNAs and related small RNAs in plants.** *Nat Genet* 2006, **38(Suppl)**:S31-36.
2. Chen X: **MicroRNA biogenesis and function in plants.** *FEBS Lett* 2005, **579(26)**:5923-5931.
3. Lu YD, Gan QH, Chi XY, Qin S: **Roles of microRNA in plant defense and virus offense interaction.** *Plant Cell Rep* 2008, **27(10)**:1571-1579.
4. Shukla LI, Chinnusamy V, Sunkar R: **The role of microRNAs and other endogenous small RNAs in plant stress responses.** *Biochim Biophys Acta* 2008, **1779(11)**:743-748.
5. Bartel B, Bartel DP: **MicroRNAs: at the root of plant development?** *Plant Physiol* 2003, **132(2)**:709-717.
6. Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism, and function.** *Cell* 2004, **116(2)**:281-297.
7. Ambros V, Lee RC, Lavanway A, Williams PT, Jewell D: **MicroRNAs and other tiny endogenous RNAs in C. elegans.** *Curr Biol* 2003, **13(10)**:807-818.
8. Voinnet O: **Origin, biogenesis, and activity of plant microRNAs.** *Cell* 2009, **136(4)**:669-687.
9. Bartel DP: **MicroRNAs: target recognition and regulatory functions.** *Cell* 2009, **136(2)**:215-233.
10. Guo L, Lu Z: **Global expression analysis of miRNA gene cluster and family based on isomiRs from deep sequencing data.** *Comput Biol Chem* **34(3)**:165-171.
11. Ebhardt HA, Fedynak A, Fahlman RP: **Naturally occurring variations in sequence length creates microRNA isoforms that differ in argonaute effector complex specificity.** *Silence* 2010, **1(12)**.
12. Naya L, Khan GA, Sorin C, Hartmann C, Crespi M, Lelandais-Briere C: **Cleavage of a non-conserved target by a specific miR156 isoform in root apexes of *Medicago truncatula*.** *Plant Signal Behav* **5(3)**:328-331.
13. Park W, Li J, Song R, Messing J, Chen X: **CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*.** *Curr Biol* 2002, **12(17)**:1484-1495.
14. Reinhart BJ, Bartel DP: **Small RNAs correspond to centromere heterochromatic repeats.** *Science* 2002, **297(5588)**:1831.
15. Griffiths-Jones S: **The microRNA Registry.** *Nucleic Acids Res* 2004, , **32 Database:** D109-111.
16. Griffiths-Jones S: **miRBase: the microRNA sequence database.** *Methods Mol Biol* 2006, **342**:129-138.
17. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ: **miRBase: microRNA sequences, targets and gene nomenclature.** *Nucleic Acids Res* 2006, , **34 Database:** D140-144.
18. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ: **miRBase: tools for microRNA genomics.** *Nucleic Acids Res* 2008, , **36 Database:** D154-158.
19. Yang T, Xue L, An L: **Functional diversity of miRNA in plants.** *Plant Science* 2007, **172**:423-432.
20. Lelandais-Briere C, Sorin C, Declerck M, Benslimane A, Crespi M, Hartmann C: **Small RNA diversity in plants and its impact in development.** *Curr Genomics* **11(1)**:14-23.
21. Liu Q, Chen Y-Q: **Insights into the mechanism of plant development: interactions of miRNAs pathway with phytormone response.** *Biochem Biophys Res Commun* 2009, **384**:1-5.
22. Chuck G, Candela H, Hake S: **Big impacts by small RNAs in plant development.** *Curr Opin Plant Biol* 2009, **12(1)**:81-86.
23. Katiyar-Agarwal S, Jin H: **Role of small RNAs in host-microbe interactions.** *Annu Rev Phytopathol* 2010, **48**:225-246.
24. Lu XY, Huang XL: **Plant miRNAs and abiotic stress responses.** *Biochem Biophys Res Commun* 2008, **368(3)**:458-462.
25. Sunkar R, Zhu JK: **Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis.** *Plant Cell* 2004, **16(8)**:2001-2019.
26. Subramanian S, Fu Y, Sunkar R, Barbazuk WB, Zhu JK, Yu O: **Novel and nodulation-regulated microRNAs in soybean roots.** *BMC Genomics* 2008, **9**:160.
27. Zhang B, Pan X, Stellwag EJ: **Identification of soybean microRNAs and their targets.** *Planta* 2008, **229(1)**:161-182.
28. Wang Y, Li P, Cao X, Wang X, Zhang A, Li X: **Identification and expression analysis of miRNAs from nitrogen-fixing soybean nodules.** *Biochem Biophys Res Commun* 2009, **378(4)**:799-803.
29. Chen R, Hu Z, Zhang H: **Identification of microRNAs in wild soybean (Glycine soja).** *J Integr Plant Biol* 2009, **51(12)**:1071-1079.
30. Joshi T, Yan Z, Libault M, Jeong DH, Park S, Green PJ, Sherrier DJ, Farmer A, May G, Meyers BC, Xu D, Stacey G: **Prediction of novel miRNAs and associated target genes in Glycine max.** *BMC Bioinformatics* 2010, **11(Suppl 1)**:S14.
31. Song QX, Liu YF, Hu XY, Zhang WK, Ma B, Chen SY, Zhang JS: **Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing.** *BMC Plant Biol* 2011, **11**:5.
32. Desclaux D, Roumet P: **Impact of srought stress on the phenology of two soybean cultivars.** *Field and Crops Research* 1996, **46(1-3)**:61-70.
33. Desclaux D, Huynh TT, Roumet P: **Identification of soybean plant characteristics that indicate the timing of drought stress.** *Crop Science* 2000, **40**:716-722.
34. Van de Mortel M, Recknor JC, Graham MA, Nettleton D, Dittman JD, Nelson RT, Godoy CV, Abdelnoor RV, Almeida AMR, Baum TJ, Whitham SA:

Distinct biphasic mRNA changes in response to Asian soybean rust infection. *MPMI* 2007, **20(8)**:887-899.
35. Sinclair JB, Hartman GL: **Soybean rust.** In *Compendium of soybean diseases*.. 4 edition. Edited by: Hartman ea. St. Paul: American Phytopathological Society; 1999:25-26.
36. Yorinori JT, Paiva WM, Frederick RD, Costa Milan LM, Bertagnolli PF, Hartman GL, Godoy CV, Nunes Junior J: **Epidemics of soybean rust (*Phakopsora pachyrhizi*) in Brazil and Paraguay from 2001 to 2003.** *Plant Dis* 2005, **89**:675-677.
37. Oya T, Nepomuceno AL, Neumaier N, Farias JRB, Tobita S, Ito O: **Drought Tolerance Characteristics of Brazilian soybean cultivars - Evaluation and characterization of drought tolerance of various Brazilian soybean cultivars in the field.** *Plant Prod Sci* 2004, **7(2)**:129-137.
38. Fehr WR, Caviness CE: **Stages of soybean development. Special Report 80.** Ames, USA: Iowa State University of Science and Technology, Iowa Agriculture and Home Economics Experiment Station; 1977.
39. Martins PK, Jordão BQ, Yamanaka N, Farias JRB, Beneventi MA, Binneck E, Fuganti R, Stolf R, Nepomuceno AL: **Differential gene expression and mitotic cell analysis of the drought tolerant soybean (*Glycine max* L. Merrill Fabales, Fabaceae) cultivar MG/BR46 (Conquista) under two water deficit induction systems.** *Genet Mol Biol* 2008, **31**:512-521.
40. Silva DCG, Yamanaka N, Brogin RL, Arias CAA, Nepomuceno AL, Di Mauro AO, Pereira SS, Nogueira LM, Passianotto ALL, Abdelnoor RV: **Molecular mapping of two loci that confer resistance to Asian Rust in soybean.** *Theor Appl Genet* 2008, **117**:57-63.
41. Li R, Li Y, Kristiansen K, Wang J: **SOAP: short oligonucleotide alignment program.** *Bioinformatics* 2008, **24(5)**:713-714.
42. Smith TF, Waterman MS: **Identification of common molecular subsequences.** *J Mol Biol* 1981, **147(1)**:195-197.
43. Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen X, Green PJ, Griffiths-Jones S, Jacobsen SE, Mallory AC, Martienssen RA, Poethig RS, Qi Y, Vaucheret H, Voinnet O, Watanabe Y, Weigel D, Zhu JK: **Criteria for annotation of plant MicroRNAs.** *Plant Cell* 2008, **20(12)**:3186-3190.
44. Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction.** *Nucleic Acids Res* 2003, **31(13)**:3406-3415.
45. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR, Lao KQ, Livak KJ, Guegler KJ: **Real-time quantification of microRNAs by stem-loop RT-PCR.** *Nucleic Acids Res* 2005, **33(20)**:e179.
46. Kulcheski FR, Marcelino-Guimaraes FC, Nepomuceno AL, Abdelnoor RV, Margis R: **The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean.** *Anal Biochem* 2010, **406(2)**:185-192.
47. Zhang Y: **miRU: an automated plant miRNA target prediction server.** *Nucleic Acids Res* 2005, , **33 Web Server**: W701-704.
48. Du Z, Zhou X, Ling Y, Zhang Z, Su Z: **agriGO: a GO analysis toolkit for the agricultural community.** *Nucleic Acids Res* 2010, , **38 Web Server**: W64-70.
49. Jones-Rhoades MW, Bartel DP: **Computational identification of plant microRNAs and their targets, including a stress-induced miRNA.** *Mol Cell* 2004, **14(6)**:787-799.
50. Zhou L, Liu Y, Liu Z, Kong D, Duan M, Luo L: **Genome-wide identification and analysis of drought-responsive microRNAs in *Oryza sativa*.** *J Exp Bot* 2010, **61(15)**:4157-4168.
51. Kantar M, Unver T, Budak H: **Regulation of barley miRNAs upon dehydration stress correlated with target gene expression.** *Funct Integr Genomics* 2010, **10(4)**:493-507.
52. Lu Y, Yang X: **Computational Identification of Novel MicroRNAs and Their Targets in *Vigna unguiculata*.** *Comp Funct Genomics* 2010.
53. Zang YX, Kim HU, Kim JA, Lim MH, Jin M, Lee SC, Kwon SJ, Lee SI, Hong JK, Park TH, Mun JH, Seol YJ, Hong SB, Park BS: **Genome-wide identification of glucosinolate synthesis genes in *Brassica rapa*.** *FEBS J* 2009, **276(13)**:3559-3574.
54. Ramanjulu S, Bartels D: **Drought- and desiccation-induced modulation of gene expression in plants.** *Plant Cell Environ* 2002, **25(2)**:141-151.
55. Axtell MJ, Snyder JA, Bartel DP: **Common functions for diverse small RNAs of land plants.** *Plant Cell* 2007, **19(6)**:1750-1769.
56. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP: **A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*.** *Genes Dev* 2006, **20(24)**:3407-3425.
57. Choi JJ, Alkharouf NW, Schneider KT, Matthews BF, Frederick RD: **Expression patterns in soybean resistant to *Phakopsora pachyrhizi* reveal the importance of peroxidases and lipoxygenases.** *Funct Integr Genomics* 2008, **8(4)**:341-359.
58. Konishi H, Yamane H, Maeshima M, Komatsu S: **Characterization of fructose-bisphosphate aldolase regulated by gibberellin in roots of rice seedling.** *Plant Mol Biol* 2004, **56(6)**:839-848.
59. Belkhadir Y, Subramaniam R, Dangl JL: **Plant disease resistance protein signaling: NBS-LRR proteins and their partners.** *Curr Opin Plant Biol* 2004, **7(4)**:391-399.
60. Moffett P, Farnham G, Peart J, Baulcombe DC: **Interaction between domains of a plant NBS-LRR protein in disease resistance-related cell death.** *EMBO J* 2002, **21(17)**:4511-4519.

doi:10.1186/1471-2164-12-307
Cite this article as: Kulcheski *et al.*: Identification of novel soybean microRNAs involved in abiotic and biotic stresses. *BMC Genomics* 2011 **12**:307.

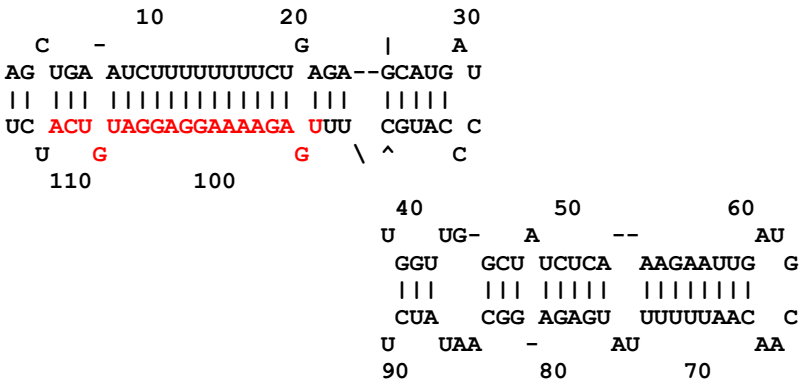
Additional file 1. Predicted precursor structures of all miRNAs identified. The miRNAs (red) and pre-miRNA sequences with chromosome and locus information. The pre-miRNA length (nt) and its direction information (sense (+) or anti-sense (-) compared to the soybean genome sequence) is provided. The fold-back structure with respect to the free energy value (dG) was predicted using the Mfold program.

Identification of novel miRNAs from soybean

```
      SEQ01  GGACAGUCUCAGGUAGACA
      Gm04:30764003-30764171, 169nt, (-)
ACUGUUUCCUGGGAUUGGCUUUGGGCUUUCCUGCACAGCUUAGGUGGAGGGCAAAGAAGACUUCCUUCUGAGGGGGCCAGAGCCAUCAGAGAGAUACCACUCUGGACGAGCUAGAAUUCUAACCUUGUGCCAGGACC
UAUGGGCCGAGGGACAGUCUCAGGUAGACAGU

      Structure 4 Folding bases 1 To 169 of 10Sep29-16-53-28     Initial dG = -67.90


        10        20             30          40           50            60
       -          G    G     -----    -     GCUU   -    .-G   AA----     AC
ACUGUUU CCUGGGAUUG CUUU GGCUU     UCCU GCACA    AGG UGGA   GGC      AGAAG   \
||||||| |||||||||| |||| |||||     |||| |||||    ||| ||||   |||      |||||
UGACAGA GGACUCUGAC GGGA CCGGG     AGGA CGUGU    UCC AUCU   CCG      UCUUC  U
      U          A    G     UAUCC    C     ----   A    \ -   GGGGAG     CU
      160       150       140       130           120            70
                                                          80            90
                                                          AG---|  CA--      AGA
                                                               AGC    UCAGAG   U
                                                               |||    ||||||
                                                               UCG    GGUCUC   A
                                                          UAAGA^  AGCA      ACC
                                                               110       100
```

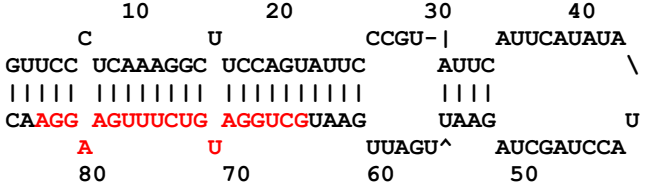
SEQ03 UGAGAAAAGGAGGAUGUCA
Gm11:29821812-29821926, 115nt, (+)
AGCUGAAUCUUUUUUUUCUGAGAGCAUGAUCCCAUGCUGGUUGGCUAUCUCAAAGAAUUGAUGCAACAAUUUUUUAUGAGAGGCAAUAUCUUUUGAGAAAAGGAGGAUGUCAUCU

Structure 1 Folding bases 1 To 115 of 10Sep29-17-04-58 Initial dG = -34.80



SEQ04a GCUGGAUGUCUUUGAAGGA
Gm08:46853906-46853991, 86nt, (+)
GUUCCCUCAAAGGCUUCCAGUAUUCCCGUAUUCAUUCAUAUAUACCUAGCUAGAAUUGAUUGAAUGCUGGAUGUCUUUGAAGGAAC

Structure 1 Folding bases 1 To 86 of 10Sep29-17-09-38 Initial dG = -32.60



SEQ04b GCUGGAUGUCUUUGAAGGA
Gm18:61624611-61624690, 80nt, (-)
UCGUGUUCCCUCAAAGGCUUCCAGUAUUCAUUCAUACCUAACUAGUUGCUUGAAUGCUGGAUGUCUUUGAAGGAAUUUGA

Structure 1 Folding bases 1 To 80 of 10Sep29-17-11-26 Initial dG = -35.40

```
         10         20         30         40
    U     C         U            UU  UACCU
UCG GUUCC UCAAAGGC UCCAGUAUUCA  CA     A
||| ||||| |||||||| ||||||||||||  ||
AGU UAAGG AGUUUCUG AGGUCGUAAGU  GU     A
   U     A         U            UC  UGAUC
         70         60         50
```

SEQ05 AACCCUCAAAGGCUUCCUAG
Gm18:61626669-61626771, 103nt, (+)
CAGUAACCCUCAAAGGCUUCCUAGACUCCAUGUUACGGUCAAAUCAAUUAAUCGUUGAUUAGGAAUAAUUAAGAGUUUCGGAAGUAACUUUGGGGGUUAACUG

Structure 1 Folding bases 1 To 103 of 10Sep29-17-12-48 Initial dG = -41.30

```
          10          20         30         40         50
    -             --       U      CAU    CGG  A       UA
CAG UAACCCUCAAAG  GCUUCC AGACUC   GUUA   UC AAUCAAU  \
||| ||||||||||||  |||||| ||||||   ||||   || |||||||
GUC AUUGGGGGUUUC  UGAAGG UUUGAG   UAAU   GG UUAGUUG  A
   A             AA       C      AAU    AA-  A       CU
 100         90         80         70          60
```

SEQ06 AGUGGAACUUUGAGGCCUGC
Gm08:46848259-46848354, 96nt, (+)
CUUUAGCAACCCUCAAAGGCUUCCACUACUCCAUAUUUCAGUCUAGUGAAUGUCCACAAACAUGGAGGAGUAGUGGAACUUUGAGGCCUGCUGAAG

Structure 2 Folding bases 1 To 96 of 10Sep29-17-14-16 Initial dG = -45.10

```
         10         20         30         40         50
         AC         GC              AUAU|   GUCUA   AA
CUUUAGCA  CCUCAAAG  UUCCACUACUCC     UUCA     GUG  U
||||||||  ||||||||  ||||||||||||     ||||     |||
GAAGUCGU  GGAGUUUC  AAGGUGAUGAGG     AGGU     CAC  G
         CC         --             ----^   ACAAA   CU
      90         80          70             60
```

```
SEQ07               AAAUGACUUGAGAGGUGUAG
SEQ07(iso1)         CGACUUUGUGAAAUGACUUG
SEQ07(iso2)         UGAAAUGACUUGAGAGGUGUAG
SEQ07(iso3)         GAAAUGACUUGAGAGGUGUAGG
SEQ07(iso4)         GAAAUGACUUGAGAGGUGUAG
SEQ07(iso5)         AAUGACUUGAGAGGUGUAGGAU
SEQ07(iso6)         AAAUGACUUGAGAGGUGUAGGA
SEQ07(iso7)         AAUGACUUGAGAGGUGUAGGA
SEQ07(iso8)         AAUGACUUGAGAGGUGUAGG
SEQ07(iso9)         AAUGACUUGAGAGGUGUAG
SEQ07(iso10)         AUGACUUGAGAGGUGUAGG
SEQ07(iso11)          UGACUUGAGAGGUGUAGGA
SEQ07(iso12)            GACUUGAGAGGUGUAGGAUA
SEQ07(iso13)               UUGAGAGGUGUAGGAUAAG

Gm01:44787899-44788252,   90nt, (+)
UUAGUUCGACUUUGUGAAAUGACUUGAGAGGUGUAGGAUAAGUGGGGAGCAAUCCUCACCUUAUAAGUCGGUUUUGUAGGGUUGAGUUAA

Structure 1 Folding bases 1 To 90 of 10Sep29-17-19-09       Initial dG = -36.60
          10         20         30         40
U  GU          UG   -       A      U     AA  GG
 UA  UCGACUUUG  AAA UGACUUG GAGGUG AGGAU  GU  \
 ||  |||||||||  ||| ||||||| |||||| |||||  ||
 AU  AGUUGGGAU  UUU GCUGAAU UUCCAC UCCUA  CG  G
A  UG          GU   G       A      -     A-  AG
          80         70         60          50

SEQ08  CUAAAGAUUGUCCAAAAGGAA
Gm14:6756499-6763456,153nt,                                                                                                    (+)
GUUUGCACUAAAGAUUGUCCAAAAGGAAUAUUGUUUUAUGCAGAAGACAAAUUACAUCGGAUAAGAAAAAUCACACCACAUAAUUGCUGCAUUGCUUUAGAUCUCUAUUUCUGCAUAAAACAAUAUUCCUUUUGGAC
AAUCUUUAGUGCAAAC
```

```
Structure 1 Folding bases 1 To 153 of 10Sep29-17-22-09     Initial dG = -73.40

        10         20         30         40          50         60
                                                  .-ACAAAUUACA| GG
GUUUGCACUAAAGAUUGUCCAAAAGGAAUAUUGUUUUAUGCAGAAG             UC  A
||||||||||||||||||||||||||||||||||||||||||||||             ||
CAAACGUGAUUUCUAACAGGUUUUCCUUAUAACAAAAUACGUCUUU             AG  U
                                                 \ ----------^ AA
 150        140        130        120        110

                                                    70         80         90
                                                   AAAAUCACACCACAUAAUU  UG
                                                                      GC  C
                                                                      ||
                                                                      CG  A
                                                   AUCUCUAGAUUU-------  UU
                                                        100
```

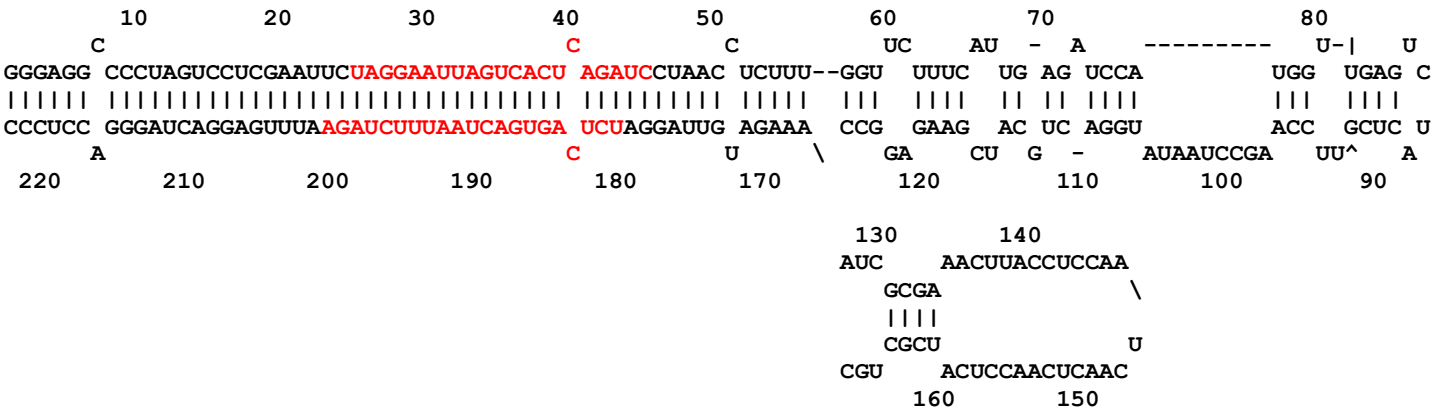
```
SEQ09  GUAGUGGAUGCCUAGAGGUCC
Gm18:61655979-61656075, 97nt, (-)
GCUUUUUAGUGGCCUCAAAGGCUUCCACUACUGCAUGUUUCUGUGGUUUAAUAUCCUGAAACAAGUGGUAGUGGAUGCCUAGAGGUCCAUUAAAAGC

Structure 1 Folding bases 1 To 97 of 10Sep29-17-30-58      Initial dG = -48.90

        10          20         30         40         50
      U     -    AA    U         UG A      UGU  UUU
GCUUUU AGUGG CCUC  AGGC UCCACUAC  C UGUUUC   GG   A
|||||| ||||| ||||  |||| ||||||||  | ||||||   ||
CGAAAA UUACC GGAG  UCCG AGGUGAUG  G ACAAAG   CC   A
      -     U    A-    U         GU A      U--  UAU
       90          80         70         60
```

```
SEQ10-5p     UAGGAAUUAGUCACUCAGAUC
SEQ10-3p     UCUCAGUGACUAAUUUCUAGA
SEQ10-3p iso1AUCUCAGUGACUAAUUUCUAG
SEQ10-3p iso2GAAUUUGAGGACUAGGGACCUC
```

SEQ10-3p iso3UAAUUUCUAGAAUUUGAGGACU

Gm15:31542836-31543058, 223nt, (+)
GGGAGGCCCCUAGUCCUCGAAUUCUAGGAAUUAGUCACUCAGAUCCUAACCUCUUUGGUUCUUUCAUUGAGAUCCAUGGUUGAGUCUACUCGUUCCAAGCCUAAUAUGGACUGCAUCGAAGAGGCCAUCGCGAAACU
UACCUCCAAUCAACUCAACCUCAUCGCUGCAAAGAUGUUAGGAUCUCAGUGACUAAUUUCUAGAAUUUGAGGACUAGGGACCUCCC

Structure 3 Folding bases 1 To 223 of 10Sep29-18-10-02 Initial dG = -110.20



SEQ11 UUGUUCGAUAAAACUGUUGUG
Gm16:5743687-5744863, 65nt, (-)
UUGCAUGGUUGUUCGAUAAAACUGUUGUGAUAAUGUACAACACAAUUAUCGAUAGCUUAUGCAAA

Structure 1 Folding bases 1 To 65 of 10Sep29-17-37-15 Initial dG = -23.40

```
          10        20        30
-      -      U      AAC       AU
 UUGCAU GGUUGU CGAUAA   UGUUGUG  A
 |||||| |||||| ||||||   |||||||
 AACGUA UCGAUA GCUAUU   ACAACAU  A
A      U      -      AAC       GU
    60         50        40
```

SEQ12 UCUCUUGAUUCUAGAUGAUGU

Gm16:27653048-27653102, 55nt, (+)

UCAUAGAGUCUAGGCUCACGGGAAAGAAGAUUCUCUUGAUUCUAGAUGAUGUUGA

Structure 1 Folding bases 1 To 55 of 10Sep29-17-40-29 Initial dG = -16.50

```
         10        20
  -   GA       C   C     AG
UC AUA  GUCUAGG UCA GGGAA  A
|| |||  ||||||| ||| |||||
AG UGU  UAGAUCU AGU CUCUU  A
  U   AG       U   U      AG
    50        40        30
```

SEQ13 UGUUGCGGGUAUCUUUGCCUC

Gm04:28578972-28579075, 104nt, (-)

GCCAGCAAAACUGUUGCGGGUAUCUUUGCCUCUGAAGGAAAGUUGUGCCUAUUAUUAUGGCUUAUUGCUUUAGUGGCGUAGAUCCCCACAACAGUUAUGCUUGC

Structure 1 Folding bases 1 To 104 of 10Sep29-17-44-58 Initiald G = -47.10

```
        10        20        30          40        50
  C    A        C   U    U    U      G --    UG  CCU
GC AGCA AACUGUUG GGG AUCU UGCC CUGAAG A  AAGU  UG   A
|| |||| |||||||| ||| |||| |||| |||||| |  ||||  ||
CG UCGU UUGACAAC CCC UAGA GCGG GAUUUC U  UUCG  AU   U
  U    A        A   C    U    U      G UA    GU  UAU
  100        90        80        70        60
```

SEQ14a UGAGAAUUUGGCCUCUGUCCA
Gm09:28264427-28264514, 88nt, (+)
AAAUUUUCUUGAGAAUUUGGCCUCUGUCCAUGUCUAAUUAUUAAUUCCAAUAAUUGAGAUGGAUAGAGCCCAAAUUCUCAAGAGAAUUU

Structure 1 Folding bases 1 To 89 of 10Sep29-17-47-47 InitialdG = -48.50

```
        10         20         30         40
                     C          G  U       AA
AAAUUUUCUUGAGAAUUUGG CUCUGUCCAU UC AAUUAUU  U
|||||||||||||||||||| |||||||||| || |||||||
UUUAAGAGAACUCUUAAACC GAGAUAGGUA AG UUAAUAA  U
                     C          G  -       CC
        80         70         60          50
```

SEQ14b UGAGAAUUUGGCCUCUGUCCA
SEQ14b(iso) AGAAUUUGGCCUCUGUCCA
Gm09:28272488-28272562, 75nt, (+)
CUUGAGAAUUUGGCCUCUGUCCAUGUCUAAUUAUUAAUUCCAAUAAUUCAGAUGGACAGAGCCCAAAUUCUCAAG

Structure 1 Folding bases 1 To 75 of 10Sep29-17-49-1 Initial dG = -38.80

```
        10         20          30
             C          UG   -       AA
CUUGAGAAUUUGG CUCUGUCCA  UCU AAUUAUU  U
|||||||||||||| |||||||||  ||| |||||||
GAACUCUUAAACC GAGACAGGU  AGA UUAAUAA  U
             C          --   C       CC
    70         60          50         40
```

SEQ15a UUAGAUUCACGCACAAACUUG
Gm02:1041996-1042084, 89nt, (+)
AAAUUGCAGGUUCGUGCGUGAAUCUAACGAAAGUUUCUCGUUCUCUCCAUUCCACUUCUGCGUUAGAUUCACGCACAAACUUGUCAUUU

Structure 1 Folding bases 1 To 89 of 10Sep29-17-53-59 Initial dG = -34.80

```
        10         20         30          40
    U       C                  A--     UUCUCGUUC
AAAU GCAGGUU GUGCGUGAAUCUAACG    AAGU          \
|||| ||||||| ||||||||||||||||    ||||
UUUA UGUUCAA CACGCACUUAGAUUGC    UUCA          U
    C       A                  GUC     CCUUACCUC
        80         70         60          50
```

SEQ15b UUAGAUUCACGCACAAACUUG

Gm10:1085223-1085322, 100nt, (+)

GAGACAGAGGCAAAUCGCAGGUUCGUGCGUGAAUCUAAUCAAAGUUUCUCAUUAACUUCUGCGUUAGAUUCACGCACAAACUUGUCAUUUCCCUUUUCUU

Structure 1 Folding bases 1 To 100 of 10Sep29-17-55-46 Initial dG = -43.90

```
          10         20         30          40
    C     C    C       C                -  -     UCU
GAGA AGAGG AAAU GCAGGUU GUGCGUGAAUCUAAU CA AAGUU    \
|||| ||||| |||| ||||||| ||||||||||||||| || |||||
UUCU UUUCC UUUA UGUUCAA CACGCACUUAGAUUG GU UUCAA    C
    -     C    C       A                C  C     UUA
          90         80         70          60
```

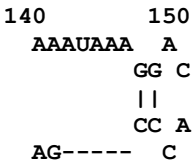
SEQ16 UUAUAGUCUGACAUCUGGAAU

Gm05:9279518-9279737, 220nt, (+)

CAUCGAGAUUUUAUAGUCUGACAUCUGGAAUUUAAAUGUCUCAACAAGGGCGUUAGAUCCAACAACGUUCCACAAUGCUAAUGGGAGAAUUCAAAUUUGGGACCUCUCUCUCAUUUCUUCUCCUUUCACGCUUUGAA
ACAAAUAAAGGACACCCGAGAUCCCCGUAGAAGCUCUUGUUGGAAAAUUUAGAUUUCAAAUGUCAGAUUAUAAAACAUGGGUG

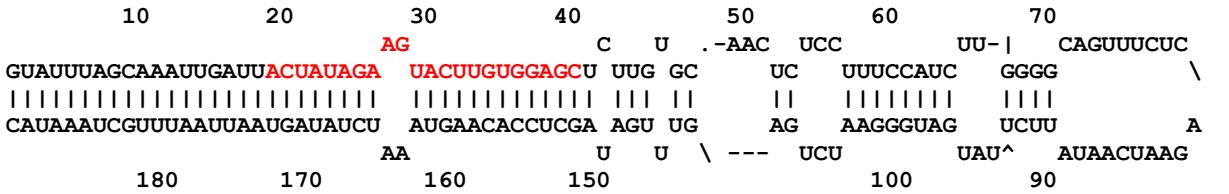
Structure 1 Folding bases 1 To 220 of 10Sep29-17-57-12 Initial dG = -67.30

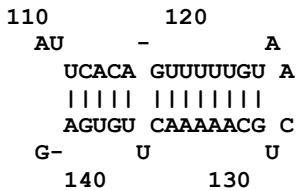
```
        10         20         30         40         50                60         70         80          90        100
    GAGA                C            GUC           GUUA-----     .-CAACAAC|  CCA    U  UAAU-        UUC     UU     C
CAUC     UUUUAUAGUCUGACAU UGGAAUUUAAAU   UCAACAAGGGC          GAUC          GUU   CAA GC     GGGAGAA    AAAU  GGGA C
||||     |||||||||||||||| ||||||||||||   ||||||||||||          ||||          |||   ||| ||     |||||||    ||||  ||||
GUGG     AAAAUAUUAGACUGUA ACUUUAGAUUUA   GGUUGUUCUCG          CUAG          CAA   GUU CG     UCCUCUU    UUUA  CUCU U
    GUAC                A            AAA           AAGAUGCCC     \ -------^  A--    U  CACUU        C--     CU     C
.        210        200        190        180        170        160                    130        120         110
```



SEQ17 ACUAUAGAAGUACUUGUGGAGC
Gm16:2916844-2917034, 191nt, (+)
GUAUUUAGCAAAAUUGAUUACUAUAGAAGUACUUGUGGAGCUCUUGUGCAACUCUCCUUUCCAUCUUGGGGCAGUUUCUCAGAAUCAAUAUUCUUAUGAUGGGAAUCUGAAUUCACAGUUUUUGUAACUGCAAAAACU
UGUGAGGUUUGAUAGCUCCACAAGUAAAUCUAUAGUAAUUAAUUUGCUAAAUAC

Structure 3 Folding bases 1 To 191 of 10Sep29-18-02-46 Initial dG = -83.00





SEQ18 CCUCAUUCCAAACAUCAUCUAA
Gm09:16565935-16566025, 91nt, (-)

GUUUGCUAGAGGUGUUUGGGAUGAGAGAAUAGAAUUUUUUCAAAUGCUUGAAAGUGAUCUCUUCCCUCAUUCCAAACAUCAUCUAACACAC

Structure 1 Folding bases 1 To 91 of 10Sep29-18-04-30 Initial dG = -36.30

```
        10         20         30        40
  U  C    -               A   U   AUU       A
GU UG UAGA GGUGUUUGGGAUGAG GAA AGA   UUUUCAA U
|| || |||| |||||||||||||||| ||| |||   |||||||
CA AC AUCU CUACAAACCUUACUC CUU UCU   GAAAGUU G
  C  A    A               C   C   AGU       C
90         80         70        60         50
```

SEQ19 UGAAGAUUUGAAGAAUUUGGGA
Gm15:16900161-16900327, 167nt, (+)
UGAAGAUUUGAAGAAUUUGGGAGAAGGACGCCGUCAAGGUCGAGGGUUUCGUGACUACAGCUUCUGAAGCACGUCAUCUUCACAUAAAUAAGACGCGUUUCAGAAGUUGUAGUCACGAAACUCUCGACCUAGACCGC
AGCCCUUCUCCCAAAUUCUUCAAAUCUUCA

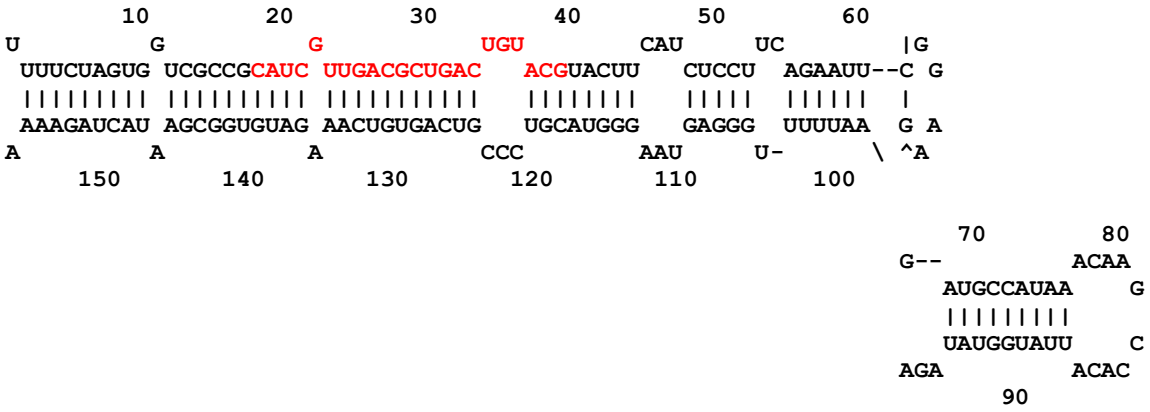
Structure 1 Folding bases 1 To 167 of 10Sep29-18-07-13 Initial dG = -119.10

```
         10         20         30        40         50         60         70         80
U                          AC-| C   A                                    A    AUCUUCA
 GAAGAUUUGAAGAAUUUGGGAGAAGG   GC GUC AGGUCGAGGGUUUCGUGACUACAGCUUCUGAAGC CGUC       C
 |||||||||||||||||||||||||    || ||| ||||||||||||||||||||||||||||||||||| ||||
 CUUCUAAACUUCUUAAACCCUCUUCC   CG CAG UCCAGCUCUCAAAGCACUGAUGUUGAAGACUUUG GCAG       A
A                          CGA^ C   A                                    C    AAUAAAU
     160        150        140        130        120        110        100        90
```

SEQ20 CAUCGUUGACGCUGACUGUACG
SEQ20(iso) CAUCGUUGACGCUGACUGU
Gm04:35428794-35428950, 157nt, (-)

UUUUCUAGUGGUCGCCGCAUCGUUGACGCUGACUGUACGUACUUCAUCUCCUUCAGAAUUCGGAAGGAUGCCAUAAACAAGCCACAUUAUGGUAUAGAAAUUUUUGGGAGUAAGGGUACGUCCCGUCAGUG
UCAAAGAUGUGGCGAAUACUAGAAAA

Structure 1 Folding bases 1 To 157 of 10Sep29-18-11-54 Initial dG = -75.00

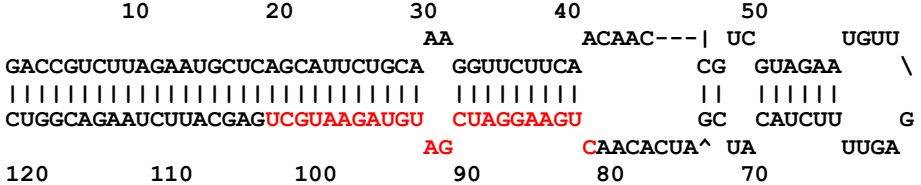


SEQ21 CUGAAGGAUCGAUGUAGAAUGCU

Gm02:39825520-39825641, 122nt, (+)

GACCGUCUUAGAAUGCUCAGCAUUCUGCAAAGGUUCUUCAACAACCGUCGUAGAAUGUUGAGUUUUCUACAUCGAUCACAACUGAAGGAUCGAUGUAGAAUGCUGAGCAUUCUAAGACGGUC

Structure 1 Folding bases 1 To 122 of 10Sep29-18-13-11 Initial dG = -70.70

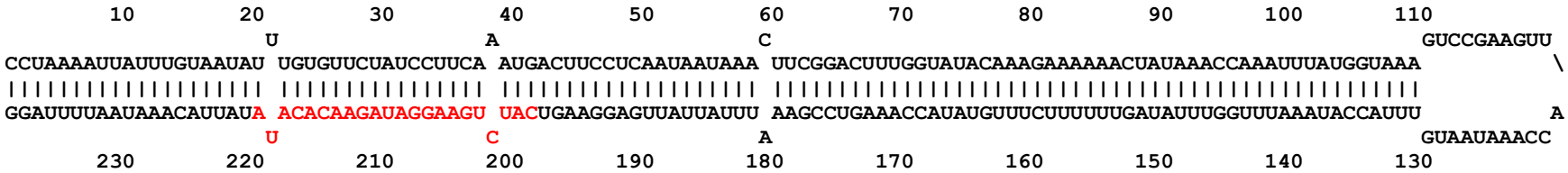


SEQ22 CAUCUGAAGGAUAGAACACAUA

Gm09:29816467-29816705, 239nt, (+)

CCUAAAAUUAUUUGUAAUAUUUGUGUUCUAUCCUUCAAAUGACUUCCUCAAUAAUAAACUUCGGACUUUGGUAUACAAAGAAAAAACUAUAAACCAAAUUUAUGGUAAAGUCCGAAGUUACCAAAUAAUGUUUACCA
UAAAUUUGGUUUAUAGUUUUUUCUUUGUAUACCAAAGUCCGAAAUUUAUUAUUGAGGAAGUCAUCUGAAGGAUAGAACACAUAUAUUACAAAUAAUUUUAGG

Structure 1 Folding bases 1 To 239 of 10Sep29-18-15-19 Initial dG = -153.80

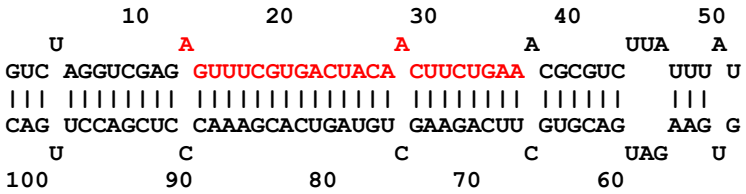


SEQ23 AGUUUCGUGACUACAACUUCUGAA

Gm15:16900193-16900294, 102nt, (-)

GUCUAGGUCGAGAGUUUCGUGACUACAACUUCUGAAACGCGUCUUAUUUAUGUGAAGAUGACGUGCUUCAGAAGCUGUAGUCACGAAACCCUCGACCUUGAC

Structure 1 Folding bases 1 To 102 of 10Sep29-18-22-02 Initial dG = -63.30

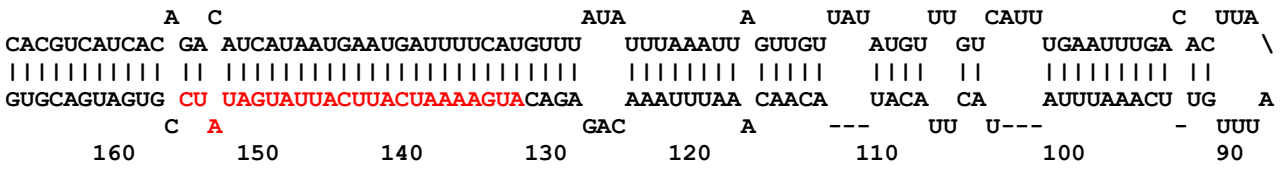


SEQ24 AUGAAAAUCAUUCAUUAUGAUAUC

Gm16:28536014-28536181, 168nt, (-)

CACGUCAUCACAGACAUCAUAAUGAAUGAUUUUCAUGUUUAUAUUUAAAUUAGUUGUUAUAUGUUUGUCAUUUGAAUUUGACACUUAAUUUGUUCAAAUUUAUACUUACAUACAACAAAUUUAAACAGAGACAUGAA
AAUCAUUCAUUAUGAUAUCCGUGAUGACGUG

Structure 1 Folding bases 1 To 168 of 10Sep29-18-24-25 Initial dG = -73.40



SEQ25a GAAAAUGAAUGAUGAGGAUGGGGA

Gm11:7787358-7787494, 137nt, (+)

GUAACUCUUUUAGAGUUACUUCUCAUCCUCACCAUUCAUUUUUUUAAGAUCUAAUGGUUAAAAUUAGUUACUUAUUACAACCAUUAGAUUUUAAGAAAAUGAAUGAUGAGGAUGGGGAGUAACUCUAAAAAAGUUAC

Structure 1 Folding bases 1 To 137 of 10Sep29-18-26-31 Initial dG = -77.60

```
        10        20        30        40        50        60
      C                       C                         AA   U  U
GUAACU UUUUAGAGUUACUUCUCAUCCUCA CAUUCAUUUUUUUUAAGAUCUAAUGGUU  AAU AG \
|||||| ||||||||||||||||||||||||| |||||||||||||||||||||||||||  ||| ||
CAUUGA AAAAUCUCAAUGAGGGGUAGGAGU GUAAGUAAAAGAAUUUUAGAUUACCAA  UUA UC U
      A                       A                         CA   U  A
     130       120       110       100       90        80        70
```

SEQ25b GAAAAUGAAUGAUGAGGAUGGGGA

Gm11:7821070-7821206, 137nt, (-)

GUAACUCUUUUAGAGUUACUCCUCAUCCUCACCAUUCAUUUUCUUAAGAUGUAAUGGUUAAAAUUAGUUACUUAUUACAACCAUUAGAUUUUAAGAAAAUGAAUGAUGAGGAUGGGGAGUAACUCUAAAAA
AGUUAC

Structure 1 Folding bases 1 To 137 of 10Sep29-18-27-35 Initial dG = -79.00

```
        10        20        30        40        50        60
      C                       C                   G     AA   U  U
GUAACU UUUUAGAGUUACUCCUCAUCCUCA CAUUCAUUUUCUUAAGAU UAAUGGUU  AAU AG \
|||||| ||||||||||||||||||||||||| |||||||||||||||||| ||||||||  ||| ||
CAUUGA AAAAUCUCAAUGAGGGGUAGGAGU GUAAGUAAAAGAAUUUUA AUUACCAA  UUA UC U
      A                       A                   G     CA   U  A
     130       120       110       100       90        80        70
```

Identification of homologues miRNAs of other plant species

gma-MIR170 UAUUGGCCUGGUUCACUCAGA

Gm02:4014001-4014153, 153nt, (-)

AUGAAGUAGUUAUUGUGAUAUUGGCCUGGUUCACUCAGACAUGUAUACCACGGCACGGUUGUGUCCUUGCUAGAAAGAUUUAAAACAAAAUACAUGUUGUGAUUGUGGUUUUGGGUUUGAUUGAGCCGUGCCAAUAU
CUCAGUGCUAUUUCAU

Structure 1 Folding bases 1 To 153 of 10Sep29-18-30-26 Initial dG = -61.30

```
         10         20         30         40         50          60         70
          U      U          C        C       AUGUAU        -   GG -      CC   C--|   A
AUGAAGUAG UAUUG GAUAUUGGC UGGUUCA UCAGAC       ACCACGG CAC  U UGUGU  UUG   UAGA A
||||||||| ||||| ||||||||| ||||||| ||||||       ||||||| |||  | |||||  |||   ||||
UACUUUAUC GUGAC CUAUAACCG GCCGAGU AGUUUG       UGGUGUU GUG  G ACAUA  AAC   AUUU G
          -     U          U        U       GGUUU-        A   UU U     A-   AAA^   A
 150         140        130        120        110         100         90         80
```

gma-MIR395a CUGAAGUGUUUGGGGGAACUC

Gm01:4818581-4818690, 110nt, (+)

UCAGGUUUUCCCUAGAGUUCCCCUGAACGCUUCAUUAAAGGGCUUUAUUAUCAUAUAGUCCCAAGUUAGUCCAUACUGAAGUGUUUGGGGGAACUCCCGGUGAUACUUGA

Structure 2 Folding bases 1 To 110 of 10Sep29-18-31-25 Initial dG = -47.50

```
          10         20         30         40         50
       U UC  UA          UG          U  AA|    UUAUUAUCAUA
UCAGGU U  CC  GAGUUCCCC  AACGCUUCA UA  GGGCU            \
|||||| |  ||  |||||||||  ||||||||| ||  |||||
AGUUCA A  GG  CUCAAGGGG  UUGUGAAGU AU  CCUGA            U
       U GU  CC          GU          C  A-^    UUGAACCCUGA
          100         90         80          70         60
```

gma-MIR395b CUGAAGUGUUUGGGGGAACUC
Gm02:1723449-1723543, 95nt, (-)
UCCCUAGAGUUUCCCUGAACACUUCAUUAAAGGGCUUUAUUGAAAUAUAAUCCAAAGUUAGUCUAUACUGAAGUGUUUGGGGGAACUCCUGGUGA

Structure 1 Folding bases 1 To 95 of 10Sep29-18-34-15 Initial dG = -38.30

```
          10         20         30          40
  -| UA          UG          U  AA      UUA-   AAAU
UC CC  GAGUUUCCC  AACACUUCA UA  GGGCU    UUG    A
|| ||  |||||||||  ||||||||| ||  |||||    |||
AG GG  CUCAAGGGG  UUGUGAAGU AU  UCUGA    AAC    U
  U^ UC          GU          C  A-      UUGA   CUAA
    90         80         70          60         50
```

gma-MIR395c CUGAAGUGUUUGGGGGAACUC
Gm08:40840226-40840312, 87nt, (+)
CCUAGAGUUCCCCUUAAUGCUUCAUUGAGGAUUCUGUUUAGGUCCAAUUUAACUAGUCCCUACUGAAGUGUUUGGGGGAACUCCGGG

Structure 1 Folding bases 1 To 87 of 10Sep29-18-35-47 Initial dG = -38.30

```
         10         20         30         40
| UA          U          U  A     CU   U  GU
CC  GAGUUCCCCU AAUGCUUCA UG GGAUU  GUU AG  C
||  |||||||||| ||||||||| || |||||  ||| ||
GG  CUCAAGGGGG UUGUGAAGU AU CCUGA  CAA UU  C
^ GC           U          C  C     U-   U  AA
      80         70         60          50
```

gma-MIR397a UCAUUGAGUGCAGCGUUGAUG
Gm08:4639046-4639153, 108nt, (-)
AGAGAAACAUCAUUGAGUGCAGCGUUGAUGAAGUUUCACUCUCAUCUCAGGUAGAUGCUUAAUUUAUAGUGUUAUUGUCAUCGACACUGCACUCAAUCAUGUUUUUCU

Structure 1 Folding bases 1 To 108 of 10Sep29-18-37-16 Initial dG = -46.10

```
         10         20         30          40
           C          C        -    UU    CU--------|    CA
AGAGAAACAU AUUGAGUGCAG GUUGAUGA AGU  CACU           CAUCU  \
|||||||||| |||||||||| |||||||| |||  ||||           |||||
UCUUUUUGUA UAACUCACGUC CAGCUACU UUA  GUGA           GUAGA  G
           C          A        G   UU    UAUUUAAUUC^    UG
       100         90         80         70         60
```

gma-MIR397b UCAUUGAGUGCAGCGUUGAUG

Gm13:34382999-34383120, 122nt, (-)

GGAGAAACAUCAUUGAGUGCAGCGUUGAUGAAGUCCUAAAUAGUAAAUAGCAACUACCCCUCGUGUUUCUCAGGUAGAUGCUUAAAUGGUUUUAUUGACGCUGCACUCAAUCAUGUUUUUUU

Structure 1 Folding bases 1 To 122 of 10Sep29-18-40-16 Initial dG = -51.10

```
         10         20         30         40         50         60
           C              UG       U  UAAAUAGUAAAU    A     CCUCGU
GGAGAAACAU AUUGAGUGCAGCGU  AUGAAG CC             AGCA CUACC      G
|||||||||| ||||||||||||||  |||||| ||             |||| |||||
UUUUUUUGUA UAACUCACGUCGCA  UAUUUU GG             UCGU GAUGG      U
           C              GU       -  UAAAU-------    A     ACUCUU
120        110        100         90                 80         70
```

gma-MIR408a AUGCACUGCCUCUUCCCUGGC

gma-MIR408a(iso) UGCACUGCCUCUUCCCUGGC

Gm02:837416-837548, 133nt, (+)

GAGACAGGACAAAGCAGGGGAACAGGCAGAGCAUGGAUGGAGCUAUCAACACAAUAUUGUCAAGAAACUGAGAGUGAGAGGAGAAAUAUGUUGUGGUUCUGCUCAUGCACUGCCUCUUCCCUGGCUCUGUCUC

Structure 1 Folding bases 1 To 133 of 10Sep29-18-42-13 Initial dG = -58.30

```
         10         20         30         40         50         60         70
       GACAA    A       C       A      GA         U      CAA     G-  AAGAA  UG
GAGACAG      AGC GGGGAA AGGCAG GCAUG  UGGAGCUA CAACA    UAUU  UC     AC  \
|||||||      ||| |||||| |||||| |||||  |||||||| |||||    ||||  ||     ||
CUCUGUC      UCG UCCCUU UCCGUC CGUAC  GUCUUGGU GUUGU    AUAA  AG     UG  A
       -----    G       C       A      UC         -      ---     AG  GAGAG  AG
 130             120        110        100         90            80
```

```
gma-MIR408b-5p      CUGGGAACAGGCAGGGCACG
gma-MIR408b-3p      AUGCACUGCCUCUUCCCUGGC
Gm03:44626696-44626827, 132nt, (-)
GACAAAGGCUGGGAACAGGCAGGGCACGAAUGGAGCUAUCAACAGAAAAUGGUAAAGUGAGAAUGAAAGGAGAGAGAGAGAGAGAGAUCUGUUGUGGCUACGCUCAUGCACUGCCUCUUCCCUGGCUCUGUC
Structure 1 Folding bases 1 To 132 of 10Sep29-18-43-09     InitialdG = -45.36

        10          20         30         40         50         60
    AA    -     C      G   C  A  G     U       AAAUGGUAAAGUGAGAAUGA
GACA  GGCU GGGAA AGGCAG GCA GA UG AGCUA CAACAGA                    \
||||  |||| ||||| |||||| ||| || || ||||| |||||||                    |
CUGU  UCGG CCCUU UCCGUC CGU CU GC UCGGU GUUGUCU                    A
    C-    U     C      A   A  C  A     -       AGAGAGAGAGAGAGAGAGGA
130        120        110        100         90         80        70


gma-MIR408c         AUGCACUGCCUCUUCCCUGGC
gma-MIR408c(iso)    UGCACUGCCUCUUCCCUGGC
Gm10:36557005-36557130, 126nt, (-)
GACAAAGCAGGGGAACAGGCAGAGCAUGGAUGGAGCUAUCAACACAAUAUUGUCAAGAAAGUGAGAAAGUGAGAGGAGAAAUCUGUUGUGGUUCUGCUCAUGCACUGCCUCUUCCCUGGCUUUGUC

Structure 1 Folding bases 1 To 126 of 10Sep29-18-44-58     Initial dG = -59.60

        10         20         30         40          50
        A      C       A     GA        U    .-A| AUA  G   AG
GACAAAGC GGGGAA AGGCAG GCAUG  UGGAGCUA CAAC   CA   UU UCA  A
|||||||| |||||| |||||| |||||  |||||||| ||||   ||   || |||
CUGUUUCG UCCCUU UCCGUC CGUAC  GUCUUGGU GUUG   GU   AA AGU  A
        G      C      A     UC        -    \ -^ GA-  G   GA
    120        110        100        90         70         60


                                        AGA   G
                                           GGA A
                                           |||
                                           UCU A
                                        ---   A
                                              80
```

gma-MIR2218a-5p GGAGAUGGGAGGGUCGGUAAAG
gma-MIR2218a-5p(iso) GGAGAUGGGAGGGUCGGUAA
gma-MIR2218a-3p UUGCCGAUUCCACCCAUUCCUA
Gm10:48574017-48574137, 121nt, (-)

GAGCUUGAGGAAGUGAUGGGAGAUGGGAGGGUCGGUAAAGGAUAACAGCGUCUCUAUGAUUAAUUGUUGUGUUGUUUAUUCUUUUGCCGAUUCCACCCAUUCCUAUGAUUUUCUUUGGUUC

Structure 1 Folding bases 1 To 121 of 10Sep29-18-46-18 Initial dG = -52.00

```
         10         20         30         40          50
      U         G      G     A  -            U--|      UCUCU  GA
GAGCU GAGGAAGU AUGGGA AUGGG GG GUCGGUAAAGGA   AACAGCG     AU  \
||||| |||||||| |||||| ||||| || ||||||||||||   |||||||     ||
CUUGG UUCUUUUA UAUCCU UACCC CC UAGCCGUUUUCU   UUGUUGU     UA  U
      U         G      -     A  U             UAU^      GUUGU  AU
          110        100         90         80         70
```

gma-MIR2218b-5p GGAGAUGGGAGGGUCGGUAA
gma-MIR2218b-3p UUGCCGAUUCCACCCAUUCCUA
Gm20:35349741-35349881, 141nt, (+)

GAGCUUGAGGAAGUGAUGGGAGAUGGGAGGGUCGGUAAAGAAUAUAUCUGAGACUCGACUCAAUCUCGAUCUCUCUCAGUGUUGUGUUGUUUUGUUUAUCCUUUUGCCGAUUCCACCCAUUCCUAUGAUUUCCUUCG
GUUC

Structure 2 Folding bases 1 To 141 of 10Sep29-18-46-46 Initial dG = -53.80

```
        10         20         30         40         50         60          70
     U         G      G     A  -          AAU|  UC      UC   U     -  C  UC
GAGCU GAGGAAGU AUGGGA AUGGG GG GUCGGUAAAG   AUA  UGAGAC  GAC CAAU CU GA  U
||||| |||||||| |||||| ||||| || ||||||||||   |||  ||||||  ||| |||| || ||
CUUGG UUCCUUUA UAUCCU UACCC CC UAGCCGUUUU   UAU  GUUUUG  UUG GUUG GA CU  C
     C         G      -     A  U           CC-^  UU      --   U    U  -  CU
  .       130        120        110        100         90          80
```

```
gma-MIR3522a          AGACCAAAUGAGCAGCUGA
gma-MIR3522a(iso1)   GAGACCAAAUGAGCAGCUGA
gma-MIR3522a(iso2)  UGAGACCAAAUGAGCAGCUGA
gma-MIR3522a(iso3)  UCGUCCUGAGACCAAAUGAGC
Gm15:4318787-4318873, 87nt, (+)
AGGAUCGUCCUGAGACCAAAUGAGCAGCUGACCACAUGAUGCAGCUAUGUUUGCUAUUCAGCUGCUCAUCUGUUCUCAGGUCGCCCU

Structure 1 Folding bases 1 To 87 of 10Sep29-18-49-35      Initial dG = -40.20

        10         20         30         40
   AU  U       C  A            CCACAUGAU   G
AGG  CG CCUGAGA CA AUGAGCAGCUGA          GCA C
|||  || ||||||| || ||||||||||||          |||
UCC  GC GGACUCU GU UACUCGUCGACU          UGU U
   C-  U       U  C            UAUCGUU--   A
       80         70         60          50
```

Identification of conserved soybean miRNAs

```
gma-MIR156d  UUGACAGAAGAUAGAGAGCAC
Gm08:3891365-3891489, 125nt, (+)
AAGGUUGUUGACAGAAGAUAGAGAGCACAGAUGAUGAUAUGCAUAUUAUAUAAUAUAUAGCAGGGAACUCAUGAUGAAUUGUGCAUCUUACUCCUUUGUGCUCUCUAUACUUCUGUCAUCACCUU
Structure 1 Folding bases 1 To 125 of 10Sep29-10-05-37 Initial dG = -50.60
        10         20         30         40          50
    U  U         -              U  -   U       .-U|    A
AAGG UG UGACAGAAG AUAGAGAGCACAGA GA UGA AUGCAUA   UAUAU \
|||| || ||||||||| |||||||||||||| || ||| |||||||   |||||
UUCC AC ACUGUCUUC UAUCUCUCGUGUUU CU AUU UACGUGU   AUAUA A
    -  U         A              C  C   C       \ -^    U
    120       110       100        90        80

                                          60
                                           GC--------  G
                                                     AG G
                                                     ||
                                                     UC A
                                           UAAGUAGUAC  A
                                                    70
```

gma-MIR156g(iso) UUGACAGAAGAUAGAGAGCAC
Gm19:8895390-8895493, 104nt, (+)
GAGAAGUACAUUGACAGAAGAUAGAGAGCACAAAACUGCUCACACACAAAAGCUUUUUUGGGUAUGAUCACCUGUGCUCUCUAUCUUCUGUCAACAAACUGUUC

Structure 1 Folding bases 1 To 104 of 10Sep29-10-13-50 Initial dG = -45.30

```
        10         20         30         40         50
|  A   ACA                      AAAC  C   C  A     G
GAG AGU   UUGACAGAAGAUAGAGAGCACA    UG UCA AC CAAAA C
||| |||   ||||||||||||||||||||||    || ||| || |||||
CUU UCA   AACUGUCUUCUAUCUCUCGUGU    AC AGU UG GUUUU U
^  G   AAC                      CC--  U   A  G     U
  100         90         80           70         60
```

gma-MIR156h UUGACAGAAGAUAGAGAGCAC
Gm02:7812526-7812628, 103nt, (+)
GGUGAUGCUGUUGACAGAAGAUAGAGAGCACAGAUGAUGAAAUGCAAGAAAGGAAAUGGCAUCUUACUCUUUUGUGCUCUCUAGUCUUCUUGUCAUCAUCAUU
Structure 1 Folding bases 1 To 103 of 10Sep29-09-32-40 Initial dG = -41.80

```
        10          20         30          40         50
|      CUGU     -      -            U  -   A    AAGAAA
GGUGAUG    UGACA GAAGAU AGAGAGCACAGA GA UGA AUGC      \
|||||||    ||||| |||||| |||||||||||| || ||| ||||
UUACUAC    ACUGU CUUCUG UCUCUCGUGUUU CU AUU UACG      G
^      U---     U      A           U  C   C    GUAAAG
 100           90         80         70         60
```

gma-MIR156i UUGACAGAAGAUAGAGAGCAC
Gm05:38621690-38621813, 124nt, (+)
AAGGUUGUUGACAGAAGAUAGAGAGCACAGAUGAUGAUAUGCAUAUUAUAUAAAAAGCAGCUAGGGAACUCAUGAAUUGUGCAUCUCACUCCUUUGUGCUCUCUAUACUUCUGUCAUCACCCUU

Structure 1 Folding bases 1 To 124 of 10Sep29-09-41-56 Initial dG = -53.40

```
        10          20         30          40         50         60
    U  U          -               U  -|  U       UUAUAUAAAAAG  GCU  G
AAGG UG UGACAGAAG AUAGAGAGCACAGA GA UGA AUGCAUA             CA   AG G
|||| || ||||||||| |||||||||||||| || ||| |||||||             ||   ||
UUCC AC ACUGUCUUC UAUCUCUCGUGUUU CU ACU UACGUGU             GU   UC A
    C  U          A               C  C^  C       UAA---------  AC-  A
  120        110        100         90         80                70
```

gma-MIR156j UUGACAGAAGAGAGUGAGCAC

Gm06:4699149-699240, 92nt, (+)

GAAAUUGACAGAAGAGAGUGAGCACACAGAGGCACUUGGUAUAGUAUAUAUACUAUUGCUUUUGCGUGCUCACUUCUCUUUCUGUCAACUUC

Structure 1 Folding bases 1 To 92 of 10Sep29-09-45-47 Initial dG = -51.00

```
        10          20         30         40
   A       -     -|         A        CU       GU
GAA UUGACAG AAGAG AGUGAGCAC CAGAGGCA  UGGUAUA  \
||| ||||||| ||||| ||||||||| ||||||||  |||||||
CUU AACUGUC UUCUC UCACUCGUG GUUUUCGU  AUCAUAU  A
   C       U     U^         C        U-       AU
 90         80         70         60         50
```

gma-MIR156k UUGACAGAAGAUAGAGAGCAC

Gm07:9347139-9347259, 121nt, (+)

GGUAAGGUUGUUGACAGAAGAUAGAGAGCACAGAUGAUGAUAUGCACAUAUACAUGGAACAGGAAUUUAAGCAAUUGCAUCUCACUCCUUUGUGCUCUCUAAGCUUCUGUCAUCCACCUCC

Structure 1 Folding bases 1 To 121 of 10Sep29-09-47-04 Initial dG = -45.60

```
        10         20          30         40         50         60
| UA    UGU          A-              U  -   U     CAUAUACAUGGAAC
GG  AGGU   UGACAGAAG  UAGAGAGCACAGA GA UGA AUGCA              \
||  ||||   |||||||||  |||||||||||| || ||| |||||
CC  UCCA   ACUGUCUUC  AUCUCUCGUGUUU CU ACU UACGU              A
^ --    CCU          GA              C  C   C     UAACGAAUUUAAGG
 .         110        100         90         80         70
```

gma-MIR156l UUGACAGAAGAUAGAGAGCAC
Gm09:37843750-37843864, 115nt, (-)
AGGGUUGUUGACAGAAGAUAGAGAGCACAGAUAGUGAUAUGCAUAAAAAUAUGGAACGGGAAAGCAAUUGCAUCUCACUCCUUUGUGCUCUCUAGGCUUCUGUCAUCCACACCUU

Structure 1 Folding bases 1 To 115 of 10Sep29-10-08-02 Initial dG = -49.30

```
          10          20         30          40         50
     -  U-          A-             U-|    U     UAAAAAUAUGGA
AGGGU UG  UGACAGAAG  UAGAGAGCACAGA  AGUGA AUGCA            \
||||| ||  |||||||||  ||||||||||||||  ||||| |||||
UUCCA AC  ACUGUCUUC  AUCUCUCGUGUUU  UCACU UACGU            A
     C  CU          GG             CC^    C     UAACGAAAGGGC
   110        100         90         80         70         60
```

gma-MIR156m UUGACAGAAGAGAGUGAGCAC
Gm14:10664512-10664600, 89nt, (-)
AAAUUGACAGAAGAGAGUGAGCACAAAGAGGCACUUGAUAUAAAUUCUAUAUCACUGCUUUUGUGUGCUCACCACUCUUUCUGUCGGUUU

Structure 1 Folding bases 1 To 89 of 10Sep29-10-09-55 Initial dG = -44.90

```
         10          20         30         40
           -     A-|        A       CU       A
AAAUUGACAG AAGAG  GUGAGCACA AGAGGCA  UGAUAUA U
|||||||||| |||||  ||||||||| |||||||  |||||||
UUUGGCUGUC UUCUC  CACUCGUGU UUUUCGU  ACUAUAU U
          U     AC^        G       C-       C
         80         70         60         50
```

gma-MIR156n UUGACAGAAGAGAGUGAGCAC
Gm17:37759446-37759535, 90nt, (+)
AAAUUGACAGAAGAGAGUGAGCACAAAGAGGCACUUGAUAUAAAUCUAUAUCACUGCUUUUGUGUGCUCACUACUCUUUCUGUCGGUUUU

Structure 1 Folding bases 1 To 90 of 10Sep29-10-11-28 Initial dG = -47.80

```
10          20         30         40
-           -     -|          A       CU       A
 AAAUUGACAG AAGAG AGUGAGCACA AGAGGCA  UGAUAUA A
 |||||||||| ||||| |||||||||| |||||||  |||||||
 UUUGGCUGUC UUCUC UCACUCGUGU UUUUCGU  ACUAUAU U
U           U     A^         G       C-       C
.         80         70         60         50
```

gma-MIR159a-5p	GAGCUCCUUGAAGUCCAAUUG	Gm09	37672410	37672586
gma-MIR159a-5p(iso)	GAGCUCCUUGAAGUCCAAUU	Gm09	37672410	37672586
gma-MIR159a-3p	UUUGGAUUGAAGGGAGCUCUA	Gm09	37672410	37672586
gma-MIR159a-3p(iso1)	UUUGGAUUGAAGGGAGCUCU	Gm09	37672410	37672586
gma-MIR159a-3p(iso2)	UUGGAUUGAAGGGAGCUCUA	Gm09	37672410	37672586
gma-MIR159a-3p(iso3)	UGGAUUGAAGGGAGCUCUA	Gm09	37672410	37672586

Gm09:37672410-37672586, 177nt, (+)

GUGGAGCUCCUUGAAGUCCAAUUGAGGAUCUUACUGGGUGAAUUGAGCUGCUUAGCUAUGGAUCCCACAGUUCUACCCAUCAAUAAGUGCUUUUGUGGUAGUCUUGUGGCUUCCAUAUCUGGGGAGCUUCA
UUUGCCUUUAUAGUAUUAACCUUCUUUGGAUUGAAGGGAGCUCUAC

Structure 1 Folding bases 1 To 177 of 10Sep29-10-28-08 Initial dG = -85.40

```
        10         20         30             40         50         60         70         80
            GA        UU     AUCU     ----    UG    U     G      C      UC       UU      -| UCAAUA
GUGGAGCUCCUU  AGUCCAA  GAGG     UACU     GGG   AAU GAGCU CUUAG UAUGGA  CCACAG   CUACC CA       A
||||||||||||  |||||||  ||||     ||||     |||   ||| ||||| ||||| ||||||  ||||||   ||||| ||
CAUCUCGAGGGA  UUAGGUU  UUCC     AUGA     UCC   UUA UUCGA GGGUC AUACCU  GGUGUU   GAUGG GU       G
            AG        UC     AAUU     UAUU    GU    C      G      U       UC       CU      U^ UUUCGU
     170        160        150        140        130        120        110        100         90
```

gma-MIR159b-5p	GAGUUCCCUGCACUCCAAGUC	Gm07	5386107	5386292
gma-MIR159b-3p	AUUGGAGUGAAGGGAGCUCCA	Gm07	5386107	5386292

Gm07:5386107-5386292, 186nt,(-)

AAACCCAACUUGGAGUUCCCUGCACUCCAAGUCUGAAAGGAUAUGAUGGUAAACCUCUACUGCUAGUUCAUGGAUACCUCUGACUUCUUAACAACAUGCGUUCGAAGUCAAGGGUUUGCAUGCCCUGGGAG
AUGAGUUUACCUUGAUCUUUUGGUAUUGGAGUGAAGGGAGCUCCAGAGGGUAUUC

```
Structure 1 Folding bases 1 To 186 of 10Sep29-10-15-16 Initial dG = -83.80

          10         20         30         40         50         60         70         80         90
AA--|   AAC             G         GU  G        AUGAU        C  UA  G    UU    GAU   UC        UU   AAC
    ACCC   UUGGAGUUCCCU CACUCCAA  CU AAAGGAU     GGUAAAC UC  CU CUAG  CAUG   ACC  UGACUUC  AAC    \
    ||||   |||||||||||| ||||||||  || |||||||     ||||||| ||  || ||||  ||||   |||  |||||||  |||
    UGGG   GACCUCGAGGGA GUGAGGUU  GG UUUUCUA     CCAUUUG AG  GA GGUC  GUAC   UGG  ACUGAAG  UUG   A
CUUA^   A--             A        AU  -       GUU--        -  UA  G    CC    GUU   GA       C-   CGU
    180          170        160         150         140         130        120        110         100


gma-MIR159e-5p            GAGCUCCUUGAAGUCCAAUU      Gm07   9524917      9525127
gma-MIR159e-3p            UUUGGAUUGAAGGGAGCUCUA     Gm07   9524917      9525127
gma-MIR159e-3p(iso1)      UUUGGAUUGAAGGGAGCUCU      Gm07   9524917      9525127
gma-MIR159e-3p(iso2)      UUUGGAUUGAAGGGAGCUC       Gm07   9524917      9525127
gma-MIR159e-3p(iso3)       UUGGAUUGAAGGGAGCUCUA     Gm07   9524917      9525127
gma-MIR159e-3p(iso4)        UGGAUUGAAGGGAGCUCUA     Gm07   9524917      9525127

Gm07:9524917-9525127, 211nt, (-)

CAAAGGGGGUUAUGGAGUGGAGCUCCUUGAAGUCCAAUUGAGGAUCUUACUGGGUGGAUUGAGCUGCUUAGCUAUGGAUCCCACAGUUCUACCCAUCAUUAAGUGCUUUUGUGGUAGUCUUGUGGCUUCCA
UAUCUGGGGAGCUUCAUUUGCCUUUAUAGUAUUAUCCUUCUUUGGAUUGAAGGGAGCUCUACACCCUUCUCCCCUUUUGU

Structure 1 Folding bases 1 To 211 of 10Sep29-10-19-50 Initial dG = -108.00

         10         20         30         40         50             60         70         80         90         100
-         UUAU-  A            GA        UU       CU     ----   UG   U     G     C      UC       UU     -| UCAUUA
 CAAAGGGGG     GG GUGGAGCUCCUU  AGUCCAA  GAGGAU  UACU    GGG  GAU GAGCU CUUAG UAUGGA  CCACAG  CUACC CA      A
 |||||||||     || ||||||||||||  |||||||  ||||||  ||||    |||  ||| ||||| ||||| ||||||  ||||||  ||||| ||
 GUUUUCCCC     CC CAUCUCGAGGGA  UUAGGUU  UUCCUA  AUGA    UCC  UUA UUCGA GGGUC AUACCU  GGUGUU  GAUGG GU      G
U         UCUUC  A            AG        UC      UU    UAUU   GU   C     G     U      UC       CU     U^ UUUCGU
 .       200        190        180        170        160        150        140        130        120        110
```

gma-MIR159f-5p GAGUUCCCUGCACUCCAAGUC
gma-MIR159f-3p AUUGGAGUGAAGGGAGCUCCA
Gm16:2794128-2794307, 180nt, (-)
ACCCAAGUUGGAGUUCCCUGCACUCCAAGUCUGAAAGGAUAUGAUGGUAAACCUCUGCUGCUAGUUCAUGGAUACCUCUGGCCUCGUAACAACAUGCGUUCGAAGUCAAGGGUUUGCAUGACCUGGGAGAU
GAGUUUACCUUGACCUUUUGGUAUUGGAGUGAAGGGAGCUCCAGAGGGU

Structure 1 Folding bases 1 To 180 of 10Sep29-10-34-39 Initial dG = -85.60

```
        10         20         30         40         50         60         70         80         90
|   AAG            G         GU UG     AUAUGAU       C  UG  G    U     GAU   UC    C   U   AAC
ACCC   UUGGAGUUCCCU CACUCCAA  C  AAAGG       GGUAAAC UC  CU CUAG UCAUG   ACC  UGGC UCG AAC   \
||||   ||||||||||||| ||||||||  |  |||||       ||||||| ||  || |||| |||||   |||  |||| ||| |||
UGGG   GACCUCGAGGGA GUGAGGUU  G  UUUCC       CCAUUUG AG  GA GGUC AGUAC   UGG  ACUG AGC UUG   A
^   A--            A         AU GU     AGUU---        -  UA  G    C     GUU   GA    A   -   CGU
.         170        160        150          140        130        120        110        100
```

gma-MIR162a UCGAUAAACCUCUGCAUCCAG
Gm06:20176238-20176339, 102nt,(-)
GUGAAGUCACUGGAUGCAGCGGUUCAUCGAUCUCUUCCUGAAUCGUUGUUUAACAUCAGAACCAUGAAUCGAUCGAUAAACCUCUGCAUCCAGCGCUCACUC

Structure 1 Folding bases 1 To 102 of 10Sep29-10-38-42 Initial dG = -43.20

```
          10         20         30         40         50
--|   A  CA          C    C       UC    C  AA  GU-  UU
  GUGA GU  CUGGAUGCAG GGUU AUCGAUC  UUC UG  UC   UG  \
  |||| ||  |||||||||| |||| |||||||  ||| ||  ||   ||
  CACU CG  GACCUACGUC CCAA UAGCUAG  AAG AC  AG   AC  U
CU^   -  C-          U    A       CU   U  CA  ACU  AA
100          90         80         70         60
```

gma-MIR162b UCGAUAAACCUCUGCAUCCAG
Gm05:7692594-7692698, 105nt, (-)
GGUGAAGUCACUGGAGGCAGCGGUUCAUCGAUCUCUUCCUGAAUUUGGUUGUGGAAGAACACAAAGCAAGAAUCGGUCGAUAAACCUCUGCAUCCAGCGCUCACU

Structure 1 Folding bases 1 To 105 of 10Sep29-10-36-20 Initial dG = -41.80

```
   10         20         30         40         50
|    A  CA      G     C     C        UC    C  AAUUUGG      GA
GGUGA GU  CUGGA GCAG GGUU AUCGAUC  UUC UG        UUGUG  A
|||||  ||  |||||  ||||  ||||  |||||||  |||  ||        |||||
UCACU CG  GACCU CGUC CCAA UAGCUGG  AAG AC        AACAC  G
^     -  C-      A    U    A        CU    A  GA-----      AA
    100          90         80         70              60
```

gma-MIR162c UCGAUAAACCUCUGCAUCCAG

Gm17:10181489-10181607, 119nt, (+)

GAGAUGAGGUGAAGUCACUGGAGGCAGCGGUUCAUCGAUCUCUUCCUGAAUUUGGUUGUGGAAGAACACAAAGCAAGAAUCGGUCGAUAAACCUCUGCAUCCAGCGCUCACUUUGCCUC

Structure 1 Folding bases 1 To 119 of 10Sep29-10-40-19 Initial dG = -48.40

```
        10         20         30         40         50         60
|  AU       A  CA      G     C     C        UC    C  AAUUUGG      GA
GAG  GAGGUGA GU  CUGGA GCAG GGUU AUCGAUC  UUC UG        UUGUG  A
|||  ||||||| ||  ||||| ||||  ||||  |||||||  |||  ||        |||||
CUC  UUUCACU CG  GACCU CGUC CCAA UAGCUGG  AAG AC        AACAC  G
^  CG       -  C-      A    U    A        CU    A  GA-----      AA
       110          100         90         80              70
```

gma-MIR166a-5p GGAAUGUUGUCUGGCUCGAGG

gma-MIR166a-3p UCGGACCAGGCUUCAUUCCCC

Gm16:1912570-1912715, 146nt, (-)

ACGGAAGCUUUGUCUUUUGAGGGGAAUGUUGUCUGGCUCGAGGACCCUUCUUCAUCUUGAUCUUGUGUAGACUACUAUGCUUGUGGUCAAGGAAUACAUAGUGUUGUCGGACCAGGCUUCAUUCCCCCCAA
UUAUAUGCUUCCAAA

Structure 1 Folding bases 1 To 146 of 10Sep29-10-59-40 Initial dG = -62.10

```
         10         20         30         40         50         60          70
AC-      UU   CUU   A        UU      CU   G   C  UCUUCA-        UU--|     A
   GGAAGC  UGU   UUG GGGGAAUG  GUCUGG  CGA GAC CU       UCUUGAUC    GUGUAG C
   ||||||  |||   ||| ||||||||  ||||||  ||| ||| ||       ||||||||    ||||||
   CCUUCG  AUA   AAC CCCCUUAC  CGGACC  GCU UUG GA       GGAACUGG    CGUAUC U
AAA      U-   UU-   C        UU      AG   G   U  UACAUAA        UGUU^     A
    140         130        120        110        100         90         80
```

gma-MIR166c-5p GGAAUGUCGUCUGGUUCGAG
gma-MIR166c-3p UCGGACCAGGCUUCAUUCCCC
Gm02:14340767-14340863, 97nt, (+)
GUUGAGGGGAAUGUCGUCUGGUUCGAGACCAUUCAUGCAAGUAGUCUCAGACAUGACUCUUCUGAGUGAUUUCGGACCAGGCUUCAUUCCCCUCAGC

Structure 1 Folding bases 1 To 97 of 10Sep29-10-43-46 Initial dG = -57.00

```
        10         20         30         40         50
|            UC             C      UGC   U    UCA
GUUGAGGGGAAUG  GUCUGGUUCGAGA CAUUCA   AAG AGUC   G
|||||||||||||  ||||||||||||| ||||||   ||| ||||
CGACUCCCCUUAC  CGGACCAGGCUUU GUGAGU   UUC UCAG   A
^            UU             A      C--   -    UAC
      90         80         70          60
```

gma-MIR166d-5p GGAAUGUUGUUUGGCUCGAGG
gma-MIR166d-3p UCGGACCAGGCUUCAUUCCCG
Gm08:14990547-14990731, 185nt, (+)

GGAUGAUGGGAAUGUUGUUUGGCUCGAGGUAACUGCAUGGUCUUAAUUUUGUUCAUCUUUUGAAGCUUUAAUUUAUUUAUGGGUUUCAAUCUUUUUUGAUCCCUUGAAACAGAAAAAGCUUUAAAGGUUGG
AUUUUGAGGCUAUCCCUUUAUGUGAUCUCGGACCAGGCUUCAUUCCCGUCAACC

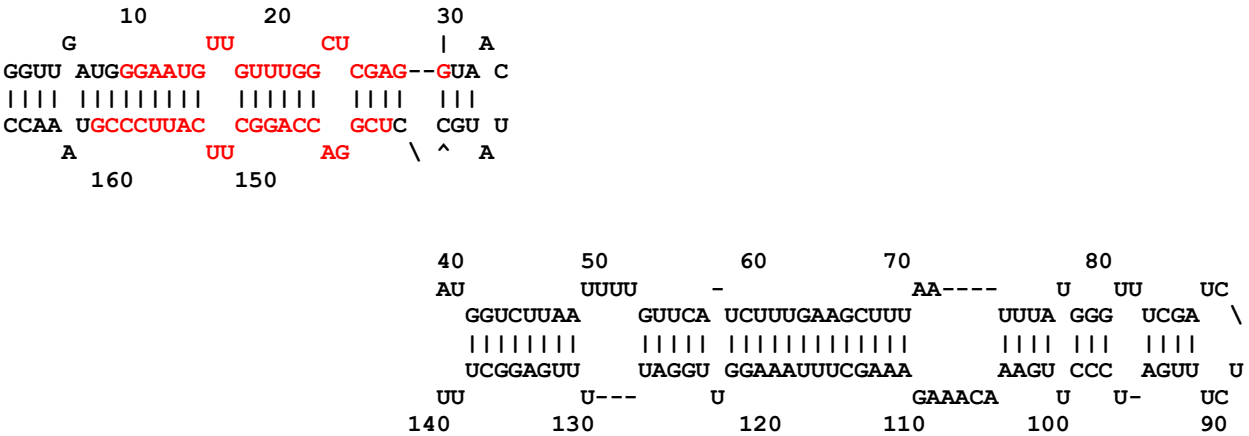
Structure 1 Folding bases 1 To 185 of 10Sep29-10-45-36 Initial dG = -72.60

```
        10         20         30             40         50         60         70         80         90
  A           UU       CU       A  UGC-----|         UUUU     U            AAUUUA    U   UU    UC
GG UGAUGGGAAUG  GUUUGG  CGAGGU AC        AUGGUCUUAA    GUUCA CUUUUGAAGCUUU      UUUA GGG  UCAA  U
|| ||||||||||  ||||||  |||||| ||        ||||||||||    ||||| |||||||||||||      |||| |||  ||||
CC ACUGCCCUUAC  CGGACC  GCUCUA UG        UAUCGGAGUU    UAGGU GGAAAUUUCGAAA      AAGU CCC  AGUU  U
  A           UU       AG       G  UAUUUCCC^         U---     U             AAGACA    U   U-    UU
   180        170        160        150        140          130        120        110        100
```

gma-MIR166e-5p GGAAUGUUGUUUGGCUCGAGG
gma-MIR166e-3p UCGGACCAGGCUUCAUUCCCG
Gm15:3688764-3688931, 168nt, (-)

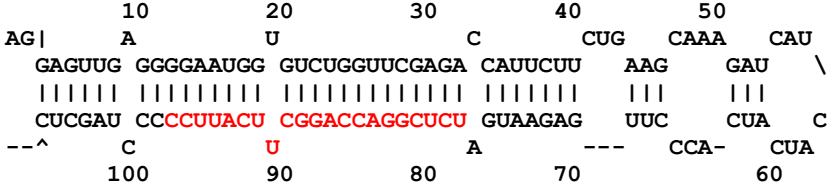
GGUUGAUGGGAAUGUUGUUUGGCUCGAGGUAACUAUGCAUGGUCUUAAUUUUGUUCAUCUUUGAAGCUUUAAUUUAUGGGUUUCGAUCUCUUUGAUCCCUUGAAACAAAGAAAGCUUUAAAGGUUGGAUUU
UGAGGCUUUCUCGGACCAGGCUUCAUUCCCGUAAACC

Structure 1 Folding bases 1 To 168 of 10Sep29-10-51-35 Initial dG = -63.90



gma-MIR166f UCUCGGACCAGGCUUCAUUCC
gma-MIR166f(iso) UCUCGGACCAGGCUUCAUUC
Gm20:43105394-43105500, 107nt, (-)
AGGAGUUGAGGGGAAUGGUGUCUGGUUCGAGACCAUUCUUCUGAAGCAAAGAUCAUCAUCAUCACCCUUGAGAAUGAUCUCGGACCAGGCUUCAUUCCCCCUAGCUC

Structure 1 Folding bases 1 To 107 of 10Sep29-11-01-57 Initial dG = -52.70

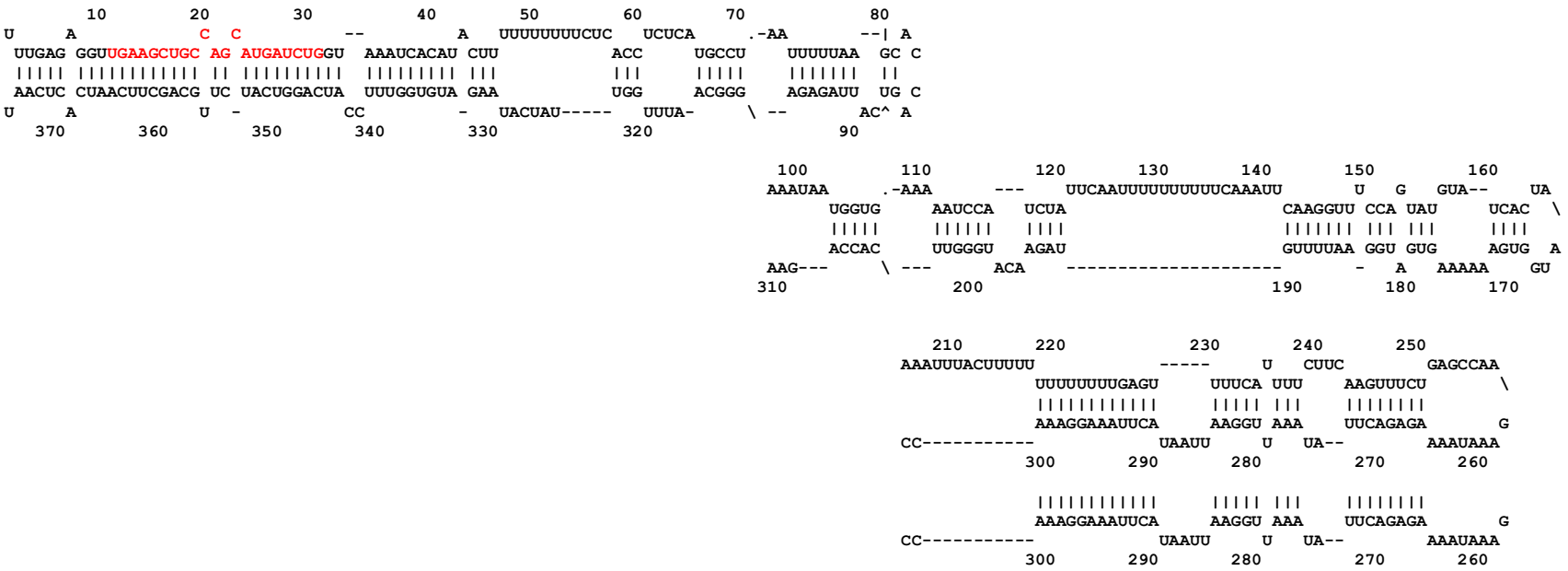


gma-MIR167c UGAAGCUGCCAGCAUGAUCUG

Gm07:39778512-39778886, 375nt, (+)

UUUGAGAGGUUGAAGCUGCCAGCAUGAUCUGGUAAAUCACAUACUUUUUUUUUUCUCACCUCUCAUGCCUAAUUUUUAAGCACCAGUCAUUAGAGAAAAUAAUGGUGAAAAAUCCAUCUAUUCAAUUUUUU
UUUUCAAAUUCAAGGUUUCCAGUAUGUAUCACUAAUGGUGAAAAAAGUGAUGGAAUUUUGUAGAACAUGGGUUAAAUUUACUUUUUUUUUUUUUUGAGUUUUCAUUUUCUUCAAGUUUCUGAGCCAAGAAAUAAAAGA
GACUUAUAAAUUGGAAUUAAUACUUAAAGGAAACCCACCAGAAGGGCAAUUUGGUUAUCAUAAGAUGUGGUUUCCAUCAGGUCAUCUUGCAGCUUCAAUCACUCAAU

Structure 1 Folding bases 1 To 375 of 10Sep29-11-04-02 Initial dG = -97.70



gma-MIR167g(iso) UGAAGCUGCCAGCAUGAUCUG

Gm10:39044877-39044954, 78nt, (+)

CAGCAGUUGAAGCUGCCAGCAUGAUCUGAGUUUACCUUCUAUUGGUAAGAACAGAUCAUGUGGCUGCUUCACCUGUUG

Structure 1 Folding bases 1 To 78 of 10Sep29-11-14-32 Initial dG = -44.90

```
        10         20         30
      U      U    G          AG      UUC
CAGCAG UGAAGC GCCA CAUGAUCUG  UUUACC   \
|||||| |||||| |||| |||||||||  ||||||
GUUGUC ACUUCG CGGU GUACUAGAC  GAAUGG   U
      C      U    -          AA      UUA
       70         60         50
```

gma-MIR167h UGAAGCUGCCAGCAUGAUCUG
Gm20:44765096-44765173, 78nt, (+)
CAGCAGUUGAAGCUGCCAGCAUGAUCUGAGUUUACCUUCUAUUGGUAAGAACAGAUCAUGUGGCUGCUUCACCUGUUG

Structure 1 Folding bases 1 To 78 of 10Sep29-11-16-29 Initial dG = -44.90

```
        10         20         30
|     U      U    G          AG      UUC
CAGCAG UGAAGC GCCA CAUGAUCUG  UUUACC   \
|||||| |||||| |||| |||||||||  ||||||
GUUGUC ACUUCG CGGU GUACUAGAC  GAAUGG   U
^     C      U    -          AA      UUA
       70         60         50
```

gma-MIR168a UCGCUUGGUGCAGGUCGGGAA
Gm09:41353226-41353350, 125nt, (-)
CACUGUGCGGUCUCUAAUUCGCUUGGUGCAGGUCGGGAACCGGUUUUCGCGCGGAAUGGAGGAGCGGUCGCCGGCGCCGAAUUGGAUCCCGCCUUGCAUCAACUGAAUCGGAGGCCGCGGUGAAC

Structure 1 Folding bases 1 To 125 of 10Sep29-11-19-40 Initial dG = -66.20

```
           10        20        30        40        50           60
---|   GU        UA     C          U     A       UC    -  AA--  GA
   CACU  GCGGUCUC  AUUCG UUGGUGCAGG CGGGA CCGGUUU  GCGC GG    UG  G
   ||||  ||||||||  ||||| |||||||||| ||||| |||||||  |||| ||    ||
   GUGG  CGCCGGAG  UAAGU AACUACGUUC GCCCU GGUUAAG  CGCG CC    GC  G
CAA^   --        GC     C          C     A       C-    G  GCUG  GA
   120         110       100        90        80         70
```

gma-MIR168b UCGCUUGGUGCAGGUCGGGAA
Gm01:48070311-48070420, 110nt, (-)
CGGUCUCUAAUUCGCUUGGUGCAGGUCGGGAACCGGUUUUCGCGCGGAAUGGAGGAACGGUCGCCGGCGGCGAAUUGGAUCCCGCCUUGCAUCAACUGAAUCGGAGGCCG

Structure 1 Folding bases 1 To 110 of 10Sep29-11-18-03 Initial dG = -60.40

```
        10         20         30         40          50
       UA     C           U     A        -    G   AA--| GA
CGGUCUC  AUUCG UUGGUGCAGG CGGGA CCGGUUU UCGC CGG    UG  G
|||||||  ||||| |||||||||| ||||| ||||||| |||| |||    ||
GCCGGAG  UAAGU AACUACGUUC GCCCU GGUUAAG GGCG GCC    GC  G
       GC     C           C     A        C    -   GCUG^ AA
        100         90         80         70          60
```

gma-MIR169a CAGCCAAGGAUGACUUGCCGG
Gm09:35771804-35771924, 121nt, (+)
AAGAGGAAGAGAGAGUGAUGCAGCCAAGGAUGACUUGCCGGCGUUAUUAUUUGCUCAUGUUCAUGCUCACCGGUUUCCUUGCCGGCAAGUUGUGUUUGGCUAUGUUUUGCUCUCUUCUUCU

Structure 1 Folding bases 1 To 121 of 10Sep29-12-01-27 Initial dG = -57.20

```
         10         20         30        40         50           60
A            AGU   GC        G UG           UUAUU   --| CU   G
 AGAGGAAGAGAG    GAU  AGCCAAG A  ACUUGCCGGCG     AUU  UG  CAU U
 ||||||||||||    |||  ||||||| |  ||||||||||     |||  ||  |||
 UCUUCUUCUCUC    UUG  UCGGUUU U  UGAACGGCCGU     UGG  AC  GUA U
-             GUU   UA       G GU           UCCUU   CC^ UC   C
120        110        100        90        80         70
```

gma-MIR169f-5p UAGCCAAGAAUGACUUGCCGG
gma-MIR169f-3p UUUCGACGAGUUGUUCUUGGC
Gm02:46876643-46876727, 85nt, (-)
GUAGCCAAGAAUGACUUGCCGGAAUGCAUGCAUUUAUUAGGUACCAAGGUGUAUUGUAUGAUUUCGACGAGUUGUUCUUGGCUAC

```
Structure 1 Folding bases 1 To 85 of 10Sep29-11-50-35       Initial dG = -39.10

         10         20         30         40
|           UG      C       G       UU-     AGGU
GUAGCCAAGAA  ACUUG CGGAAU CAUGCA    UAUU     \
||||||||||||  ||||| |||||| ||||||    ||||
CAUCGGUUCUU  UGAGC GCUUUA GUAUGU    GUGG     A
^           GU      A       -       UAU     AACC
     80         70          60         50
```

```
gma-MIR169g          CAGCCAAGAAUGACUUGCCGG
gma-MIR169g(iso)     AGCCAAGAAUGACUUGCCGG
Gm09:5263992-5264096, 105nt,     (+)
GAGUGAUUUGCAGCCAAGAAUGACUUGCCGGAAUGCUUGCAUUAGGCAUAAUAUAUAGUUGUAUACUUUAUAAUCCGGCAAGUUGUUUUUGGCUACACUUUUUUC
```

```
Structure 1 Folding bases 1 To 105 of 10Sep29-11-54-56 Initial dG = -42.90

         10         20         30           40
   U  UU  C          UG             |       C
GAG GA  UG AGCCAAGAA  ACUUGCCGGA--AUGCUUG \
|||  ||  || |||||||||  ||||||||||  |||||||
CUU UU  AC UCGGUUUUU  UGAACGGCCU  UACGGAU A
   U  UC  A          GU           \ ^      U
   100         90         80

                                  50         60
                                   AA     U   UG
                                     UAUA AGU  \
                                     |||| |||
                                     AUAU UCA  U
                                    A-     U   UA
                                       70
```

gma-MIR169h CAGCCAAGAAUGACUUGCCGG
gma-MIR169h(iso) AGCCAAGAAUGACUUGCCGG
Gm14:5324798-5324911, 114nt, (+)
GAGUGAUUUGCAGCCAAGAAUGACUUGCCGGAAUGCAUAUAUAUGCAUUAGGUACCAACAUAUAUAGUUGUAUUGUAUAAUUUCGGGCAAGUUGUUUUUGGCUACAUUUAUCUC

Structure 4 Folding bases 1 To 114 of 10Sep29-11-57-42 Initial dG = -40.60

```
      10        20        30        40         50
   UGAUU  C         UG       G   UGC|    U     -   G   CC
GAG     UG AGCCAAGAA  ACUUGCC GAA   AUAUA AUGCA UUA GUA  A
|||     || |||||||||  ||||||| |||   ||||| ||||| ||| |||
CUC     AC UCGGUUUUU  UGAACGG CUU   UAUGU UAUGU GAU UAU  A
   UAUUU  A         GU       G   UAA^    -     U   A   AC
  110       100        90        80         70        60
```

gma-MIR169i CAGCCAAGGAUGACUUGCCGG
Gm10:40332790-40332926, 137nt, (-)
AGAGAGUGUAGUGCAGCCAAGGAUGACUUGCCGGCAUUAGCCAAGUGAAUGAGCAUCAUAUAUAUAUAUAUAUAUAUAUAUAUAUAUAUGACUCAUGUUCUUGUCGGCAAGUUGGCCUUGGCUAUAUUGGA
CUCUCU

Structure 1 Folding bases 1 To 137 of 10Sep29-12-03-25 Initial dG = -71.40

```
        10        20        30        40        50        60        70
       G     C        AU            UU   CAAGUGA     CA               A
AGAGAGU UAGUG AGCCAAGG  GACUUGCCGGCA  AGC       AUGAG  UCAUAUAUAUAUAUAU U
||||||| ||||| ||||||||  ||||||||||||  |||       |||||  ||||||||||||||||
UCUCUCA GUUAU UCGGUUCC  UUGAACGGCUGU  UUG       UACUC  AGUAUAUAUAUAUAUA A
       G     A        GG            UC   -------     --               U
     130       120       110       100                 90        80
```

gma-MIR169j CAGCCAAGGAUGACUUGCCGG

Gm13:368441-368563, 123nt,(-)
GAAAGUAGAGUGCAGCCAAGGAUGACUUGCCGGAGCAAUAAGCAAUAAUAUCUAUCUAUUUUAUCUUUCUUUUCUUCUUAUUCUUUCACGCCGGCAAGUUGUUCUUGGCUACAUUUUGUUUUC

Structure 1 Folding bases 1 To 123 of 10Sep29-12-06-19 Initial dG = -44.80

```
        10         20         30         40         50         60
            C          UG          AGCAAU   C     UAUCUAUCUAUUUU
GAAAGUAGAGUG AGCCAAGGA  ACUUGCCGG      AAG AAUAA              \
|||||||||||| |||||||||  |||||||||      ||| |||||
CUUUUGUUUUAC UCGGUUCUU  UGAACGGCC      UUC UUAUU              A
            A          GU          GCACU-   -     CUUCUUUUCUUUCU
 120       110        100         90          80         70
```

gma-MIR169k CAGCCAAGGGUGAUUUGCCGG
Gm15:14150069-14150183, 115nt, (+)
GGAGUGCAGCCAAGGGUGAUUUGCCGGCACAGGCACUAAUUAGUUCAAUAUUGAAUAGUUAGUUCUGUUUGAUUGAUUUACUUCUGUGCCGGCAAGUUUCUCUUGGCUACAUUUC

Structure 1 Folding bases 1 To 115 of 10Sep29-12-10-27 Initial dG = -56.80

```
        10         20         30         40         50
      C          U               CACU        C    UU    AG
GGAGUG AGCCAAGGG GAUUUGCCGGCACAGG    AAUUAGUU AAUA  GAAU  \
|||||| ||||||||| ||||||||||||||||    |||||||| ||||  ||||
CUUUAC UCGGUUCUC UUGAACGGCCGUGUCU    UUAGUUAG UUGU  CUUG  U
      A          U               UCAU        U    --    AU
   110        100         90         80         70          60
```

gma-MIR169l CAGCCAAGGAUGACUUGCCGG
Gm17:4861816-4861963, 148nt, (-)
AGAGGUAGAAAGUAGAGUGCAGCCAAGGAUGACUUGCCGGAGCAAUAAGAAUAAACAAUAAAAUGUAUCUAUUUAUAACAUUCUUCUUCUUCUUCUUCUUUCACGCCGGCAAGUUGUUCUUGGCUACA
UUUUGUUUUCUUCUUCU

Structure 1 Folding bases 1 To 148 of 10Sep29-12-14-23 InitialdG = -52.80

```
 10        20        30        40              50        60        70
     U             C         UG         ------| C  U     UAAACAAUAA     AUCUA
AGAGG AGAAAGUAGAGUG AGCCAAGGA  ACUUGCCGG      AG AA AAGAA          AAUGU     \
|||||  |||||||||||||  |||||||||  |||||||||      || ||  |||||          |||||
UCUUC UCUUUUGUUUUAC UCGGUUCUU  UGAACGGCC      UC UU UUCUU          UUACA     U
     U              A          GU          GCACUU^ -  C     CUUCUUCUUC     AUAUU
      140       130       120       110        100        90        80
```

gma-MIR171d UUGAGCCGUGCCAAUAUCACG

Gm06:48920631-48920715, 85nt, (-)

CGGGAUAUUGGUCCGGUUCAAUAAGAAAGCAAUGCUCAAAAUGUUAUUGGGUCCUGUUUUUUCAUUGAGCCGUGCCAAUAUCACG

Structure 1 Folding bases 1 To 85 of 10Sep29-12-17-50 Initial dG = -39.80

```
        10        20        30        40
| G        UC         A        AU      AAU
CG GAUAUUGG  CGGUUCAAU AGAAAGCA  GCUCAA   \
||  ||||||||  |||||||||  ||||||||  ||||||
GC CUAUAACC  GCCGAGUUA UUUUUUGU  UGGGUU   G
^ A        GU         C        CC      AUU
    80        70        60        50
```

gma-MIR171e CGAUGUUGGUGAGGUUCAAUC

Gm13:26271135-26271232, 98nt, (+)

GAGAAAGCGAUGUUGGUGAGGUUCAAUCCGAAGACGGAUUUACAUGUAGAAGCAGUAAAAUACGAUCUCAGAUUGAGCCGCGCCAAUAUCACUUUAUC

Structure 1 Folding bases 1 To 98 of 10Sep29-12-19-16 Initial dG = -41.20

```
        10        20        30          40
  G    C          A         C A   -  GA-|    A   A
GA AAAG GAUGUUGGUG GGUUCAAUC G AGA CG   UUUAC UGU G
||  |||| ||||||||||  |||||||||  | ||| ||   |||||  |||
CU UUUC CUAUAACCGC CCGAGUUAG C UCU GC   AAAUG ACG A
  A    A          G         A -   A  AUA^    -   A
       90        80        70         60         50
```

gma-MIR171f CGAUGUUGGUGAGGUUCAAUC
Gm17:9101701-9101798, 98nt, (-)
GAGAAAGCGAUGUUGGUGAGGUUCAAUCCGAAGACGGAUUUACAUGUAGAAGCAGUAAAAAUACGAUCUCAGAUUGAGCCGCGCCAAUAUCACUUUAUC

Structure 1 Folding bases 1 to 98 of 10Oct10-16-34-43 Initial dG = -41.20

```
        10         20         30          40
  G    C           A          C A   -  GA-|    A   A
GA AAAG GAUGUUGGUG GGUUCAAUC G AGA CG   UUUAC UGU G
|| |||| |||||||||| ||||||||| | ||| ||   ||||| |||
CU UUUC CUAUAACCGC CCGAGUUAG C UCU GC   AAAUG ACG A
  A    A           G          A -   A  AUA^    -   A
       90         80         70          60         50
```

gma-MIR172b-5p GUAGCAUCAUCAAGAUUCAC
Gm13:40401688-40401809, 122nt, (-)
GCGGAUGUAGCAUCAUCAAGAUUCACAUGCAAAUGAAGGUGGGUGGGACUAUGAUGCAAUCCAAGUGCUCUGCCAAUCCAUCGGUCUUUUUGAUGUGAGAAUCUUGAUGAUGCUGCAUCAGC
Structure 1 Folding bases 1 To 122 of 10Sep29-12-45-16 Initial dG = -57.40

```
        10         20          30          40         50         60
  G                    A    -    U-  A-|      G  ACUAU  U   AUC
GC GAUGUAGCAUCAUCAAGAUUC CAUG CAAA  GA  GGUGGGU GG     GA GCA   \
|| |||||||||||||||||||| |||| ||||  ||  ||||||| ||     || |||
CG CUACGUCGUAGUAGUUCUAAG GUGU GUUU  CU  CUACCUA CC     CU CGU   C
  A                    A    A    UU  GG^      A  GU---  -   GAA
120       110        100        90         80           70
```

gma-MIR172c GGAAUCUUGAUGAUGCUGCAG
gma-MIR172c(iso) GAAUCUUGAUGAUGCUGCAG
Gm18:2968986-2969138, 153nt, (+)
AAAUCAGUCACUGUUUGCCGGUGGAGCAUCAUCAAGAUUCACAAGCUUUAGGGGCAUUAAUUUGUUUGAGGUGGUUCCUUAUUGCUCCAAAACCAAUUAGCCCUUUUGCUAUGGGAAUCUUGAUGAUGCUGCAGCAG
CAAUAAAUGACUAAUA

Structure 1 Folding bases 1 To 153 of 10Sep29-12-47-32 Initial dG = -61.40

```
        10         20         30         40          50         60          70
AAAUC     CUGU     CGG  G                A  -|  UUU      AU     UG-     AG UG  UC
     AGUCA     UUGC    UG AGCAUCAUCAAGAUUC CA AGC   AGGGGC  UAAUU   UUUG  G  GU  \
     |||||     ||||    || |||||||||||||||| || |||   ||||||  |||||   ||||  |  ||
     UCAGU     AACG    AC UCGUAGUAGUUCUAAG GU UCG   UUCCCG  AUUAA   AAAC  C  UA  C
AUAA-      AAAU     ACG  G                 G  A^  UU-      --     CCA    CU GU  UU
 150         140        130        120        110         100          90         80
```

gma-MIR172d(iso1) GGAAUCUUGAUGAUGCUGCAG
gma-MIR172d(iso2) GAAUCUUGAUGAUGCUGCAG

Gm14:5548752 - 5548901 150nt (+)

AAAACAGUCGCUGAUUGCAGAUGGAGCAUCAUCAAGAUUCACAAGCUUCAGGGGUUUUUUGUUUGGGGUGGUCCCUUAUUGCUCCCAAAUGAAUUAAGCCCUUUGAUAUGGGAAUCUUGAUGAUGCUGCAGCAGCAA
UAAACGACUAACA

Structure 1 Folding bases 1 To 150 of 10Sep29-12-46-31 InitialdG = -66.90

```
        10         20         30         40          50           60          70
AAAAC     CUG     AGA  G                A  AGCU  -         --  UG      -|UG  CC
     AGUCG    AUUGC    UG AGCAUCAUCAAGAUUC CA    UC AGGGGUUU  UU  UUUGGG G  GU  \
     |||||    |||||    || |||||||||||||||| ||    || ||||||||  ||  |||||| |  ||
     UCAGC    UAACG    AC UCGUAGUAGUUCUAAG GU    AG UUCCCGAA  AA  AAACCC C  UA  C
ACAA-      AAA      ACG  G                 G  AU--  U        UU  GU      U^GU  UU
.         140        130        120        110          100         90         80
```

gma-MIR172e(iso1) GGAAUCUUGAUGAUGCUGCAG
gma-MIR172e(iso2) GAAUCUUGAUGAUGCUGCAG

Gm11:35957808-35957960, 153nt, (-)

AAAACAGUCACUGUUUGCCGGUGGAGCAUCAUCAAGAUUCACAAGCUUUAGGGGGCAUUAAUUUGUUUGAGGUGGUCCCUUAUUGAUCCAAACCAAUUAGCCCUUUUGCUAUGGGAAUCUUGAUGAUGCUGCAGCAG
CAAUAAAUGACUAAUA

Structure 1 Folding bases 1 To 153 of 10Sep29-12-40-39 Initial dG = -58.30

```
        10         20         30         40          50         60         70         80
AAAAC     CUGU    CGG  G                A  -|  UUU       AU     U     AGGU    CCU
     AGUCA    UUGC   UG AGCAUCAUCAAGAUUC CA AGC   AGGGGGC  UAAUU GUUUG    GGUC   \
     |||||    ||||   || |||||||||||||||| || |||   |||||||  ||||| |||||    ||||
     UCAGU    AACG   AC UCGUAGUAGUUCUAAG GU UCG   UUUCCCG  AUUAA CAAAC    CUAG   U
AUAA-     AAAU    ACG  G                 G  A^  U--       --     C     ----    UUA
 150        140       130       120       110         100          90
```

gma-MIR172g GCAGCACCAUCAAGAUUCAC
Gm10:31592576-31592689,114nt, (-)
GCAGGUGCAGCACCAUCAAGAUUCACAUGCAAUUUACACCCUAUAAGAGAUUUUUGACAAAAUUUCUUUCCUAGGAGAGUUUUGACUUGAGAAUCUUGAUGAUGCUGCAUCAGC

Structure 1 Folding bases 1 To 114 of 10Sep29-12-23-59 InitialdG = -45.20

```
        10         20         30         40          50
  A         C           A  UG-   UUUACAC    U--|         U
GC GGUGCAGCA CAUCAAGAUUC CA   CAA       CCUA   AAGAGAUUUU G
|| |||||||||| |||||||||| ||   |||       ||||   ||||||||||
CG CUACGUCGU GUAGUUCUAAG GU   GUU       GGAU   UUCUUUAAAA A
  A         A           A  UCA   UUGAGA-    CCU^         C
  110       100        90        80         70        60
```

gma-MIR172h-5p GCAGCAGCAUCAAGAUUCACA
gma-MIR172h-5p(iso) GCAGCAGCAUCAAGAUUCAC
gma-MIR172h-3p AGAAUCUUGAUGAUGCUGCAU
Gm10:43474733-43474823, 91nt,(+)
GCAGGUGCAGCAGCAUCAAGAUUCACACACAGAUUUACCUCCUUGGGGGCGUGUGUUUCGGUGCUGAGAAUCUUGAUGAUGCUGCAUCAGC
Structure 1 Folding bases 1 To 91 of 10Sep29-12-27-57 InitialdG = -44.10

```
        10         20         30          40
  A         G           A  -   A  UU--   -|   U
GC GGUGCAGCA CAUCAAGAUUC CA CAC GA    UAC CUCC \
|| |||||||||| |||||||||| || ||| ||    ||| |||| |
CG CUACGUCGU GUAGUUCUAAG GU GUG CU    GUG GGGG U
  A         A           A  C   G  UUGU   C^   G
90        80        70        60        50
```

```
gma-MIR172i          GCAGCAGCAUCAAGAUUCACA
gma-MIR172i(iso)     GCAGCAGCAUCAAGAUUCAC
Gm15:2892962-2893122, 161nt, (-)
GCAGGUGCAGCAGCAUCAAGAUUCACACCGCCUAAUUUGCUAGGACUUCAGGACUGCACACGCUUAAUUAUACAUACAUAUAUACAUAUAUGUUAGCUCUUUGUGGAGUGCGGAAUAAAGUUCUAUUUUAG
AUGUGGGAAUCUUGAUGAUGCUGCAUCAGC
Structure 1 Folding bases 1 To 161 of 10Sep29-12-49-36     Initial dG = -60.10

        10         20         30         40         50         60         70         80
| A          G            A    CGC     UUUGC         CAGGA       A-    UUAAUUAUACAU        U
GC GGUGCAGCA CAUCAAGAUUC CAC    CUAA      UAGGACUU      CUGCAC  CGC             ACAUAUA A
|| ||||||||| |||||||||||| |||    ||||      ||||||||      ||||||  |||             |||||||
CG CUACGUCGU GUAGUUCUAAG GUG    GAUU      AUCUUGAA      GGCGUG  GUG             UGUAUAU C
^ A          A            G    UA-     UU---          AUAA-       AG    UUUCUCGAU---        A
 .        150        140         130           120         110        100            90


gma-MIR172j          GCAGCAGCAUCAAGAUUCACA
gma-MIR172j(iso)     GCAGCAGCAUCAAGAUUCAC
Gm20:40895747-40895836,   90nt, (-)
GCAGGUGCAGCAGCAUCAAGAUUCACACACAGAUUCCACCUCCUUGGGGGAGUGUUUAGGUGCUGAGAAUCUUGAUGAUGCUGCAUCAGC
Structure 1 Folding bases 1 To 90 of 10Sep29-12-51-11      Initial dG = -44.50

        10         20          30         40
  A          G            A  -|  AGAUUC    CU  U
GC GGUGCAGCA CAUCAAGAUUC CA CAC       CAC   CC U
|| ||||||||| |||||||||||| || |||       |||   ||
CG CUACGUCGU GUAGUUCUAAG GU GUG       GUG   GG G
  A          A            A  C^  GAUUU-    AG   G
.         80         70         60          50


gma-MIR319a(iso)     UUGGACUGAAGGGAGCUCCCU
Gm05:40832097-40832279, 183nt, (+)
AAGGUAAGAGAGCUUUCUUCAGUCCACUCAUGGGUGACAGUAAGAUUCAAUUAGCUGCCGACUCAUUCAUCCAAAUGUUGAGUGUAAGCGAAUAAAUAUACUCAGCAGAUGAGUGAAUGAUGCGGGAGACA
AAUUGAAUCUUAAGUUUCCUGUACUUGGACUGAAGGGAGCUCCCUUUUCCUU

Structure 1 Folding bases 1 To 183 of 10Sep29-13-02-53     InitialdG = -88.00
```

```
          10         20         30         40         50          60         70         80         90
     U-|  A      UU           CUC      U   AG             AG-  G   AC        UC  AA           AGCG
AAGG  AAG GAGCU  CUUCAGUCCA   AUGGG GAC  UAAGAUUCAAUU   CU CCG  UCAUUCA  CA  UGUUGAGUGUA    A
||||  ||| |||||  ||||||||||   ||||| |||  ||||||||||||   || |||  |||||||  ||  ||||||||||
UUCC  UUC CUCGA  GAAGUCAGGU   UGUCC UUG  AUUCUAAGUUAA   GA GGC  AGUAAGU  GU  ACGACUCAUAU    A
     UU^  C      GG           UCA      U   A-             ACA  G   GU       GA  AG           AAAU
 180        170        160        150         140        130        120        110        100
```

gma-MIR319b(iso) UUGGACUGAAGGGAGCUCCCU

Gm08:1647811-1647990, 180nt, (-)

GGUAAGAGAGCUUUCUUCAGUCCACUUAUGGGUGACAAUAAGAUUUCAAUUAGCUGCCGACUCAUUCAUCCAAAUGCUGAGUGAAAGCGAAGAAAGAUACUCAGCAAAUGAGUGAAUGAUGCGGGAGACAA
AUUGAUUCUUAAGUUUCCUGUACUUGGACUGAAGGGAGCUCCCUUUUUC

Structure 1 Folding bases 1 To 180 of 10Sep29-13-05-47 InitialdG = -78.50

```
          10         20         30         40         50          60         70         80         90
GGU-|  A      UU           CU       U   AA      UU       AG-  G   AC        UC  AA          AAAGCG
    AAG GAGCU  CUUCAGUCCA  UAUGGG GAC  UAAGA  UCAAUU   CU CCG  UCAUUCA  CA  UGCUGAGUG       A
    ||| |||||  ||||||||||  ||||||| |||  |||||  ||||||   || |||  |||||||  ||  |||||||||
    UUC CUCGA  GAAGUCAGGU  AUGUCC UUG  AUUCU  AGUUAA   GA GGC  AGUAAGU  GU  ACGACUCAU       A
CUUU^  C      GG           UC       U   A-      U-       ACA  G   GU        GA  AA          AGAAAG
.        170        160        150         140         130        120        110        100
```

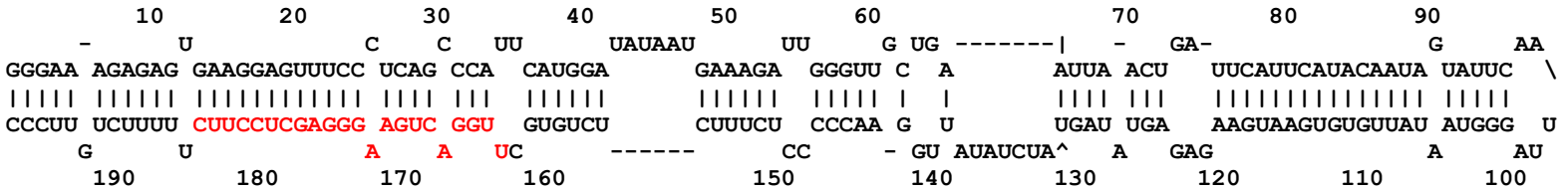
gma-MIR319d UUGGACUGAAGGGAGCUCCUUC

gma-MIR319d(iso) GGACUGAAGGGAGCUCCUUC

Gm02:43885398-43885595, 198nt, (+)

GGGAAAGAGAGUGAAGGAGUUUCCCUCAGCCCAUUCAUGGAUAUAAUGAAAGAUUGGGUUGCUGAAUUAACUGAUUCAUUCAUACAAUAGUAUUCAAUUAGGGUAAUAUUGUGUGAAUGAAGAGAGUAUAGUAUCUA
UAUUGGAACCCCCUCUUUCUCUGUGCUUGGACUGAAGGGAGCUCCUUCUUUUUCUGUUCCC

Structure 2 Folding bases 1 To 198 of 10Sep29-12-55-08 Initial dG = -83.80

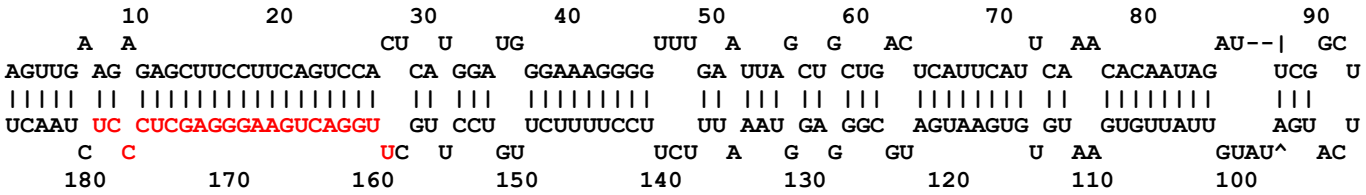


gma-MIR319e UUGGACUGAAGGGAGCUCCCU

Gm02:45704227-45704412, 186nt, (+)

AGUUGAAGAGAGCUUCCUUCAGUCCACUCAUGGAUGGGAAAGGGGUUUGAAUUAGCUGCUGACUCAUUCAUUCAAACACAAUAGAUUCGGCUUCAUGAUAUGUUAUUGUGAAUGUGUGAAUGAUGCGGGAGGUAAAU
UUCUUCCUUUUCUUGUCCUUGCUUGGACUGAAGGGAGCUCCCUCUAACU

Structure 1 Folding bases 1 To 186 of 10Sep29-12-58-00 Initial dG = -87.20



gma-MIR319f UUGGACUGAAGGGGAGCUCCUUC

gma-MIR319f(iso) UGGACUGAAGGGGAGCUCCUUC

Gm04:46348798-46348991, 194nt, (+)

AAGAGAGUGAAGGAGCUUCCUUCAGCCCACGCAUGGGUUCGGGGGAUUGAAGGGUUGCUGAAGCAUCUGCUGACUCAUUCAUACACACAGAAUAGUGAUUCAAAUUGCUAAGGUAAUUGUGUGAAUGAAGC
AGGAGAUAUUUUGCAUCCCUCUUUCUUUGUGCUUGGACUGAAGGGGAGCUCCUUCUUUCUGUU

Structure 2 Folding bases 1 To 194 of 10Sep29-12-59-20 Initial dG = -84.40

```
        10         20         30         40         50         60         70         80         90        100
  G    U          -         C   C   UG  UUC GG   UUGAA   U   U   GC    G   AC           CACAGAA   UG| U
AA AGAG GAAGGAGCU UCCUUCAG CCA GCA  GG   G  GGA     GGG UGC GAA  AUCU CUG  UCAUUCAUACA       UAG  AU C
|| |||| ||||||||| |||||||| ||| |||  ||   |  |||     ||| ||| |||  |||| |||  ||||||||||       |||  ||
UU UCUU CUUCCUCGA GGGAAGUC GGU CGU  UU   C  UCU     CCC ACG UUU  UAGA GAC  AGUAAGUGUGU       AUC  UA A
  G    U          G        A   U   GU  --- UU   -----   U   U   A-    G   GA           UAAUGGA   GU^ A
  190        180        170        160               150        140        130        120        110
```

gma-MIR319g UUGGACUGAAGGGAGCUCCCU

Gm11:1374020-1374198, 179nt, (+)

AGGUAAGAGAGCUCUCUUCAGCCCACUCAUAGGUGAUAAUAGGAUUUAAUUAGCUGCCGACUCAUUCAUACACAUGCUGAGUGAAUUAAUGAAUAAUACUCAGUAAAUGAGUGAAUGAUACGGGAGACAAAUUGAAU
CUUAUGUUUUCUGUACUUGGACUGAAGGGAGCUCCCUUUUCU

Structure 1 Folding bases 1 To 179 of 10Sep29-13-10-14 Initial dG = -74.20

```
        10         20         30         40         50         60         70         80         90
|  U   A             C   CUC     U    A           AG-  G   AC        A  CA        GA    AU
AGG AAG GAGCUCUCUUCAG CCA   AUAGG GAUA UAGGAUUUAAUU   CU CCG  UCAUUCAU CA  UGCUGAGU  AUUA  \
||| ||| |||||||||||||| |||   ||||| |||| ||||||||||||   || |||  |||||||| ||  ||||||||  ||||
UCU UUC CUCGAGGGAAGUC GGU   UGUCU UUGU AUUCUAAGUUAA   GA GGC  AGUAAGUG GU  AUGACUCA  UAAU  G
^  U   C              A   UCA     U    -            ACA  G   AU        A  AA        --    AA
       170        160        150        140        130        120        110        100
```

gma-MIR319h UUGGACUGAAGGGAGCUCCCU

Gm11:32902062-32902247, 186nt, (+)

AGUUUAAGAGAGCUUCCUUCAGUCCACUCAUGGAUGGGUAGAGGGUUUGAAUUAGCUGCUGACUCAUUCAUUCAAACACAAUAGAAUUAGUAUCAUGGUAUGCUAUUGUGAAUGUGUGAAUGAUGCAGGAG
GUAAAGUUCAUCCUUUUCUUGUCUUUGCUUGGACUGAAGGGAGCUCCCUUUUACU

Structure 1 Folding bases 1 To 186 of 10Sep29-14-00-51 Initial dG = -83.30

```
        10         20         30         40         50          60         70         80         90
|  UU   A                 CU  U   UG  U      UU     --  G  G   AC        U  AA        AAUUA    C
AGU  AAG GAGCUUCCUUCAGUCCA  CA GGA  GG AGAGGG  UGAAU  UA CU CUG  UCAUUCAU CA  CACAAUAG     GUAU A
|||  ||| ||||||||||||||||  || |||  || ||||||  |||||  || || |||  |||||||| ||  ||||||||     ||||
UCA  UUC CUCGAGGGAAGUCAGGU  GU UCU  UC UUUUCC  ACUUG  AU GA GAC  AGUAAGUG GU  GUGUUAUC     UAUG U
^  UU   C                 UC  U   GU  -      U-     AA  G  G   GU        U  AA        G----    G
    180        170        160        150         140        130        120        110           100
```

gma-MIR319i UUGGACUGAAGGGAGCUCCCU

Gm14:47959350-47959535, 186nt, (+)

AGUUGAAGAGAGCUUCCUUCAGUCCACUCAUGGAUGGGUAAGGGGUUAGAAUUAGCUGCUGACUCAUUCAUUCAAACACAAUAGAUCUGGCAUCAUGAUAUGCUAUUGUGAAUGUGUGAAUGAUGCGGGAG
GUAAAUUUCUUCCUUUUCCUGUCUGUGCUUGGACUGAAGGGAGCUCCCUCUAACU

Structure 1 Folding bases 1 To 186 of 10Sep29-14-03-37 Initial dG = -87.50

```
        10        20        30        40        50        60        70        80            90
     A  A                 CU          U      UU   A   G  G   AC        U  AA        ---   -| G
AGUUG AG GAGCUUCCUUCAGUCCA  CAUGGAUGGG AAGGGG  AGA UUA CU CUG  UCAUUCAU CA  CACAAUAG   AUC UG C
|||||  ||  ||||||||||||||||  ||||||||||  ||||||  ||| ||| ||  |||  ||||||||  ||  ||||||||   ||| ||
UCAAU UC CUCGAGGGAAGUCAGGU  GUGUCUGUCC UUUCCU  UUU AAU GA GGC  AGUAAGUG GU  GUGUUAUC   UAG AC A
     C  C                 UC          U      UC   A   G  G   GU        U  AA        GUA   U^ U
    180       170       160       150       140       130       120       110       100
```

gma-MIR319j UUGGACUGAAGGGAGCUCCUUC

gma-MIR319j(iso) GGACUGAAGGGAGCUCCUUC

Gm14:45953433-45953649, 217nt, (-)

GGAGGGAAAGAGAGUGAAGGAGCUUCCCUCAGCCCAUUCAUGGAGAUAACGAAAGAUUGGGUUGCUGAAUUAACUGCUAGCUCACACAUUCAUUCAUACAAUAGUAUUCAAUUAGGGUAAUAUUGUGUGAA
UGAAGCGGGAGUAUAUAGUAUCUAUAUUGGAACCCUCUUUCUCUGUGCUUGGACUGAAGGGAGCUCCUUCUUUUUCUGCUCCCCUCC

Structure 1 Folding bases 1 To 217 of 10Sep29-14-05-19 Initial dG = -101.20

```
          10        20        30        40        50        60        70            80        90       100       110
       A-|     U             C    C   UU        UAACGA    UU     G UG AU  AC     ---    ACACA               G     AA
GGAGGGA  AGAGAG GAAGGAGCUUCC UCAG CCA  CAUGGAGA      AAGA  GGGUU C  A  UA  UGCUA   GCUC     UUCAUUCAUACAAUA UAUUC  \
|||||||  ||||||  ||||||||||||  ||||  |||  ||||||||      ||||  |||||  |  |  ||  |||||   ||||     ||||||||||||||| |||||
CCUCCCU  UCUUUU CUUCCUCGAGGG AGUC GGU  GUGUCUCU      UUCU  CCCAA G  U  AU  AUGAU   UGAG     AAGUAAGUGUGUUAU AUGGG  U
       CG^     U             A    A   UC          ------    --     - GU AU  CU     AUA    GGCG-                A     AU
     210       200       190       180             170          160       150       140        130       120
```

gma-MIR319k UUGGACUGAAGGGAGCUCCUUC

gma-MIR319k(iso1) UGGACUGAAGGGAGCUCCUUC

gma-MIR319k(iso2) GGACUGAAGGGAGCUCCUUC

Gm17:9436178-9436279, 102nt, (-)
UGAAAGGUGCAGAAAUAGGAGUUCCCUUGCAGCCCAAAACACCCUGCAUGAACUACUUCAUGUUGUUUUGGACUGAAGGGAGCUCCUUCUUCUUCACUUUCA

Structure 1 Folding bases 1 To 102 of 10Sep29-14-06-49 Initial dG = -51.90

```
        10         20         30         40         50
U|    G  C    AU            G   C        CCCU       C
 GAAAG UG AGAA  AGGAGUUCCCUU CAG CCAAAACA    GCAUGAA U
 ||||| || ||||  ||||||||||||| ||| ||||||||    |||||||
 CUUUC AC UCUU  UCCUCGAGGGAA GUC GGUUUUGU    UGUACUU A
A^    -  U    CU            -   A        ----       C
100         90         80         70             60
```

gma-MIR3191 UUGGACUGAAGGGAGCUCCCU
Gm18:4278883-4279072, 190nt, (-)
AUAGUUUAAGAGAGCUUCCUUCAGUCCACUCAUGGAUGGAUAGAGGGUUUGAAUUAGCUGCUGACUCAUUCAUUCAAACACAAUAGAACGGGUGUCAUGGUAUGCUAUUGUGAAUGCCUGAAUGAUGCAGG
AGGUAAAGUUCAUCCUUUUCUUGUCUGUGCUUGGACUGAAGGGAGCUCCCUUUUACUGU

Structure 1 Folding bases 1 To 190 of 10Sep29-14-08-20 Initial dG = -82.90

```
        10         20         30         40         50          60         70         80         90
|    UU   A                  CU       UG AU      UU     --  G  G   AC       UU  AA        AACGG UGU
AUAGU  AAG GAGCUUCCUUCAGUCCA  CAUGGA  G  AGAGGG  UGAAU  UA CU CUG  UCAUUCA  CA  CACAAUAG     G   C
|||||  ||| ||||||||||||||||  ||||||  |  ||||||  |||||  || || |||  |||||||  ||  ||||||||     |
UGUCA  UUC CUCGAGGGAAGUCAGGU  GUGUCU  U  UUUCCU  ACUUG  AU GA GAC  AGUAAGU  GU  GUGUUAUC     U   A
^    UU   C                  UC       GU CU       --      AA  G  G   GU       CC  AA        GUA-- GGU
.       180        170        160        150         140        130        120        110         100
```

gma-MIR396a-5p UUCCACAGCUUUCUUGAACUG
gma-MIR396a-5p(iso1) UCCACAGCUUUCUUGAACUG
gma-MIR396a-3p UUCAAUAAAGCUGUGGGAAG
gma-MIR396a-3p(iso1) GUUCAAUAAAGCUGUGGGAAG
gma-MIR396a-3p(iso2) GUUCAAUAAAGCUGUGGGA

```
Gm13:26338134-26338273,   140nt,                                                                                                (-)
UCAUGGCUCUCUUUGUAUUCUUCCACAGCUUUCUUGAACUGCAUCCAAAGAGUUCCUUUGCAUGCAUGCCAUGGCACUCUUACUCCCAAAUCUUGUUUUGCGGUUCAAUAAAGCUGUGGGAAGAUACAGAUAGGGUC
AAC

Structure 1 Folding bases 1 To 140 of 10Sep29-14-12-04     Initial dG = -61.90

        10         20         30         40          50         60
UCA       C        UC             C           .-UCCA|     U  UUU   U
   UGGCUCU UUUGUAU  UUCCACAGCUUU UUGAACUGCA       AAGAGU CC   GCA G
   ||||||| |||||||  |||||||||||| ||||||||||       |||||| ||   |||
   ACUGGGA AGACAUA  AGGGUGUCGAAA AACUUGGCGU       UUCUCA GG   CGU C
CA-       U        GA             U           \ ----^     C  UAC   A
.        130        120        110        100        80         70


                                                      90
                                                ACUCC   A
                                                     CAA \
                                                     |||
                                                     GUU U
                                                UUU--   C


gma-MIR396b-5p             UUCCACAGCUUUCUUGAACUU
gma-MIR396b-3p             GCUCAAGAAAGCUGUGGGAGA
gma-MIR396b-3p(iso1)       CUCAAGAAAGCUGUGGGAGA
Gm13:26329931-26330056,   126nt, (+)
CUCAAGUCCUGGUCAUGCUUUUCCACAGCUUUCUUGAACUUCUUAUGCAUCUUAUAUCUCUCCACCUCCAGGAUUUUAAGCCCUAGAAGCUCAAGAAAGCUGUGGGAGAAUAUGGCAAUUCAGGCU

Structure 1 Folding bases 1 To 126 of 10Sep29-14-14-01     Initial dG = -49.30

        10            20         30         40         50         60
CUCA  U    ----|     C                   A     UAU  AUCUUAU   U  CCA
    AG CCUG    GUCAUG UUUUCCACAGCUUUCUUGA CUUCU   GC       AUC CU   C
    || ||||    |||||| |||||||||||||||||||| |||||   ||       ||| ||
    UC GGAC    CGGUAU AGAGGGUGUCGAAAGAACU GAAGA   CG       UAG GA   C
----  -    UUAA^     A                   C     UCC  AAUUU--   -  CCU
         120        110        100         90         80          70
```

```
gma-MIR396c  UUCCACAGCUUUCUUGAACUU
Gm13:43804777-43804893,   117nt, (+)
CAACAAGUCCUGUUAUGCUUUUCCACAGCUUUCUUGAACUUCUUAUGCCUAGUGCAAUUAUUGAUGUGGCAUAGAAGUUUAAGAAAAAUGUGGAAAAACAUGUCAAAUCUAGGACUU

Structure 1 Folding bases 1 To 117 of 10Sep29-14-16-58     Initial dG = -46.80

        10              20        30        40        50
CAAC        U------|   C         GC              U     UAGUG   U
    AAGUCCUG       UAUG UUUUCCACA  UUUCUUGAACUUCU AUGCC     CAA \
    ||||||||       |||| |||||||||  |||||||||||||| |||||     |||
    UUCAGGAU       GUAC AAAAGGUGU  AAAGAAUUUGAAGA UACGG     GUU U
----        CUAAACU^   A         AA               -     UGUA-   A
         110       100        90        80         70         60
```

```
gma-MIR396d-5p            UUCCACAGCUUUCUUGAACUU
gma-MIR396d-5p(iso) UUCCACAGCUUUCUUGAACU
gma-MIR396d-3p(iso1)      GCUCAAGAAAGCUGUGGGAGA
gma-MIR396d-3p(iso2)      CUCAAGAAAGCUGUGGGAGA
Gm17:9053051-9053155, 105nt,     (-)
GUCAUGCUUUUCCACAGCUUUCUUGAACUUCUUAUGCAUCUUAUAUCUCUCCACUUCCAGCAUUUUAAGCCCUAGAAGCUCAAGAAAGCUGUGGGAGAAUAUGGC

Structure 1 Folding bases 1 To 105 of 10Sep29-14-30-13     Initial dG = -41.76

        10        20        30        40        50
|     C                   A     UAU  AUCUUAUAUCUCUCC
GUCAUG UUUUCCACAGCUUUCUUGA CUUCU   GC               \
|||||| |||||||||||||||||| |||||   ||
CGGUAU AGAGGGUGUCGAAAGAACU GAAGA   CG               A
^     A                   C     UCC  AAUUUUACGACCUUC
   100        90        80        70        60
```

gma-MIR396e(iso1) UUCCACAGCUUUCUUGAACUG
gma-MIR396e(iso2) UUCCACAGCUUUCUUGAACU
gma-MIR396e(iso3) UCCACAGCUUUCUUGAACUG
Gm17:35366535-35366668, 134nt, (-)
GGUCUUUUUCGUGAUCUUCCACAGCUUUCUUGAACUGUGUUGUGAGGCUUCUCUCCAAUGAAGGUUUAUACCCUAUGCAAAAGAAAUUCUAUGAGCACAAUUCAAGAUAGCUGUGGAAAAUCACUGAGAUG
AUC
Structure 2 Folding bases 1 To 134 of 10Sep29-14-32-05 Initial dG = -50.50

```
        10         20         30         40         50         60
    U    C     C          U       C      -   AG C    CUCCA   AA|  U
GGUC UUUU GUGAU UUCCACAGCU UCUUGAA UGUGUU GUG  G UUCU     AUG  GGU U
|||| |||| ||||| |||||||||| ||||||| |||||| |||  | ||||     |||  |||
CUAG AGAG CACUA AAGGUGUCGA AGAACUU ACACGA UAU  U AAGA     UAU  CCA A
    U    U     A          U       A      G   CU A    AAACG   C-^  U
  130        120        110        100         90         80         70
```

gma-MIR396f UCCACAGCUUUCUUGAACUG
Gm14:13971419-13971566, 148nt, (+)
GAAUGGUCUUUUUCGUGAUCUUCCACAGCUUUCUUGAACUGUGUUGUGUGAGGUUUCUCCAAGUGAAGGUUUAAGAUCCCUUAUGCAACAUAAAUUCUUUGAGCACAAUUCAAGAUAGCUGUGGAAAAUCA
CUGAGAUGAUCUCGUUC
Structure 2 Folding bases 1 to 148 of Gm14 Initial dG = -56.90

```
          10         20         30         40          50             60
      --    U    C     C          U       C               UU   ----  AAG
GAAU  GGUC UUUU GUGAU UUCCACAGCU UCUUGAA UGUGU--UGUGUGAGG  UCU    CC   \
||||  |||| |||| ||||| |||||||||| ||||||| |||||  |||||||||  |||    ||
CUUG  CUAG AGAG CACUA AAGGUGUCGA AGAACUU ACACG  ACGUAUUCC  AGA    GG   U
    CU    U    U     A          U       A     \          CU   AUUU  AAG
      140        130        120        110               80         70
```

```
 90
A| UAAA
 CA    \
 ||
 GU    U
A^ UUCU
100
```

```
gma-MIR396g                UUCCACAGCUUUCUUGAACUU
gma-MIR396g(iso)           UUCCACAGCUUUCUUGAACU
Gm15:556702 - 556796       95nt   Frame: +1/-1
UGUUAUGCUUUUCCACAGCUUUCUUGAACUUCUUAUGCCUAAUGCAGCUAUUGAUGUGGCAUUGAAGUUUAAGAAAAAUGUGGAAAAACAUGUCA


Structure 1 Folding bases 1 To 95 of 10Sep29-14-20-45       Initial dG = -39.00

         10         20         30         40         50
U|U    C          GC             UU     UAAU   GCU
 G UAUG UUUUCCACA  UUUCUUGAACUUC  AUGCC    GCA   A
 | |||| |||||||||  |||||||||||||  |||||    |||
 C GUAC AAAAGGUGU  AAAGAAUUUGAAG  UACGG    UGU   U
A^U    A          AA             U-     ----   AGU
    90         80         70          60


gma-MIR396h-5p             UUCCACAGCUUUCUUGAACUG
gma-MIR396h-5p(iso1)       UUCCACAGCUUUCUUGAACU
gma-MIR396h-5p(iso2)       UCCACAGCUUUCUUGAACUG
gma-MIR396h-3p             GUUCAAUAAAGCUGUGGGAAG
gma-MIR396h-3p(iso1)       UUCAAUAAAGCUGUGGGAAG
gma-MIR396h-3p(iso2)       GUUCAAUAAAGCUGUGGGA
Gm17:9044850-9044984,                                                                  135nt,                                                                  (+)
UGGCCCUCUUUGUAUUCUUCCACAGCUUUCUUGAACUGCAUCCAUAGAGUUCCUUUGCAUGCAUGCCAAGGCACUCUUGCUCCCACACCUUGUUUUGCGGUUCAAUAAAGCUGUGGGAAGAUACAGAUAGGGUCA


Structure 2 Folding bases 1 To 135 of 10Sep29-14-22-31 Initial dG = -63.10

          10         20         30         40         50         60
U       C        UC             C           UCCAU  A| UCCUUU   UGCA   C
 GGCCCU UUUGUAU  UUCCACAGCUUU UUGAACUGCA     AG GU      GCA    UGC A
 |||||| |||||||  |||||||||||| ||||||||||     || ||      |||    |||
 CUGGGA AGACAUA  AGGGUGUCGAAA AACUUGGCGU     UC CA      CGU    ACG A
A       U        GA             U           UUUGU  -^ CACCCU   UCUC   G
   130        120        110        100         90          80         70
```

```
gma-MIR482a-5p(iso1)       AUUUGUGGGAAUGGGCUGAUUGG
gma-MIR482a-5p(iso2)       GGAAUGGGCUGAUUGGGAAGC
gma-MIR482a-5p(iso3)       GAAUGGGCUGAUUGGGAAGC
 gma-MIR482a-5p(iso4)       AAUGGGCUGAUUGGGAAGC
Gm02:7783819-7783913, 95nt, (+)
UCAGAAUUUGUGGGAAUGGGCUGAUUGGGAAGCAAUGUGUGCUGGUGCAAUGCAUUUAAUUUCUUCCCAAUUCCGCCCAUUCCUAUGAUUUCUGA

Structure 1 Folding bases 1 To 95 of 10Sep29-14-34-28      Initial dG = -49.50

        10         20         30         40
      U UG           U-          C   GU-|   UGGU
UCAGAA U  UGGGAAUGGGC  GAUUGGGAAG AAU   GUGC    \
|||||| |  ||||||||||| |||||||||| |||   ||||
AGUCUU A  AUCCUUACCCG  UUAACCCUUC UUA   UACG    G
      U GU           CC          U   AUU^   UAAC
    90         80         70         60         50



gma-MIR482b-5p                  UAUGGGGGGAUUGGGAAGGAAU
gma-MIR482b-5p(iso1)            UAUGGGGGGAUUGGGAAGGAA
gma-MIR482b-5p(iso2)            UAUGGGGGGAUUGGGAAGGA
gma-MIR482b-5p(iso3)            AUGGGGGGAUUGGGAAGGA
gma-MIR482b-3p                  UCUUCCCUACACCUCCCAUACC
Gm20:35360312-35360406,   95nt, (+)
GGGGGAAGGCAUGGGUAUGGGGGGAUUGGGAAGGAAUAUCCAUAAGCAAAAUAUGCUAUUUCUUCCCUACACCUCCCAUACCACUGUUUUUCCUC

Structure 1 Folding bases 1 To 95 of 10Sep29-14-40-10      InitialdG = -50.40

        10         20         30         40
           UG           AUU-|          UC    AGC
GGGGGAAGGCA  GGUAUGGGGGG    GGGAAGGAAUA  CAUA    \
||||||||||||  ||||||||||||    ||||||||||||  ||||
CUCCUUUUUGU  CCAUACCCUCC    CCCUUCUUUAU  GUAU    A
           CA           ACAU^          C-    AAA
    90         80         70         60         50
```

```
gma-MIR482c  AUUUGUGGGAAUGGGCUGAUUGG
Gm18:61452904-61453003,   100nt, (-)
GCAAUCAGAAUUUGUGGGAAUGGGCUGAUUGGGAAGUAAUGAGAUUGAGCAAUACAUUUAAUUUCUUCCCAAUUCCGCCCAUUCCUAUGAUUUCUGAUGC

Structure 1 Folding bases 1 To 100 of 10Sep29-14-35-41     Initial dG = -48.10

        10         20         30         40         50
|  A      U UG           U-         GUAAU          CAA
GCA UCAGAA U  UGGGAAUGGGC  GAUUGGGAA     GAGAUUGAG    \
||| |||||| |  ||||||||||  |||||||||     |||||||||
CGU AGUCUU A  AUCCUUACCCG  UUAACCCUU     CUUUAAUUU   U
^  -      U GU           CC          -----          ACA
.          90         80         70              60


gma-MIR482d-5p              UAUGGGGGGAUUGGGAAGGAAU
gma-MIR482d-5p(iso1)        UAUGGGGGGAUUGGGAAGGAA
gma-MIR482d-5p(iso2)       UAUGGGGGGAUUGGGAAGGA
gma-MIR482d-3p             UCUUCCCUACACCUCCCAUACC
Gm10:48569629-48569723,   95nt, (-)
GGGGGAAGACAUGGGUAUGGGGGGAUUGGGAAGGAAUAUCCAUAAGCAAAAUAUGUUAUUUCUUCCCUACACCUCCCAUACCACUGUUUUUCCUC

Structure 1 Folding bases 1 To 95 of 10Sep29-14-37-51       Initial dG = -51.10

        10         20         30         40
           UG           AUU-|          UC    AGC
GGGGGAAGACA  GGUAUGGGGGG    GGGAAGGAAUA  CAUA   \
||||||||||||  ||||||||||||    ||||||||||||  ||||
CUCCUUUUUGU  CCAUACCCUCC    CCCUUCUUUAU  GUAU   A
           CA           ACAU^          U-    AAA
    90         80         70         60         50


gma-MIR1507a        UCUCAUUCCAUACAUCGUCUGA
gma-MIR1507a(iso1)  UCUCAUUCCAUACAUCGUCUGAC
gma-MIR1507a(iso2)  UCUCAUUCCAUACAUCGUCUG
gma-MIR1507a(iso3)  UACUCAUUCCUACAUCGUCU
```

gma-MIR1507a(iso4) UCUCAUUCCAUACAUCGUC
gma-MIR1507a(iso5) UCAUUCCAUACAUCGUCUGA
gma-MIR1507a(iso6) CAUUCCAUACAUCGUCUGA
gma-MIR1507a(iso7) CAUUCCAUACAUCGUCUGACGA
Gm13:25849777-25849883, 107nt, (+)
CAGUGUUUGGCAGAGGUGUAUGGAGUGAGAGAAGGGAAAGGGUAUUUUCCGAUUCUGUCGUUACUCUCUUCCCUCUCUCAUUCCAUACAUCGUCUGACGAACGUAUC

Structure 1 Folding bases 1 To 107 of 10Sep29-14-43-01 Initial dG = -54.70

```
         10         20         30         40         50
CAG-      G    -                  A     -      UUUUC    U
    UGUUUG CAGA GGUGUAUGGAGUGAGAGA GGGAA AGGGUA     CGAU \
    |||||| |||| |||||||||||||||||| ||||| ||||||     ||||
    GCAAGC GUCU CUACAUACCUUACUCUCU CCCUU UCUCAU     GCUG C
CUAU      A    G                  -     C      U----    U
     100         90         80          70             60
```

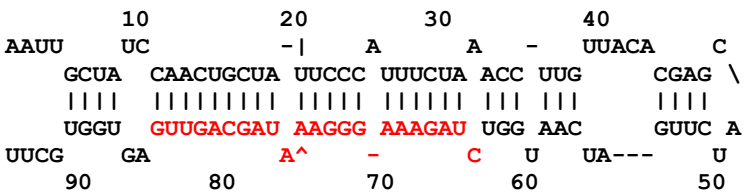
gma-MIR1507b UCUCAUUCCAUACAUCGUCUG
gma-MIR1507b(iso1) UCUCAUUCCAUACAUCGUCUGAC
gma-MIR1507b(iso2) UCUCAUUCCAUACAUCGUCUGA
gma-MIR1507b(iso3) UCUCAUUCCAUACAUCGUCU
gma-MIR1507b(iso4) UCUCAUUCCAUACAUCGUC
gma-MIR1507b(iso5) UCAUUCCAUACAUCGUCUGA
gma-MIR1507b(iso6) CAUUCCAUACAUCGUCUGA
gma-MIR1507b(iso7) CAUUCCAUACAUCGUCUGACGA
Gm17:6190604-6190701, 98nt, (+)
GUUUGACAGAGAUGUAUGGAGUGAGAGAAGGGAAAUGAUAUUUUCCGAUCCCAUCGUUACUCUCUUCCCUCUCUCAUUCCAUACAUCGUCUGACGAAC

Structure 1 Folding bases 1 To 98 of 10Sep29-14-45-20 Initial dG = -47.70

```
    10         20         30         40         50
     A    -|                 A     AU  UAUUUUC   UC
GUUUG CAGA GAUGUAUGGAGUGAGAGA GGGAA  GA       CGA  \
|||||| |||| |||||||||||||||||| |||||  ||       |||
CAAGC GUCU CUACAUACCUUACUCUCU CCCUU  CU       GCU  C
     A    G^                  -     CU  CAUU---   AC
       90         80         70          60
```

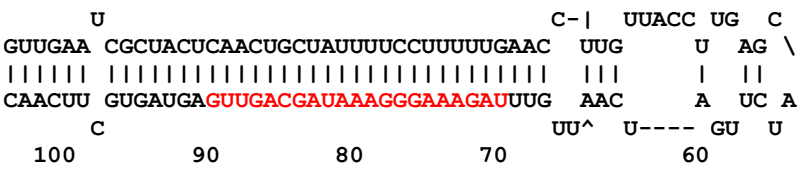
gma-MIR1508a(iso1) CUAGAAAGGGAAAUAGCAGUUG
gma-MIR1508a(iso2) CUAGAAAGGGAAAUAGCAGU
gma-MIR1508a(iso3) UAGAAAGGGAAAUAGCAGUUG
gma-MIR1508a(iso4) AGAAAGGGAAAUAGCAGUUG
gma-MIR1508a(iso5) GAAAGGGAAAUAGCAGUUG
Gm16:32903737-32903831, 95nt, (+)
AAUUGCUAUCCAACUGCUAUUCCCAUUUCUAAACCUUGUUACACGAGCAUCUUGAUCAAUGGUCUAGAAAGGGAAAUAGCAGUUGAGUGGUGCUU

Structure 1 Folding bases 1 To 95 of 10Sep29-16-03-59 Initial dG = -37.10



gma-MIR1508b UAGAAAGGGAAAUAGCAGUUG
gma-MIR1508b(iso1) AGAAAGGGAAAUAGCAGUUG
gma-MIR1508b(iso2) GAAAGGGAAAUAGCAGUUG
Gm09:28530168-28530271, 104nt, (+)
GUUGAAUCGCUACUCAACUGCUAUUUUCCUUUUUGAACCUUGUUACCUUGAGCAUCUUGAUCAAUUGUUUAGAAAGGGAAAUAGCAGUUGAGUAGUGCUUCAAC

Structure 1 Folding bases 1 To 104 of 10Sep29-16-02-07 Initial dG = -52.40



gma-MIR1509a UUAAUCAAGGAAAUCACGGUCG
Gm17:10099759-10099869, 111nt, (+)
CUGCAUCUUCUUAAUCAAGGAAAUCACGGUCGCGUGUGUGCCGGAAAGAAAGUGGCCUGUGAUCUCCGGUUUCUCUUUCUCGACCGUGUUUCCUUGGUUAACGAUAUGUGC
Structure 1 Folding bases 1 To 111 of 10Sep29-16-05-32 Initial dG = -42.00

```
         10          20         30           40         50
CU    CU-  -             U        CGUGUGU---      AAGAA-| U
  GCAU   UC UUAAUCAAGGAAA CACGGUCG           GCCGGA      AG G
  ||||   || |||||||||||||| ||||||||           ||||||      ||
  CGUG   AG AAUUGGUUCCUUU GUGCCAGC           UGGCCU      UC G
--    UAU  C              -         UCUUUCUCUU      CUAGUG^ C
 110        100         90          80         70         60
```

gma-MIR1509b(iso) UUAAUCAAGGAAAUCACGGUUG
Gm05:7774098-7774206, 109nt, (-)
CUGCAUCUUUUUAAUCAAGGAAAUCACGGUUGAGUGUGAAGGAGAGAAAGUGGCUUCAGAUUUCCGGGUUUUCCUUCUCCACUGUGUUUCCUUGGUUAAAGAUAUGUGC

Structure 1 Folding bases 1 To 109 of 10Sep29-16-06-55 Initial dG = -40.40

```
        10         20         30         40         50
CU|   C                U       U  GUGU          AAAG   CUUC
  GCAU UUUUUAAUCAAGGAAA CACGGU GA    GAAGGAGAG    UGG    A
  |||| |||||||||||||||| |||||| ||    |||||||||    |||
  UGUA AGAAAUUGGUUCCUUU GUGUCA CU    CUUCCUUUU    GCC    G
CG^   U                  -      C  ----         GG--   UUUA
       100         90          80             70          60
```

gma-MIR1510a(iso1) UGUUGUUUUACCUAUUCCACCC
gma-MIR1510a(iso2) UGUUGUUUUACCUAUUCCACC
gma-MIR1510a(iso3) UGUUGUUUUACCUAUUCCAC
gma-MIR1510a(iso4) UGUUGUUUUACCUAUUCCA
gma-MIR1510a(iso5) UUGUUUUACCUAUUCCACCCAUU
gma-MIR1510a(iso6) UUGUUUUACCUAUUCCACCCAU
Gm16:31518908-31519000, 93nt, (+)
UUAUGGAACUGGAGGGAUAGGUAAAACAAUGACUGCUGUAUAAGUAAUUGUUAUAGUUAGUUGUUGUUUUACCUAUUCCACCCAUUCCAUGUA

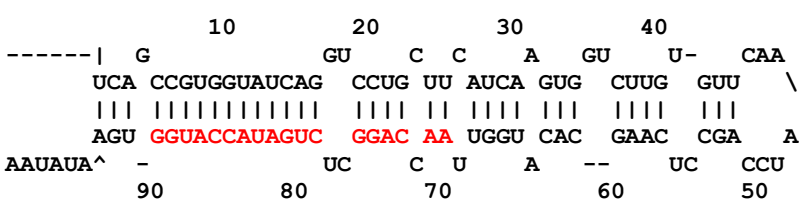
```
Structure 1 Folding bases 1 To 93 of 10Sep29-16-12-02      Initial dG = -46.30

         10        20        30          40
U-       CU  A                UG   -     -|   G
  UAUGGAA  GG GGGAUAGGUAAAACAA  ACU GCUGU AUAA U
  |||||||  || ||||||||||||||||  ||| ||||| ||||
  GUACCUU  CC CCUUAUCCAUUUUGUU  UGA UGAUA UGUU A
AU       AC  A                 GU   U     U^   A
  90        80        70        60        50
```

```
gma-MIR1510b-5p           AGGGAUAGGUAAAACAACUACU
gma-MIR1510b-5p(iso1)     GGGAUAGGUAAAACAACUAC
gma-MIR1510b-5p(iso2)       GAUAGGUAAAACAACUACU
gma-MIR1510b-5p(iso3)     AGGGAUAGGUAAAACAACUAC
gma-MIR1510b-5p(iso4)     AGGGAUAGGUAAAACAACUA
gma-MIR1510b              UGUUGUUUUACCUAUUCCACC
gma-MIR1510b-3p(iso1)     UGUUGUUUUACCUAUUCCACCA
gma-MIR1510b-3p(iso2)     UGUUGUUUUACCUAUUCCAC
gma-MIR1510b-3p(iso3)     UGUUGUUUUACCUAUUCCA
Gm02:6599299-6599392, 94nt, (+)
UUUAUGGAAGUGGAGGGAUAGGUAAAACAACUACUUCUGUAAAAGUAAUUGUUAUAGUUAGUUGUUGUUUUACCUAUUCCACCAAUUCCAUCUA
```

```
Structure 1 Folding bases 1 To 94 of 10Sep29-16-09-45       Initial dG = -43.30

        10        20        30         40
UUU      G   A                U   U-      -| G
   AUGGAA UGG GGGAUAGGUAAAACAAC ACU  CUGUAA AA U
   |||||| ||| |||||||||||||||| |||  |||||| ||
   UACCUU ACC CCUUAUCCAUUUUGUUG UGA  GAUAUU UU A
AUC      A   A                U   UU      G^ A
   90        80        70        60        50
```

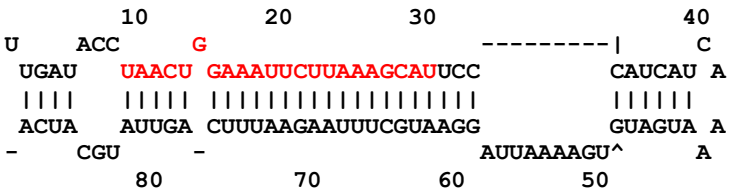
gma-MIR1511(iso) AACCAGGCUCUGAUACCAUGG
Gm18:21161236-21161334, 99nt, (+)
UCAGCCGUGGUAUCAGGUCCUGCUUCAUCAAGUGGUCUUGUGUUCAAAUCCAGCCUCAAGCACAUGGUUAACCAGGCUCUGAUACCAUGGUGAAUAUAA

Structure 1 Folding bases 1 To 99 of 10Sep29-16-13-47 Initial dG = -38.60



gma-MIR1512b UAACUGGAAAUUCUUAAAGCAU
Gm02:8618692-8618781, 89nt,(-)
UUGAUACCUAACUGGAAAUUCUUAAAGCAUUCCCAUCAUCAAAAUGAUGUGAAAAUUAGGAAUGCUUUAAGAAUUUCAGUUAUGCAUCA

Structure 1 Folding bases 1 To 89 of 10Sep29-16-15-43 Initial dG = -35.30



gma-MIR1513a UGAGAGAAAGCCAUGACUUAC
Gm07:43245809-43245901, 93nt,(+)
GGAUCAGAUAUGAGAGAAAGCCAUGACUUACACACACAUUGAAAUCUUAGUUUUAAAUGUGUAUAAGUCAUGGUUUUCUCUCAUAUCUAAUCC

Structure 1 Folding bases 1 To 93 of 10Sep29-16-21-09 Initial dG = -53.30

```
        10         20         30         40
|   C                          C     CA       C
GGAU AGAUAUGAGAGAAAGCCAUGACUUA ACACA  UUGAAAU U
|||| |||||||||||||||||||||||| |||||  |||||||
CCUA UCUAUACUCUCUUUUGGUACUGAAU UGUGU  AAUUUUG U
^   A                           A     A-       A
  90         80         70         60         50
```

gma-MIR1513b UAUGAGAGAAAGCCAUGAC
Gm17:1401433-1401523, 91nt, (-)
GAUCUGAUAUGAGAGAAAGCCAUGACUUACACACGCAUUGAAAUCUAAGUUAAAUAUGUGUGUAAGUCAUGGCAUUAUCUCAUAUCUAAUC

Structure 1 Folding bases 1 To 91 of 10Sep29-16-18-55 Initial dG = -43.80

```
        10         20         30         40
|  CU         G  A                 C   GAAAUC
GAU  GAUAUGAGA AA GCCAUGACUUACACACG AUU      \
|||  ||||||||| || |||||||||||||||| |||
CUA  CUAUACUCU UU CGGUACUGAAUGUGUGU UAA      U
^  AU         A  A                 A   AUUGAA
90         80         70         60         50
```

gma-MIR1513c AAAGCCAUGACUUACACACGC
Gm20:223679-223766, 88nt, (-)
GUUUCUAUGCGUUUGUAAAUCAUGACUUUCUCUUGUAUCUCAUCUGAGUAUAUGAGAGAAAGCCAUGACUUACACACGCAUAUGAAAC
Structure 1 Folding bases 1 To 88 of 10Sep29-16-17-18 Initial dG = -44.70

```
         10         20         30         40
     -       U     A     A        UG   -|  A
GUUUC UAUGCGU UGUAA UCAUG CUUUCUCU  UAU CUC U
||||| ||||||| ||||| ||||| ||||||||  ||| |||
CAAAG AUACGCA ACAUU AGUAC GAAAGAGA  AUA GAG C
     U       C     C     C        GU   U^  U
       80         70         60         50
```

```
gma-MIR1514a(iso)   UUCAUUUUUAAAAUAGGCAUUG
Gm07:43175810-43175908,   99nt, (-)
CUUUGCUAUUUUCAUUUUUAAAAUAGGCAUUGGGGUCCCUUCUUGUUCCUCCUUUUCCUUUCCUAUCCCAAUGCCUAUUUUAAAAUGAAAACAACGAUA

Structure 1 Folding bases 1 To 99 of 10Sep29-16-22-57      Initial dG = -31.80

        10         20         30         40         50
CU|  CUA           U                 UCCCUUCUUGUUCC
  UUG    UUUUCAUUUU AAAAUAGGCAUUGGGG               U
  |||    |||||||||| ||||||||||||||||
  AGC    AAAAGUAAAA UUUUAUCCGUAACCCU               C
AU^  AAC           -                 AUCCUUUCCUUUUC
        90          80         70         60

gma-MIR1523(iso)    UAUGGGAUAAAUGUGAGCUC
Gm02:12253303-12253397,   95nt, (-)
AGGACCCAUUAUGGGAUAAAUGUGAGCUCAGGAGCGAUGAAUAAAUCCUUUCAUCACUCCUCAUUACUCCCGAGCUCACAUUUAUUUCAAAAUGA

Structure 1 Folding bases 1 To 95 of 10Sep29-16-24-01      InitialdG = -42.80

        10         20                   30         40
AGGACC    A  GG              ----------|    C      UAA
      CAUU UG  AUAAAUGUGAGCUC           AGGAG GAUGAA   A
      |||| ||  ||||||||||||||           ||||| ||||||
      GUAA AC  UAUUUACACUCGAG           UCCUC CUACUU   U
A-----     A  UU               CCCUCAUUAC^    A      UCC
         90         80         70         60         50


gma-MIR2109a-5p (iso1)     UGCGAGUGUCUUCGCCUCUGA
gma-MIR2109a-3p            GGAGGCGUAGAUACUCACACCU
gma-MIR2109a-3p (iso1)     GGAGGCGUAGAUACUCACACC
gma-MIR2109a-3p (iso2)     GGAGGCGUAGAUACUCACAC
gma-MIR2109a-3p (iso3)     AGGCGUAGAUACUCACACC
Gm04:28532441-28532537,   97nt,  (-)
AGCCAGUGAAAUCGGUGCGAGUGUCUUCGCCUCUGAGAGAGAUACUAUGAGAUCUCAAGCCUCGGAGGCGUAGAUACUCACACCUCUUUUUCUGGCU

Structure 1 Folding bases 1 To 97 of 10Sep29-16-25-45       Initial dG = -54.50
```

```
        10         20         30           40
      U    UC    C         U          A---|    ACU
AGCCAG GAAA  GGUG GAGUGUCU CGCCUCUGAG    GAGAU   A
|||||| ||||  |||| |||||||| ||||||||||    |||||
UCGGUC UUUU  CCAC CUCAUAGA GCGGAGGCUC    CUCUA   U
      U    CU    A         U          CGAA^    GAG
      90         80         70         60         50
```

gma-MIR4345(iso1) AGACGGAACUUACAAAGAU
gma-MIR4345(iso2) ACUUACAAAGAUUGUUCAGGUAGA
Gm14:49069099-49069193, 95nt, (+)
GGAAGCUAAGACGGAACUUACAAAGAUUGUUCAGGUAGACAUUUGAGAAUAAAUGUCUGUCUUGAACAAUCUUUUUAAGUUUCGUCUUAACUUUC
Structure 1 Folding bases 1 To 95 of 10Sep29-16-27-25 Initial dG = -51.80

```
        10         20         30         40
     C              C             -|         A
GGAAG UAAGACGGAACUUA AAAGAUUGUUCAGG UAGACAUUUG G
||||| |||||||||||||| ||||||||||||| ||||||||||
CUUUC AUUCUGCUUUGAAU UUUCUAACAAGUUC GUCUGUAAAU A
     A              U             U^         A
    90         80         70         60         50
```

gma-MIR4376-3p AGCAUCAUAUCUCCUGCAUAG
Gm13:40845925-40846034, 110nt, (+)
AAGGUUUGCUACGCAGGAGAGAUGACGCUGUCCCUUGCACCCAUCCUAGCUUCCCUUGAGUAGGUAAGAGCAAGGCCAGCCAGCAUCAUAUCUCCUGCAUAGUAAACCUU

Structure 1 Folding bases 1 To 110 of 10Sep29-16-33-31 Initial dG = -52.40

```
        10         20         30           40         50
           C        G    C    UC---|     ACCCAU    G   CC
AAGGUUUGCUA GCAGGAGA AUGA GCUG     CCUUGC      CCUA CUU  \
|||||||||||| |||||||| |||| ||||     ||||||      |||| |||
UUCCAAAUGAU CGUCCUCU UACU CGAC     GGAACG      GGAU GAG  C
           A        A    A    CGACC^     AGAAU-    -   UU
.       100         90         80         70          60
```

gma-MIR4413b UAAGAGAAUUGUAAGUCACU
Gm13:5170460-5170527, 68nt, (+)
CAUCAAUAAGAGAAUUGUAAGUCACUUGAUUAGGAAAUUUUACGGAGACUUACAAUUCCGUAUUGAUG
Structure 1 Folding bases 1 To 68 of 10Sep29-16-35-09 Initial dG = -23.00

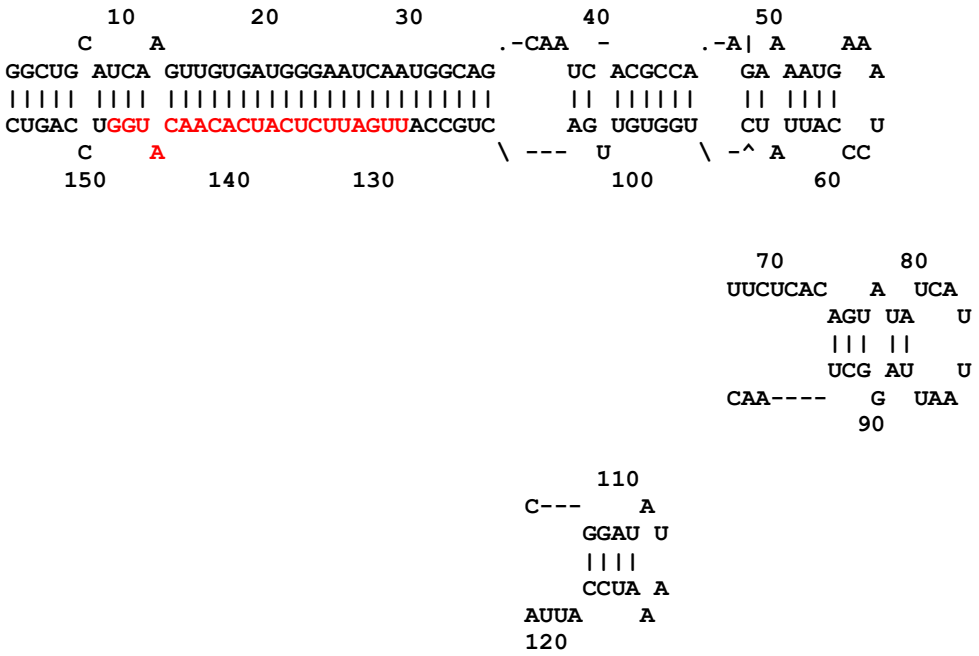
```
        10         20         30
|       AGA            A  -   UUAG
CAUCAAUA   GAAUUGUAAGUC CU UGA    G
||||||||   |||||||||||| || |||
GUAGUUAU   CUUAACAUUCAG GG AUU    A
^       GC-            A  C   UUAA
       60         50        40
```

gma-MIR4413a AAGAGAAUUGUAAGUCACUG
gma-MIR4413a(iso1) UAAGAGAAUUGUAAGUCACUG
gma-MIR4413a(iso2) UAAGAGAAUUGUAAGUCACU
Gm19:1788518-1788620, 103nt, (+)
GCAUCCUCAUCAAUAAGAGAAUUGUAAGUCACUGUAUUAAUUAGGAACUGUUGAUUAGAUGCAUGAUACAGUGACUUACAAUUCUCUUAUUAAUGAUUUGUGC
Structure 1 Folding bases 1 To 103 of 10Sep29-16-36-36 Initial dG = -45.40

```
         10        20        30        40        50
    CC-|   C                           AUUAGGAA   UU
GCAU   UCAU AAUAAGAGAAUUGUAAGUCACUGUAUUA        CUG  \
||||   |||| ||||||||||||||||||||||||||||        |||
CGUG   AGUA UUAUUCUCUUAACAUUCAGUGACAUAGU        GAU  G
    UUU^   A                          ACGUA---   UA
 100        90        80        70        60
```

gma-MIR4415b UUGAUUCUCAUCACAACAUGG
Gm08:23142767-23142922, 156nt, (+)
GGCUGCAUCAAGUUGUGAUGGGAAUCAAUGGCAGCAAUCACGCCAAGAAAAUGAAAUCCCAUUAUCUUCUCACAGUAUAUCAUUAAUUAGGCUAACUGGUGUUGACGGAUAUAAAUCCAUUACUGCCAUUG
AUUCUCAUCACAACAUGGUCCAGUC

Structure 1 Folding bases 1 To 156 of 10Sep29-16-37-59 Initial dG = -63.90

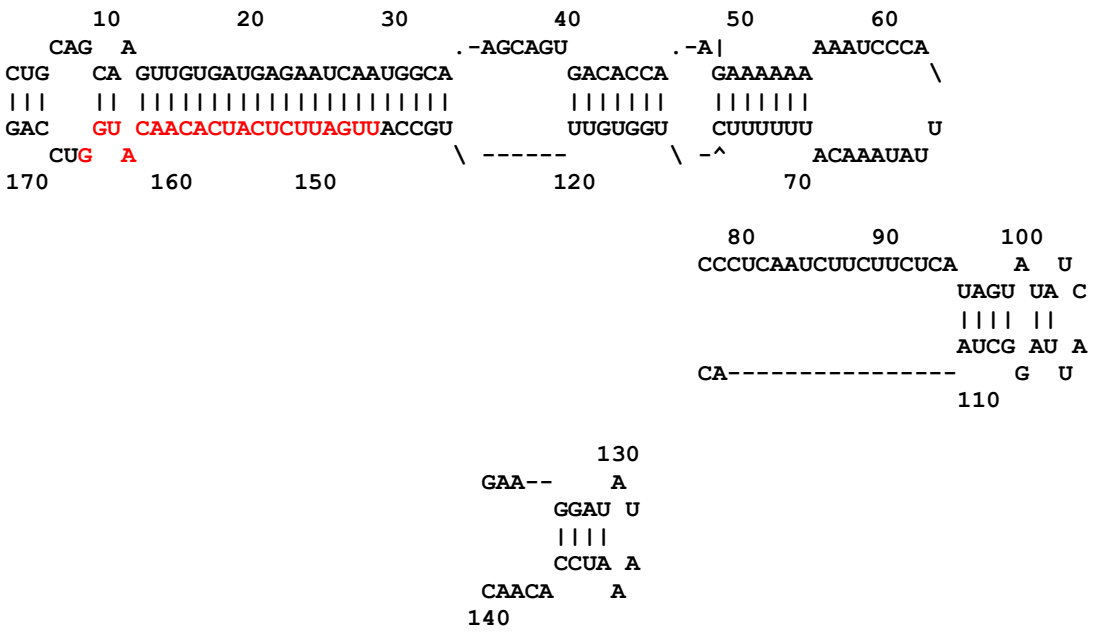


gma-MIR4415a-3p UUGAUUCUCAUCACAACAUGG

Gm18:60474198-60474369, 172nt, (+)

CUGCAGCAAGUUGUGAUGAGAAUCAAUGGCAAGCAGUGACACCAAGAAAAAAAAAAUCCCAUUAUAAACAUUUUUUCCCCUCAAUCUUCUUCUCAUAGUAUAUCAUUAGGCUAACUGGUGUUGAAGGAUAUAAAUCCA
CAACUGCCAUUGAUUCUCAUCACAACAUGGUCCAG

Structure 1 Folding bases 1 To 172 of 10Sep29-16-39-07 Initial dG = -55.90


10 20 30 40 50 60
CAG A .-AGCAGU .-A| AAAUCCCA
CUG CA GUUGUGAUGAGAAUCAAUGGCA GACACCA GAAAAAA \
||| || |||||||||||||||||||||| ||||||| |||||||
GAC GU CAACACUACUCUUAGUUACCGU UUGUGGU CUUUUUU U
CUG A \ ------ \ -^ ACAAAUAU
170 160 150 120 70
80 90 100
CCCUCAAUCUUCUUCUCA A U
UAGU UA C
|||| ||
AUCG AU A
CA---------- G U
110
130
GAA-- A
GGAU U
||||
CCUA A
CAACA A
140

Additional file 2. Predicted *Glycine max* mRNA targets for conserved miRNAs.

miRNA ID	Locus target[a]	Target description[a]	miRNA/mRNA pairing[b]
gma-MIR156d,h,i,k,l	Glyma02g13370	SBP-domain protein	||||||||-||||||||:
	Glyma02g30670	SBP-domain protein	||||||||-||||||||||
	Glyma03g27200	SBP-domain protein	-||||||||-|||||||||
	Glyma03g29900	SBP-domain protein	||||||||-||||||||||
	Glyma04g37390	SBP-domain protein	-||||||||-|||||||||
	Glyma05g00200	SBP-domain protein	-||||||||-|||||||||
	Glyma05g38180	SBP-domain protein	-||||||||-|||||||||
	Glyma06g17700	SBP-domain protein	-||||||||-|||||||||
	Glyma07g31880	SBP-domain protein	-||||||||||||||||||
	Glyma08g01450	SBP-domain protein	-||||||||-|||||||||
	Glyma11g36980	SBP-domain protein	||||||||-||||||||||
	Glyma13g24590	SBP-domain protein	-||||||||||||||||||
	Glyma13g35000	SBP-domain protein	-|||||-||-||||||||||
	Glyma17g08840	SBP-domain protein	-||||||||-|||||||||
	Glyma18g36960	SBP-domain protein	||||||||-||||||||||
	Glyma19g32800	SBP-domain protein	||||||||-||||||||||
gma-MIR156j,m,n	Glyma02g13370	SBP-domain protein	||||||-||||||||||:
	Glyma02g30670	SBP-domain protein	||||||-|||||||||||
	Glyma03g27200	SBP-domain protein	||||||-||||||||||-|
	Glyma03g29900	SBP-domain protein	||||||-|||||||||||
	Glyma04g37390	SBP-domain protein	||||||-||||||||||-|
	Glyma05g00200	SBP-domain protein	||||||-||||||||||-|
	Glyma05g38180	SBP-domain protein	||||||-||||||||||-|
	Glyma06g17700	SBP-domain protein	||||||-||||||||||-|
	Glyma07g31880	SBP-domain protein	-|||||-||-|||||||||
	Glyma08g01450	SBP-domain protein	||||||-||||||||||-|
	Glyma11g36980	SBP-domain protein	||||||-|||||||||||
	Glyma13g24590	SBP-domain protein	-|||||-||-|||||||||
	Glyma13g35000	SBP-domain protein	-|||||-|||||||||||
	Glyma17g08840	SBP-domain protein	||||||-||||||||||-|
	Glyma18g36960	SBP-domain protein	||||||-|||||||||||
	Glyma19g32800	SBP-domain protein	||||||-|||||||||||
gma-MIR159a-3p	Glyma04g10930	plant organelle RNA recognition domain	|:||||||:|||||||::||:
	Glyma06g36770	ankyrin repeat-containig	|-|||:|:||||||||:|||:
gma-MIR159b-3p	Glyma13g04030	MYB-related protein	||||||||||||||||||||
	Glyma13g25720	MYB-related protein	||||||||||||||||||||
	Glyma15g33630	SF3 transcription factor X1-Like	||||||||||||||||||||
	Glyma19g37570	protein tyrosine kinase	-|||||||||||||:|-||||
	Glyma20g11040	MYB-related protein	||||||||||||||||||||
gma-MIR159e-5p	Glyma15g37290	LRR-containing protein	||-||||||-|||||:||||
	Glyma14g40230	hydrolase activity	|:||||:|-|:||||||||:
	Glyma17g37900	hydrolase activity	|:||||:|-|:||||||||:
	Glyma18g01780	hydrolase activity	|:||||||||:||||-|||:

gma-MIR162a,b,c	Glyma07g06590	60S ribosomal protein L5	||-|||||||||||-||||:
gma-MIR166a-3p	Glyma05g06070	Myb-like DNA-binding domain	|||:|||--||||||||||:|
gma-MIR166c-5p	Glyma09g34720	predicted calmodulin-binding protein	|||||||||-||-|::||||
gma-MIR166d-5p	Glyma17g21240	LRR-containing protein	|:||||-|:||||||::||||||
gma-MIR168a,b	Glyma10g30650	Calcium-responsive transcription coactivator	||||||-:||||-||:||||||
	Glyma11g04200	protein tyrosine kinase	||:||||-||||||||||-||
gma-MIR169f-5p	Glyma14g01080	PPR-containig protein	--|||:|||:||||||||||||
gma-MIR169k	Glyma09g07960	transcription factor NF-Y alpha-related	|-|||||||||:||||||||-
	Glyma13g16770	transcription factor NF-Y alpha-related	|:|||||||||:||||||||-
	Glyma15g18970	transcription factor NF-Y alpha-related	|-|||||||||:||||||||-
	Glyma17g05920	transcription factor NF-Y alpha-related	|:|||||||||:||||||||-
	Glyma19g35620	ADP-ribosylation factor GTPase activator	::||:||||:|:|||||||||
gma-MIR170	Glyma08g41980	iron/ascorbate family oxidoreductases	|:||:|||||||:||||||-|
	Glyma18g13610	iron/ascorbate family oxidoreductases	|:||:|||||||:||||||-|
gma-MIR171d	Glyma10g40780	serine/threonine protein kinase	:|||:|||||||::||:|||||
	Glyma11g17490	GRAS family transcription factor	||-|||||||||:|||||||||
	Glyma20g26510	serine/threonine protein kinase	:|||:|||||||::||:|||||
gma-MIR172b-5p	Glyma01g39520	transcription factor activity	|-|||||-||||||||||:|
	Glyma03g33470	transcription factor activity	|-|||||-||||||||||:|
	Glyma05g09400	protein kinase C activation	-||-||||||||||||:||:
	Glyma11g05720	transcription factor activity	|-|||||-||||||||||:|
	Glyma11g10790	RNA-binding protein	|||:||:|||:|||||||:-
	Glyma14g01950	A2L zinc ribbon domain	||||||-|||||||||:-||
	Glyma19g36200	transcription factor activity	|-|||||-||||||||||:|
gma-MIR172c	Glyma01g39520	AP2 domain-containing transcription factor	|||||||||||||:||||||
	Glyma03g33470	AP2 domain-containing transcription factor	|||||||||||||:||||||
	Glyma05g18170	AP2 domain-containing transcription factor	||||||||||||||-||||
	Glyma05g31790	GTPase Rab2, small G protein superfamily	-||||-|||||-|||||||||
	Glyma08g15040	GTPase Rab2, small G protein superfamily	-||||-|||||-|||||||||
	Glyma11g05720	AP2 domain-containing transcription factor	|||||||||||||:||||||
	Glyma11g15650	AP2 domain-containing transcription factor	|||||||||||||:|||||:
	Glyma12g07800	AP2 domain-containing transcription factor	|||||||||||||:|||||:
	Glyma13g40470	AP2 domain-containing transcription factor	|||||||||||||:|||||:
	Glyma15g04930	AP2 domain-containing transcription factor	|||||||||||||:|||||:
	Glyma17g18640	AP2 domain-containing transcription factor	||||||||||||||-||||
	Glyma19g36200	AP2 domain-containing transcription factor	|||||||||||||:||||||
gma-MIR172g	Glyma06g15630	ubiquitin-protein ligase activity	|||||||||-|||||-:|||
	Glyma10g27970	ATP binding cassette protein	||||||:||||--|||||||
gma-MIR172h-3p	Glyma01g39520	AP2 domain-containing transcription factor	-|||||||||||||:|||||
	Glyma03g33470	AP2 domain-containing transcription factor	-|||||||||||||:|||||
	Glyma11g05720	AP2 domain-containing transcription factor	-|||||||||||||:|||||
	Glyma11g15650	AP2 domain-containing transcription factor	-|||||||||||||:||||||
	Glyma12g07800	AP2 domain-containing transcription factor	-|||||||||||||:||||||
	Glyma13g40470	AP2 domain-containing transcription factor	-|||||||||||||:||||||
	Glyma15g04930	AP2 domain-containing transcription factor	-|||||||||||||:||||||
	Glyma19g36200	AP2 domain-containing transcription factor	-|||||||||||||:|||||

gma-MIR172h-5p	Glyma06g13450	putative ATP-dependent Clp-type protease	||||||--|||||||||||
	Glyma10g08730	nitrate, fromate, iron dehydrogenase	||-|:-|||||:|||||||||
	Glyma10g30570	targeting protein for Xklp2	||||::||||||:||:||:|||
	Glyma11g05580	GTP-binding ADP-ribosylation factor	-|||::||||||||||||||-
	Glyma11g06830	ubiquitin-protein ligase	--||||||-|||||||||||
	Glyma15g12600	protease inhibitor	||-|:|||||||-|:||||||
gma-MIR2118b-5p	Glyma02g36360	replication termination factor 2	|||||-|||:||||||:||-
	Glyma17g08320	replication termination factor 2	|||||-|||:||||||:||-
gma-MIR395a,b,c	Glyma06g11150	sulfate transporter	||||||--|||||||||||
	Glyma10g38760	ATP sulfurylase	|||||||:||||||:|||||-
	Glyma17g10050	gibberellin regulated protein	|-|||||||||::-|||||||
	Glyma18g02240	sulfate transporter	-|||||:||||||||||||-
	Glyma18g03110	phospholipase C-related	|-|||:|:|||||:|:|||||
	Glyma20g28980	ATP sulfurylase	|||||||:||||||:|||||-
gma-MIR396a-3p	Glyma11g12580	adenylate cyclase-associated protein	:||:|||::|-|||||||||
	Glyma12g04790	adenylate cyclase-associated protein	:||:|||::|-|||||||||
	Glyma16g00260	core-2/I-branching enzyme	||||||-|||||||:|||||
gma-MIR396a-5p	Glyma01g34650	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma01g44470	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma03g02500	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma03g35010	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma04g40880	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma06g13960	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma07g04290	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma09g07990	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma10g07790	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma11g01060	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma11g11820	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma12g01730	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma13g16920	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma13g21630	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma14g10090	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma15g19460	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma16g00970	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma17g05800	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma17g35090	growth-regulating transcription factor	-||||||||||||-|||||||
	Glyma19g37740	growth-regulating transcription factor	-||||||||||||-|||||||
gma-MIR396b-3p	Glyma02g01370	AMP dependent ligase/synthetase	|||:||||-|||||||||-|
	Glyma13g12120	anticodon binding domain	|||:||||-|||||||||-|
	Glyma16g03560	PPR-containing protein	|||:|||-:|||-||||||||
gma-MIR396b-5p	Glyma12g30730	stress responsive A/B Barrel Domain	|||||||||:|||-|||||-|
	Glyma17g08020	heat shock protein 70KDa	|||-|::||||:||||||||:
	Glyma18g01000	rubber elongation factor protein	||-||||:-||||:|||||||
gma-MIR396f	Glyma04g14970	drug transporter-related	||||||||-||:-:||||||
	Glyma08g37250	SEL-1-Like proteins	:|-||-:|||||||||||||
	Glyma18g01030	protein binding	:|-|||-||||||||||||

	Glyma18g47180	SEL-1-Like proteins	:|-||-:||||||||||||
gma-MIR396h-3p	Glyma11g12580	adenylate cyclase-associated protein	:||:|||::|-|||||||||
	Glyma12g04790	adenylate cyclase-associated protein	:||:|||::|-|||||||||
	Glyma12g28570	core-2/I-branching enzyme	||||||-|||||||:|||||:
	Glyma16g00260	Core-2/I-branching enzyme	||||||-|||||||:|||||:
gma-MIR397a,b	Glyma01g26750	multicopper oxidases	:||||||||||||||||||
	Glyma02g38990	multicopper oxidases	-|||||||||||||-|||||
	Glyma03g14450	multicopper oxidases	:|||||:||||||||||||
	Glyma03g15800	multicopper oxidases	:|||||:||||||||||||
	Glyma07g16080	multicopper oxidases	-|||||||||||||:|||||
	Glyma07g17170	multicopper oxidases	-||||||||||||:|||||-|
	Glyma08g47380	multicopper oxidases	-|||||||||||||:|||||
	Glyma11g07430	multicopper oxidases	||||||:|||||-||-|||||
	Glyma11g14600	multicopper oxidases	-|||||:||||||||||||
	Glyma12g06480	multicopper oxidases	-|||||:||||||||||||
	Glyma14g37040	multicopper oxidases	-||||||||||||||-|||||
	Glyma18g02690	multicopper oxidases	||||||:||||||||||-|
	Glyma18g38710	multicopper oxidases	-|||||:||||||||||||
	Glyma18g40070	multicopper oxidases	-|||||||||||||:|||||
	Glyma18g41910	multicopper oxidases	-||||||||||||:|||||-|
gma-MIR408a	Glyma03g26060	plastocyanin-like domain	||:|||||||||||||||--
	Glyma04g42120	plastocyanin-like domain	-:||||||||||||||||||
	Glyma06g12680	plastocyanin-like domain	-:||||||||||||||||||
	Glyma07g13840	plastocyanin-like domain	||:|||||||||||||||--
gma-MIR408b-5p	Glyma11g20520	transcription factor HEX	||||||||||:|-||||||-
	Glyma12g08080	transcription factor HEX	||||||||||:|-||||||-
gma-MIR1508b	Glyma14g02680	Ca2+/calmodulin-dependent protein kinase	|||||-:||||||:||||:|:
	Glyma16g28020	PPR-containig protein	|-|||||-||||||||||||
	Glyma19g43790	protein phosphatase	||:|||:|||||||::|||||-
gma-MIR1510b	Glyma02g04750	ATP binding	|||||-||||||||:|||||-
	Glyma04g39740	transmembrane receptor activity	|||||||||||-|||||||||
	Glyma09g38390	oxidoreductase activity	-|||||||-||||-|||||||
	Glyma13g25440	LRR-cointaing protein	|||||:||:|||||:||-|||
	Glyma13g26230	LRR-cointaing protein	||-|||||:||||||||-|||
	Glyma13g26250	LRR-cointaing protein	|||||:||:|||||:||-|||
	Glyma16g22620	ATP binding	|||||-||||||||:|||||-
	Glyma16g23800	LRR-cointaing protein	||-||:|||||-|||||||||
	Glyma17g02100	F-box domain-containing protein	||||||||||||||||||-||
	Glyma19g07660	LRR-cointaing protein	||||||:||||||||||:|||
	Glyma19g07700	LRR-cointaing protein	|||||||||||-|||||:|||
	Glyma20g26970	predicted E3 ubiquitin ligase	|||-||:||||||||:||::|
gma-MIR1512b	Glyma13g43710	CASC3/Barentsz eIF4AIII binding	|||:||-|||||||::|||||
	Glyma15g01650	CASC3/Barentsz eIF4AIII binding	|||:||-|||||||::|||||
gma-MIR1513a	Glyma08g10360	F-box domain-containing protein	|-|||||||||||||:|||-|
	Glyma08g27820	F-box domain-containing protein	||||||||||||||||||-|||
	Glyma08g27850	F-box domain-containing protein	|||||||||||||||||-||-|

	Glyma08g27950	F-box domain-containing protein	|:||-||||||||||||||
	Glyma10g26670	F-box domain-containing protein	|-|||||||||||||:|||-|
	Glyma18g52630	protein geranylgeranyltransferase	|||||||||||||||||||-|
gma-MIR1513b	Glyma08g10360	F-box domain-containing protein	|||||||||||:|||-|||
	Glyma08g27950	F-box domain-containing protein	|||||||||||||||-||-
	Glyma15g18380	transcription factor activity	-||||||||||-|||:||:
	Glyma13g00200	transcription factor activity	-||||||||||-|||:||:
	Glyma17g06290	transcription factor activity	-||||||||||-|||:||:
gma-MIR1513c	Glyma02g16510	F-box domain-containing protein	|:||-|||||-||||||||||
	Glyma08g10360	F-box domain-containing protein	|||||||-||||||||||||
	Glyma08g27850	F-box domain-containing protein	|:||-|||||||||||||||
	Glyma08g27950	F-box domain-containing protein	|:||-|||||||||||||||
	Glyma17g02100	F-box domain-containing protein	-:||-|||||||||||||||
gma-MIR4376a-3p	Glyma02g07160	Predicted membrane protein	||-||::|-|||||||||||
	Glyma02g15650	inositol 5-phosphatase	||:-|:|-||||||||||||
	Glyma05g01180	ribosomal protein S4	||:|||--||||||||||:|
	Glyma07g32780	inositol 5-phosphatase	||:-|:|-||||||||||||
	Glyma16g26110	predicted membrane protein	||-||::|-|||||||||||
gma-MIR4413a	Glyma09g07290	PPR-containing protein	|||||||-|||||||||||
	Glyma13g30610	ATP-dependent RNA helicase	-:||-|:||||||||||||
	Glyma13g36030	GH3 auxin-responsive promoter	||-|||:||::||||||:||
	Glyma16g25410	PPR-containing protein	|||||||-||||:|||||||
	Glyma16g27600	PPR-containing protein	|||||||-|||||||||||
gma-MIR4413b	Glyma08g47900	multifunctional chaperone	|||||:|||:||-|:|:|||
	Glyma09g07290	PPR-containing protein	||||||-||||||||||||
	Glyma11g07490	transcriptional repressor activity	|||||:|||:||-|:|:|||
	Glyma13g30610	ATP-dependent RNA helicase	:||-|:||||||||||||:
	Glyma14g04000	isomerase activity	:||||:|||::|||:|||||
	Glyma16g25410	PPR-containing protein	||||||-||||:|||||||
	Glyma16g27600	PPR-containing protein	||||||-||||||||||||
	Glyma17g03670	N-acetyltransferase activity	-|||:|||:-|||||||||
gma-MIR4415a-3p	Glyma08g43800	importin alpha-related	-:||||||||||||||||||:|:
	Glyma08g44150	RNA-binding protein	|::-|||||||||:|:|||||
	Glyma08g44170	RNA-binding protein	|::-|||||||||:|:|||||
	Glyma13g03650	multicopper oxidases	-||-|||||||||:||||||
	Glyma18g08590	RNA-binding protein	|::-|||||||||:|:|||||
	Glyma20g12150	multicopper oxidases	-||-|||||||||:||||||
	Glyma20g12230	multicopper oxidases	-||-|||||||||:||||||

[a] Data from Phytozome version 6.0. [b] Pairing obtained in psRNATarget Server: "|" indicates a
Watson-Crick base pairing; ":" is a G:U base pairing, and "-" indicates a mismatch.

Additional file 3. The soybean transcript loci which were identified as new-miRNA families target by degradome sequencing. The miRNA target site is indicated in red and underlined while the degradome sequence is highlighted.

```
Glyma13g01690 TCCTGATGGTCTACCTGAGACTGATCTGGATGCCACACAGGATATTCCTTCCC
                     -|||||||||||||||-:|
miR-Seq01          19 ACAGAUGGACUCUGACAGG 1

Glyma16g26070 GAGGGGCAGGGCACGAGGTTCCACTTCATAAACCCCG
                    |||||-|-|:|:||||||||
miR-Seq06         20 CGTCCGGAGTTTCAAGGTGA 1

Glyma04g01020 TTCTGCTACTCTTCTCAAGTCATCTCCTGTTCTTGACAAGT
                    ||||-|:||||||||||-|
miR- Seq07        20 GATGTGGAGAGTTCAGTAAA 1

Glyma19g27280 GTAGCTGCATTTTTTAGGTCATTTAATTTCTTGTTATTCAGA
                   ||:||::|:|:|:|||||||
miR- Seq07       20 GATGTGGAGAGTTCAGTAAA 1

Glyma15g13500 TGTGATGCATCAGTTTTGTTGAACAACACTGCCACCATAG
                    ::||-|||||||:|:||||||
miR-Seq11         21 GTGTTGTCAAAATAGCTTGTT 1

Glyma09g02600 GTGACGCATCAGTTTTGTTGAACAACACTGCCACCATAG
                    |:||-|||||||:|:||||||
miR-Seq11         21 GTGTTGTCAAAATAGCTTGTT 1

Glyma08g20670 TTGGTGCATCATCTAGGATTAAGAGTACATGCATCTATGGTG
                    :||||||||||:||:|||||-
miR- Seq12        21 TGTAGTAGATCTTAGTTCTCT 1

Glyma07g01260 TTGGTGCATCATCTAGGATTAAGAGTACATGCATCTATGGTG
                    :||||||||||:||:|||||-
miR- Seq12        21 TGTAGTAGATCTTAGTTCTCT 1

Glyma17g20860 TTTCAATTCCAGATGTCTGATTATAAAATTGTGGTGACTTCA
                    ||||||||||||-||:|||||
miR-Seq16         21 TAAGGTCTACAGTCTGATATT 1

Glyma05g09440 TTTCAATTCCAGATGTCTGATTATAAAATTGTGGTGACTTCT
                    ||||||||||||-||:|||||
miR-Seq16         21 TAAGGTCTACAGTCTGATATT 1
```

CAPÍTULO V

The role of MIR4415 in soybean response to asian soybean rust infection and water deficit stress

Franceli Rodrigues Kulcheski[1], Pablo Andre Manavella[2], Detlef Weigel[2] and Rogério Margis[1,*]

* Corresponding author. E-mail: rogerio.margis@ufrgs.br

[1] Laboratory of Genomes and Plant Population, Center for Biotechnology, building 43431, Federal University of Rio Grande do Sul - UFRGS, P.O. Box 15005, CEP 91501-970, Porto Alegre, RS, Brazil

[2] Department of Molecular Biology, Max Planck Institute for Developmental Biology, D-72076 Tübingen, Germany

Manuscript in preparation

Abstract

MicroRNAs (miRNAs) are endogenous and small non-coding RNAs that regulate the expression of protein-coding-genes at the post-transcriptional level. Plant miRNAs are thought to largely regulate transcripts by single, highly complementary target sites. Consequently, plant miRNAs are predicted to have only a limited number of messenger RNA (mRNA) targets. In plants, the majority of annotated *MIRNA* are family- or species-specific, however little is known about their targets and the biologic pathways in which they are involved. Employing 5'RACE and RT-qPCR we investigated the candidate target for a soybean-specific MIR4415, which was previously detected by our group using high-throughput sequencing. We observed that MIR4415 expression profile varied between different soybean genotypes submitted to water-deficit and asian soybean rust stresses. In the present study, the ascorbate oxidase (AO) mRNA was shown to be regulated by MIR4415. AO is an apoplastic enzyme that catalyses the oxidation of ascorbic acid (AA) to monodehydroascorbate, a radical which rapidly degrades to dehydroascorbate, playing a major role in controlling the redox state of the apoplast. In our study we observed that AO is induced in both genotypes when infected by *P. pachyrhizi*, providing evidence that it is a component of the oxidative stress pathway to be considered in studies of soybean-rust interaction.

Introduction

MicroRNAs (miRNAs) are endogenous and small non-coding RNAs that regulate the expression of protein-coding-genes at the post-transcriptional level. In plants, MIRNA genes are transcribed by RNA polymerase II enzymes (Pol II) generating primary miRNA (pri-miRNA). The pri-miRNA forms an imperfect fold-back structure, which is processed into a stem-loop precursor (pre-miRNA) by nuclear RNaseIII-like enzymes called DICER-LIKE proteins (e.g., DCL1) [1]. The resulting pre-miRNA contains a miRNA:miRNA* intermediate duplex, formed by a self-complementary fold-back structure. A mature miRNA sequence can range from 19 to 24 nucleotides (nt) in length and act as a regulatory molecule in post-transcriptional gene silencing by base pairing with target mRNAs. This leads to mRNA cleavage or translational repression, depending on the degree of complementarity between the miRNA and its target transcript [2, 3]. Plant miRNAs are thought to largely regulate transcripts by single, highly complementary target sites [1]. Consequently, plant miRNAs are predicted to have only a limited number of messenger RNA (mRNA) targets.

In plants, a minority of annotated *MIRNA* gene families are conserved between plant families, while the majority are family- or species-specific, suggesting that most known *MIRNAs* genes arose relatively recently in evolutionary time [4]. A high proportion of species-specific or nonconserved *MIRNA* genes were also observed in *Physcomitrella patens*, *Selaginella moellendorffii*, *Oryza sativa*, *Medicago truncatula*, and *Glycine max* [5-12]. Actually, data from additional plant genomes indicate that the majority of the species subjected to high-throughput smRNA sequencing have been found to possess non-conserved miRNAs (www.mirbase.org; [13]). Given that a large number of miRNA families are specie-specific or restricted to closely related species, it

is reasonable to suggest that plants harbor relatively large numbers of recently spawned *MIRNA* loci.

Usually species-specific miRNAs target mRNAs have a wider range of functions than the targets of conserved miRNAs. For example, only about 1/9 of a newly discovered set of non-conserved miRNAs in arabidopsis target transcription factors; instead, their targets include genes involved in metabolism, signal transduction, protein modification, and RNA or carbohydrate binding [14, 15]. It will be interesting to learn about these miRNAs biological roles, since it is known that miRNAs are involved in a variety of biological and metabolic process, such as development, hormone signaling, abiotic stress responses, and pathogen responses [5, 16-20].

In *Glycine max,* the MIR4415 that was identified by Joshi et al [10] and also by our group using high-throughput sequencing technology [11] is an example of species-specific miRNA. In our previous work, we identified that this mature miRNA was originated from two different loci localized in different soybean chromosomes [11]. MIR4415 are highly specific, since copies of this miRNA were not observed in other species, indicating their recently origin. Their precursors have large fold-back sequences, which are also highly conserved. We observed that MIR4415 expression profile varied between different soybean genotypes submitted to water-deficit and asian soybean rust stresses.

Since soybean is one of the most economically important cultures, investigation of the different reactions that occur at a molecular level can improve our knowledge about how these plants cope with different stresses that affect their development. For this reason, employing modified 5'RACE and RT-qPCR, we performed the validation of the predicted MIR4415 target gene, and also attempted to understand the biological relevance of this miRNA during water and soybean rust stress.

Material and Method

Plant materials and treatments

Water deficit assay

For water deficit treatment, we used the soybean (*Glycine max* (L.) Merrill) cultivars 'Embrapa 48' as a drought-tolerant standard and 'BR 16' as a sensitive standard [21]. Plants were grown in a greenhouse at Embrapa-Soybean (Londrina, Brazil) using a hydroponic system compound for plastic containers (30 liters) and an aerated pH 6.6-balanced nutrient solution. Seeds were pre-germinated on moist filter paper in the dark at 25°C ± 1°C and in 65% ± 5% relative humidity. Plantlets were then placed in polystyrene supports so the roots of the seedlings were completely immersed in the nutrient solution. Each seedling tray was maintained in a greenhouse at 25°C ± 2°C and in 60% ± 5% relative humidity under natural daylight (photosynthetic photon flux density (PPFD) =1.5 x 10^3 µmoles m^{-2} s^{-1}) for a 12 h day. After 15 days, seedlings with the first trifoliate leaf fully developed (V2 developmental stage) [22] were submitted to different water-deficit treatments according to Martins et al. [23]. The nutrient solution was removed from each plastic container where the roots were kept in the tray in the dark without nutrient solution or water for 0 minutes (T0 or control) and 125 minutes (T125 or treated). At the end of each water-deficit period, the roots of the seedlings were immediately frozen in liquid nitrogen and stored at -80°C until RNA extraction. The experimental design was a factorial (cultivars x duration of water deficit) with three replicates. Each replicate was composed of five plantlets that were sampled in bulk. So the respective samples analyzed by RT-qPCR for the water deficit-stress are: roots of drought-sensitive seedlings submitted to 0 minutes of stress (Drought-Sensitive Root Control (DSRC)); roots of drought-sensitive seedlings

submitted to 125 minutes of stress (Drought-Sensitive Root Treated (DSRT)); roots from drought-tolerant seedlings submitted to 0 minutes of stress (Drought-Tolerant Root Control (DTRC)); and roots of drought-tolerant seedlings submitted to 125 minutes and 150 minutes of stress (Drought-Tolerant Root Treated (DTRT)).

Asian Soybean Rust assay

The ASR reaction was evaluated in soybean plants in a greenhouse at Embrapa-Soybean (Londrina, Brazil) using a field population of *Phakopsora pachyrhizi* collected from soybean fields in the state of Mato Grosso, which were maintained for over 10 generations on the susceptible cv. BRSMS-Bacuri. ASR identification was confirmed by ITS-sequencing analysis as described by Silva et al. [24], and it revealed a similarity to the MUT Zimbabwe isolate. The soybean plants were grown in a pot-based system. The 'Embrapa 48' genotype was used as a susceptible host plant, which develops a susceptible lesion (TAN) after *Phakopsora pachyrhizi* infection. The 'PI561356' genotype was used as the resistant host, which carries an ASR resistance gene mapped onto linkage group G (Ricardo V. Abdelnoor, EMBRAPA Soja, personal communication) and develops a reddish-brown (RB) lesion with few or no spores.

Urediniospores were collected from infected BRSMS-Bacuri plants in a separate greenhouse by tapping infected leaves over a plastic tray. The urediniospores were then diluted in distilled water with 0.05% Tween-20 to a final concentration of 3×10^5 spores/mL. This spore suspension was sprayed onto three plants per pot at the V2 to V3 growth stages [22]. A solution without the spores was used for the mock inoculations. Following the ASR or mock inoculations, water-misting bags were placed over all plantlets for one day to aid the infection process and to prevent cross-contamination of the mock-infected plants. The third trifoliolate leaves of six plants were collected 12

hours after inoculation (hai) for RNA extraction. The experiment followed a completely randomized design with the three replicates as blocks and a full factorial treatment structure consisting of three treatment factors: hai (12 hours), genotype (resistant or susceptible), and inoculation type (ASR or mock). The samples analised were: leaves of rust-susceptible seedlings with mock inoculation (Rust-Susceptible Leaf Control (RSLC)); leaves of rust-susceptible seedlings with rust-spore inoculation (Rust-Susceptible Leaf Treated (RSLT)); leaves of rust-resistant seedlings with mock inoculation (Rust-Resistant Leaf Control (RRLC)); and leaves of rust-resistant seedlings with rust-spore inoculation (Rust-Resistant Leaf Treated (RRLT)).

RNA extraction and cDNA synthesis

For each treatment, total RNA was isolated by extraction with Trizol, (Invitrogen) and the quality was evaluated by electrophoresis on a 1.0% agarose gel. The cDNA synthesis for miRNA analysis was carried out by multiplex technique [25, 26] from approximately 2 µg of total RNA. Each reaction was primed with a pool of 0.5 µM 10 gene-specific stem-loop primers [27] (IDT). The cDNA synthesis for mRNAs we employed 1 µM of oligonucleotide dT24V (Invitrogen) per reaction. Before transcription, RNA and primers were mixed with RNase-free water to a total volume of 10 µL and incubated at 70°C for 5 minutes followed by ice-cooling. Then, 6 µL 5x RT-Buffer (Promega), 1 µL of 5 mM dNTP (Ludwig) and 1 µL MML-V RT Enzyme 200 U (Promega) were added for a final volume of 30 µL. The synthesis for miRNAs and mRNAs cDNA were performed at 42°C for 30 minutes and 40°C for 60 minutes respectively on a Veriti Thermal Cycler (Applied Biosystem), and inactivation of the enzyme was completed at 85°C for five minutes. All cDNA samples were 50-fold diluted with RNase-free water before being used as a template in RT-qPCR analysis.

The stem-loop primers, used for miRNAs cDNA synthesis, were designed according to Cheng et al. [27]. The stem-loop sequence consists of 44 conserved and six variable nucleotides that are specific to the 3' end of the miRNA sequence (5' GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACNNNNNN 3').

Prediction of miRNA targets

Target prediction for miRNAs is straightforward because it is assumed that most of them match their targets with almost perfect complementarities [1, 3]. The putative target genes for MIR4415ab-3p and -5p were searched for by using the web-based computer psRNA Target Server (http://plantgrn.noble.org/psRNATarget/) [28] which can identify putative targets that may be regulated at post-transcriptional or at translational levels. The total scoring for an aligment was calculated based on the miRNA length, and the sequences were considered to be miRNA targets if the total score were less than 3.0 points (mismatch=1 and G:U=0.5). Results from these analyses were individually inspected in the soybean genome on the Phytozome web server (http://www.phytozome.net/), where the loci and protein annotation were obtained.

Cleavage site mapping of target mRNA

Total RNA was extracted from root and leaves of soybean seedlings with trizol and posterior purified using FastTrack™ MAG Micro mRNA Isolation Kit (Invitrogen). A Gene-Racer kit (Invitrogen) was used for 5'RACE (5' rapid amplification of cDNA ends), except the decapping protocol was not carried out, and the adapter was linked directly to mRNA. Two nested gene-specific reverse oligonucleotides were used for 5′ RACE. The PCR products were resolved on 2% agarose gels and detected by ethidium

bromide staining. The products from the second (nested) round of 5'RACE were gel purified using a Wizard preps PCR purification kit (Promega) and ligated into pGEM-T easy (Promega). Plasmids were transformed into *Escherichia coli* DH5α cells and purified. Clones were investigating with colony PCR to verify they contained inserts of the correct size and were sequenced. Primers used for this analysis are shown in Supplementary file 1.

Reverse transcription quantitative PCR reaction (RT-qPCR)

All quantitative PCR were performed in a BIO-RAD CFX384 Real-time PCR Detection System (BIO-RAD) using SYBR Green I (Invitrogen) to detect double-strand cDNA synthesis. Reactions were done in a volume of 10 µL containing 5 µL of diluted cDNA (1:50), 1X SYBR Green I (Invitrogen), 0.1 mM dNTP, 1X PCR Buffer, 3 mM $MgCl_2$, 0.25 U Platinum Taq DNA Polymerase (Invitrogen) and 200 nM of each reverse and forward primers. The employed primers for the RT-qPCR analysis are listed in Supplementary file 1. The miRNA primers were designed based on the fully tested miRNA sequence (forward), and the reverse primer was the universal reverse primer for miRNA [27]. All samples were analyzed as technical quadruplicates with a no-template control also included. The conditions were set as the following: an initial polymerase activation step for 5 min at 94°C, 40 cycles of 15 sec at 94°C for denaturation, 10 sec at 60°C for annealing and 25 sec at 72°C for elongation. At the end of cycling protocol, a melting-curve analysis was included and fluorescence measured from 65 to 99°C. Threshold and baselines were manually determined using the BIO-RAD CFX Manager™ Software. We used MIR156b and MIR1520d as reference genes [29].To calculate the relative expression of the miRNAs, we used the $2^{-\Delta\Delta Ct}$ method. Student's t-

test was performed to compare pair-wise differences in expression and the means were considered significantly different when $P < 0.05$.

Results and Discussion

Prediction and validation of MIR4415-target mRNA

The potencial miRNAs targets were predicted by the psRNA-target server using both MIR4415ab-3p and -5p as queries against the soybean transcriptome. The results of the analysis pointed to a *Glycine max* ascorbate oxidase transcript. We selected the binding site sequence and submitted to a BLAST in the Phytozome. The results revealed three different transcripts of ascorbate oxidase enzyme, which are originated from the respective genes: *Glyma13g03650*; *Glyma20g12150* and *Glyma14g04530*. These three genes are highly conserved, and an exact binding site for the MIR4415 was observed in *Glyma13g03650* and *Glyma20g12150*, while a single nucleotide mismatch was observed to *Glyma14g04530* (Figure 1 -B). This nucleotide is localized at the last position of the 5'end in the mRNA binding site, which theoretically could not affect the potential miRNA binding activity.

The next step was to confirm the cleavage site in vivo using the modified 5'RACE technique. Since the conservation among the three transcripts is really highly, it was very difficult to design gene-specific (employed in the first round of PCR) and nested primers (used in the second PCR) for this step. Since the position for these primers should be fixed, the primers could hybridize in all three different genes. However, after the cloning and sequencing of the correct bands amplified by PCR, we performed a blast and identified only sequences from Glyma13g03650 being originated from the MIR4415 cleavage (Figure 1 – A). Using 5'RACE, we amplified cDNA corresponding to this mRNA fragment. Sequence analysis of 19 independent cDNA

clones produced identical results and placed the 5' end of the cleaved fragment at nucleotide 319 of the Glyma13g03650 mRNA. This nucleotide position is located in the first exon at 5'extremity of Glyma13g03650 mRNA (Figure 1).

The experimental validation of a plant miRNA target usually relies on the detection of the endonucleolytic cleavage guided by the small RNA. Many AGO proteins, such as arabidopsis AGO1 [30], cleave the target RNA between positions 10 and 11 of the miRNA 5' end [31, 32]. Here we provide evidence showing that MIR4415 directs Glyma13g03650 mRNA cleavage in vivo between the 10^{th} and 11^{th} nucleotides of the complementary MIR4415 binding site (Figure 1 – B). This result is consistent with previous findings of others soybean miRNAs and their respective targets [10, 12].

Expression profile of MIR4415 and respective targets during water deficit and asian soybean rust stress

To confirm the results from 5'RACE analysis we performed RT-qPCR in order to quantify the relative expression of mature MIR4415 and the transcripts of Glyma13g03650, Glyma20g12150 and Glyma14g04530. Employing the qPCR technique, our goal was to investigate the miRNA and mRNAs profiles, once the miRNA and its presumptive target mRNA accumulation are usually inferred to be inversely correlated. However, for all RNAs samples it was not possible to amplify Glyma14g04530, and for this reason it will not appear in the expression profile analyses.

In a previous study, MIR4415 showed to be differently expressed in different genotypes under water deficit and asian soybean rust stress [11] and for this reason we reproduced the investigation of this miRNA plus Glyma13g03650 and Glyma20g12150 under the same conditions.

In the water deficit assay, we observed that the expression patterns of MIR4415 presented an effective rise in its expression level in the sensitive plants, while this higher level was also observed in the tolerant genotype under control condition and after water stress (Figure 2). Surprisingly Glyma13g03650 and Glyma20g12150 presented the same profile in the sensitive genotype that the MIR4415 and an equivalent increase was observed in the tolerant genotype under stress. Glyma13g03650 was more expressed than Glyma20g12150, but the expression pattern for all conditions was similar. This result was unexpected, and revealed that is not possible to predict the function of a miRNA only validating the cleavage sites in a determined mRNA.

The RT-qPCR analyses during asian soybean infection (Figure 3) showed a strong decrease of MIR4415 during the rust infection when compared with the control in the susceptible plants. In this same condition, Glyma13g03650 and Glyma20g12150 were up-regulated when the plants were under pathogen infection. This scenario fits with those of genes down or upregulated by miRNAs, and was expected, since Glyma13g03650 and Glyma20g12150 were validated as MIR4415 target. Nevertheless, the same prognostic was not confirmed in the resistant genotype. While the MIR4415 didn’t vary significantly, Glyma13g03650 and Glyma20g12150 presented an increased expression level in the stressed condition comparing to the mock. This intriguing situation, allow to consider that a different mechanism take place in the resistant genotype, with other factors than MIR4415 being involved in the pathogen recognition and molecular responses. This hypothesis might be quite likely since the two genotypes, susceptible and resistant, produce different responses after the pathogen infection.

The role of MIR4415 target

The discovery of a miRNA target and the validation of the in vivo cleavage is an important achievement, but the identification of which biological process it is being involved is the crucial question. In this way, employing 5'RACE and RT-qPCR technologies, we identified a new target for the specie-specific MIR4415ab.

Ascorbate oxidase (AO) is an apoplastic enzyme that catalyses the oxidation of ascorbic acid (AA) to monodehydroascorbate, a radical which rapidly degrades to dehydroascorbate, playing a major role in controlling the redox state of the apoplast [33, 34]. Ascorbic acid is considered to be the most abundant antioxidant in the apoplast: the apoplast contains up to 5% of the leaf's ascorbate pool [35], a pool which is generally highly oxidized in contrast to the symplastic pool. Regulation of the apoplastic ascorbate pool occurs via transport of ascorbate and dehydroascorbate between the cytosol and the apoplast [34, 36-38].

Since the ascorbic acid oxidase was discovery by Albert von Szent-Györgyi in 1931 [39], more than 80 years have passed, however the clear biological function of AO remains controversial. Several studies that were developed during this time have different hypothesis for AO function. The most widely opinion prevailed that AO plays a role in cell elongation because of its extracellular localization and its high activity in rapidly expanding tissues [40, 41].

Further studies on the effects of AO in tobacco have shown that this enzyme has a role in the perception of the environment or stress responses [38, 42] and may be linked to the cellular ascorbate redox state (defined as the ratio of reduced to total ascorbate).The ascorbate redox state has been shown to control processes including stomatal function [43], cell division [40, 44] and changes as a result of stress [34, 45]. Studies focusing light-induced expression of AO support the hypothesis of a mechanism acting to remove excess oxygen produced in the photosynthetic process [46]. Besides

this, Ballestrini et al. [47] showed that the expression of a *Lotus japonicus* AO gene is induced in the symbiotic interaction with both nitrogen-fixing bacteria and arbuscular mycorrhizal fungi. In this framework, high AO expression is viewed as a possible strategy to down-regulate oxygen diffusion in root nodules, and a component of AM symbiosis [47].

In soybean, a study involving the detection of genes that were inducted during infection with *Pseudomonas syringae* showed AO were among the genes up-regulated during the interaction plant-pathogen. The transcripts of ascorbate oxidase gene were inducted in both compatible and incompatible interaction, and the author proposed that GmAO could represent a new class of pathogenesis-related protein [48]. Although this study was conducted with the bacteria *P. syringae*, the results regarding to the induction of AO when the plant is infected by the pathogen resemble our results regarding the induction of AO, which also was observed in the compatible and incompatible interaction *P. pachyrhizi*-soybean.

Other work that corroborates in part with our findings was performed by Soria-Guerra et al. [49]. Employing transcriptome profiling analysis they were looking for genes controlling resistance to soybean rust. The experiment was conducted in resistant and susceptible *Glycine tomentella* genotypes triggered by *P. pachyrhizi* infection. Among 38,400 genes monitored using a soybean microarray, ascorbate oxidase was significant up-regulated in the resistant genotype when infected with *P. pachyrhizi* at 12hpi. The rust-susceptible genotypes demonstrated an inverse regulation [49].

The results show clear evidence that AO is differentially expressed in infected and control conditions. However, to understand the role of this enzyme in the establishment of the pathogen on the host is still something to be defined; most intriguing is to understand how the MIR4415 is involved in this pathway. Looking at

our results, we can observe that AO is induced in both genotypes when infected by *P. pachyrhizi*. Thinking about the activity of AO, it is known that the level of AA is diminished when AO is induced. Because ascorbic acid is a major antioxidant buffer and free-radical scavenger, consequently it also acts as a barrier against pathogen attack [38]. Susceptible plants with reduced accumulation of AA in the apoplast can be one of the factors that facilitate the establishment of the fungus on the leaf tissue, because production of reactive oxygen species and host-cell killing is a prerequisite for successful fungal development and pathogenesis of necrotrophic nourishment [50, 51]. And why this scenario would not be seen on the resistant genotype? One possibility would be the presence of major genes for resistance (R genes) that have a specific and differentiated response to infection by the fungus. Nevertheless this hypothesis needs additional approaches to be clarified.

Conclusion

The detection of cleaved products of miRNA target genes is necessary in the study of mechanisms by which miRNAs regulate their target genes. Employing 5'RACE and RT-qPCR techniques, we validated a new target of specie-specific miRNA. In the present study, Ascorbate oxidase mRNA was showed to be degraded by MIR4415, and apparently this mechanism is significantly affected by the asian soybean rust infection in a susceptible genotype. However, additional studies have yet to be performed in order to clarify how this regulation occurs in soybean.

References

1. Voinnet O: **Origin, biogenesis, and activity of plant microRNAs**. *Cell* 2009, **136**(4):669-687.

2. Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism, and function**. *Cell* 2004, **116**(2):281-297.

3. Bartel DP: **MicroRNAs: target recognition and regulatory functions**. *Cell* 2009, **136**(2):215-233.

4. Cuperus JT, Fahlgren N, Carrington JC: **Evolution and functional diversification of MIRNA genes**. *Plant Cell* 2011, **23**(2):431-442.

5. Axtell MJ, Snyder JA, Bartel DP: **Common functions for diverse small RNAs of land plants**. *Plant Cell* 2007, **19**(6):1750-1769.

6. Heisel SE, Zhang Y, Allen E, Guo L, Reynolds TL, Yang X, Kovalic D, Roberts JK: **Characterization of unique small RNA populations from rice grain**. *PLoS One* 2008, **3**(8):e2871.

7. Lelandais-Briere C, Naya L, Sallet E, Calenge F, Frugier F, Hartmann C, Gouzy J, Crespi M: **Genome-wide *Medicago truncatula* small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules**. *Plant Cell* 2009, **21**(9):2780-2796.

8. Sunkar R, Zhou X, Zheng Y, Zhang W, Zhu JK: **Identification of novel and candidate miRNAs in rice by high throughput sequencing**. *BMC Plant Biol* 2008, **8**:25.

9. Szittya G, Moxon S, Santos DM, Jing R, Fevereiro MP, Moulton V, Dalmay T: **High-throughput sequencing of *Medicago truncatula* short RNAs identifies eight new miRNA families**. *BMC Genomics* 2008, **9**:593.

10. Joshi T, Yan Z, Libault M, Jeong DH, Park S, Green PJ, Sherrier DJ, Farmer A, May G, Meyers BC *et al*: **Prediction of novel miRNAs and associated target genes in *Glycine max***. *BMC Bioinformatics* 2010, **11 Suppl 1**:S14.

11. Kulcheski FR, de Oliveira LFV, Molina LG, Almerão MP, Rodrigues FA, Marcolino J, Barbosa BJ, Stolf-Moreira R, Nepomuceno AL, Marcelino-Guimarães FC *et al*: **Identification of novel soybean microRNAs involved in abiotic and biotic stresses**. *BMC Genomics* 2011, **12**:307-322.

12. Subramanian S, Fu Y, Sunkar R, Barbazuk WB, Zhu JK, Yu O: **Novel and nodulation-regulated microRNAs in soybean roots**. *BMC Genomics* 2008, **9**:160.

13. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ: **miRBase: tools for microRNA genomics**. *Nucleic Acids Res* 2008, **36**(Database issue):D154-158.

14. Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL *et al*: **High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes**. *PLoS One* 2007, **2**(2):e219.

15. Willmann MR, Poethig RS: **Conservation and evolution of miRNA regulatory programs in plant development**. *Curr Opin Plant Biol* 2007, **10**(5):503-511.

16. Covarrubias AA, Reyes JL: **Post-transcriptional gene regulation of salinity and drought responses by plant microRNAs**. *Plant Cell Environ* 2010, **33**(4):481-489.

17. Krol J, Loedige I, Filipowicz W: **The widespread regulation of microRNA biogenesis, function and decay**. *Nat Rev Genet* 2010, **11**(9):597-610.

18. Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, Carrington JC, Weigel D: **Control of leaf morphogenesis by microRNAs**. *Nature* 2003, **425**(6955):257-263.

19. Arenas-Huertero C, Perez B, Rabanal F, Blanco-Melo D, De la Rosa C, Estrada-Navarrete G, Sanchez F, Covarrubias AA, Reyes JL: **Conserved and novel miRNAs in the legume *Phaseolus vulgaris* in response to stress**. *Plant Mol Biol* 2009, **70**(4):385-401.

20. Belkhadir Y, Subramaniam R, Dangl JL: **Plant disease resistance protein signaling: NBS-LRR proteins and their partners**. *Curr Opin Plant Biol* 2004, **7**(4):391-399.

21. Oya T, Nepomuceno AL, Neumaier N, Farias JRB, Tobita S, Ito O: **Drought Tolerance Characteristics of Brazilian soybean cultivars - Evaluation and characterization of drought tolerance of various Brazilian soybean cultivars in the field**. *Plant Prod Sci* 2004, **7**(2):129-137.

22. Fehr WR, Caviness CE: **Stages of soybean development. Special Report 80**. In. Ames, U.S.A.: Iowa State University of Science and Technology, Iowa Agriculture and Home Economics Experiment Station; 1977.

23. Martins PK, Jordão BQ, Yamanaka N, Farias JRB, Beneventi MA, Binneck E, Fuganti R, Stolf R, Nepomuceno AL: **Differential gene expression and mitotic cell analysis of the drought tolerant soybean (*Glycine max* L. Merrill Fabales, Fabaceae) cultivar MG/BR46 (Conquista) under two water deficit induction systems**. *Genet Mol Biol* 2008, **31**:512-521.

24. Silva DCG, Yamanaka N, Brogin RL, Arias CAA, Nepomuceno AL, Di Mauro AO, Pereira SS, Nogueira LM, Passianotto ALL, Abdelnoor RV: **Molecular mapping of two loci that confer resistance to Asian Rust in soybean**. *Theor Appl Genet* 2008, **117**:57-63.

25. Lao K, Xu NL, Yeung V, Chen C, Livak KJ, Straus NA: **Multiplexing RT-PCR for the detection of multiple miRNA species in small samples**. *Biochem Biophys Res Commun* 2006, **343**(1):85-89.

26. Tang F, Hajkova P, Barton SC, Lao K, Surani MA: **MicroRNA expression profiling of single whole embryonic stem cells**. *Nucleic Acids Res* 2006, **34**(2):e9.

27. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR *et al*: **Real-time quantification of microRNAs by stem-loop RT-PCR**. *Nucleic Acids Res* 2005, **33**(20):e179.

28. Dai X, Zhao PX: **psRNATarget: a plant small RNA target analysis server**. *Nucleic Acids Res* 2011, **39** (Web Server issue):W155-159.

29. Kulcheski FR, Marcelino-Guimaraes FC, Nepomuceno AL, Abdelnoor RV, Margis R: **The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean**. *Anal Biochem* 2010, **406**(2):185-192.

30. Baumberger N, Baulcombe DC: ***Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs**. *Proc Natl Acad Sci U S A* 2005, **102**(33):11928-11933.

31. Elbashir SM, Lendeckel W, Tuschl T: **RNA interference is mediated by 21- and 22-nucleotide RNAs**. *Genes Dev* 2001, **15**(2):188-200.

32. Llave C, Xie Z, Kasschau KD, Carrington JC: **Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA**. *Science* 2002, **297**(5589):2053-2056.

33. Barbehenn RV, Jaros A, Yip L, Tran L, Kanellis AK, Constabel CP: **Evaluating ascorbate oxidase as a plant defense against leaf-chewing insects using transgenic poplar**. *J Chem Ecol* 2008, **34**(10):1331-1340.

34. Garchery C, Gest N, Do PT, Alhagdow M, Baldet P, Menard G, Rothan C, Massot C, Gautier H, Aarrouf J *et al*: **A diminution in ascorbate oxidase activity affects carbon allocation and improves yield in tomato under water deficit**. *Plant Cell Environ* 2012, **36**(1):159-175.

35. Veljovic-Jovanovic SD, Pignocchi C, Noctor G, Foyer CH: **Low ascorbic acid in the vtc-1 mutant of *Arabidopsis* is associated with decreased growth and intracellular redistribution of the antioxidant system**. *Plant Physiol* 2001, **127**(2):426-435.

36. Horemans N, Foyer CH, Asard H: **Transport and action of ascorbate at the plant plasma membrane**. *Trends Plant Sci* 2000, **5**(6):263-267.

37. Pignocchi C, Fletcher JM, Wilkinson JE, Barnes JD, Foyer CH: **The function of ascorbate oxidase in tobacco**. *Plant Physiol* 2003, **132**(3):1631-1641.

38. Pignocchi C, Foyer CH: **Apoplastic ascorbate metabolism and its role in the regulation of cell signalling**. *Curr Opin Plant Biol* 2003, **6**(4):379-389.

39. Szent-Györgyi A: **On the function of hexuronic acid in the respiration of the cabbage leaf**. *J Biol Chem* 1931, **90**:385-393.

40. Kato N, Esaka M: **Expansion of transgenic tobacco protoplasts expressing pumpkin ascorbate oxidase is more rapid than that of wild-type protoplasts**. *Planta* 2000, **210**(6):1018-1022.

41. Mertz D: **Distribution and cellular localization of ascorbic acid oxidase in themaize root tip,** . *Amer J Bot* 1961 **48** 405-413.

42. Pignocchi C, Kiddle G, Hernandez I, Foster SJ, Asensi A, Taybi T, Barnes J, Foyer CH: **Ascorbate oxidase-dependent changes in the redox state of the apoplast modulate gene transcript accumulation leading to modified hormone signaling and orchestration of defense processes in tobacco**. *Plant Physiol* 2006, **141**(2):423-435.

43. Fotopoulos V, De Tullio MC, Barnes J, Kanellis AK: **Altered stomatal dynamics in ascorbate oxidase over-expressing tobacco plants suggest a role for dehydroascorbate signalling**. *J Exp Bot* 2008, **59**(4):729-737.

44. Potters G, de Gara L, Asard H, N. H: **Ascorbate and glutathione: guardians of the cell cycle, partners in crime?** *Plant Physiol Biochem* 2002, **40**:537–548.

45. Gest N, Page D, Birtic S, Gouble B, Gilbert L, Garchery C, Causse M, Stevens R: **Response of the fruit antioxidant system to the post-chilling period in two different tomato lines** *Funct Plant Sci Biotechnol* 2010, **4**:76–83.

46. De Tullio MC, Liso R, Arrigoni O: **Ascorbic acid oxidase: an enzyme in search of a role** *Biol Plantarum* 2004, **48**(2):161-166.

47. Balestrini R, Ott T, Güther M, Bonfante P, Udvardi MK, De Tullio MC: **Ascorbate oxidase: the unexpected involvement of a 'wasteful enzyme' in the symbioses with nitrogen-fixing bacteria and arbuscular mycorrhizal fungi** *Plant Physiol Biochem* 2012, **59**:71-79.

48. Chou WM, Shigaki T, Dammann C, Liu YQ, Bhattacharyya MK: **Inhibition of phosphoinositide-specific phospholipase C results in the induction of pathogenesis-related genes in soybean**. *Plant Biol* 2004, **6**(6):664-672.

49. Soria-Guerra RE, Rosales-Mendoza S, Chang S, Haudenshield JS, Padmanaban A, Rodriguez-Zas S, Hartman GL, Ghabrial SA, Korban SS: **Transcriptome analysis of resistant and susceptible genotypes of Glycine tomentella during *Phakopsora pachyrhizi* infection reveals novel rust resistance genes**. *Theor Appl Genet*, **120**(7):1315-1333.

50. Govrin EM, Levine A: **The hypersensitive response facilitates plant infection by the necrotrophic pathogen *Botrytis cinerea***. *Curr Biol* 2000, **10**(13):751-757.

51. Waller F, Achatz B, Baltruschat H, Fodor J, Becker K, Fischer M, Heier T, Huckelhoven R, Neumann C, von Wettstein D *et al*: **The endophytic fungus *Piriformospora* indica reprograms barley to salt-stress tolerance, disease resistance, and higher yield**. *Proc Natl Acad Sci U S A* 2005, **102**(38):13386-13391.

Figure legends

Figure 1 – Prediction and the cleavage assay for the potential mRNA target of MIR4415. A) Results from the 5'RACE PCR. The products from first (1st) and nested (2nd) rounds of PCR were analyzed by agarose gel electrophoresis. DNA from the expected size PCR products were gel purified before cloning, and clones were then sequenced and mapped in *Glyma13g03650* gene (data from Phytozome version 9.0). **B)** Predicted cleavage site confirmed by 5'RACE into the Glyma13g03650 transcript. Numbers in bold indicate the proportion of clones out of the total number analyzed that mapped to the canonical MIR4415 cleavage position (indicated by arrow). Pairing obtained in psRNATarget Server: "|" indicates a Watson-Crick base pairing; ":" is a G:U base pairing, and "○" indicates a mismatch.

Figure 2 - Differential relative expression of MIR4415 and the respective target genes *Glyma13g03650* and *Glyma20g12150* during deficit water stress assay. For the water deficit-stress assay, the four different samples are: DSRC (drought-sensitive seedlings root control); DSRT (drought-sensitive seedlings root treated); DTRC (drought-tolerant seedlings root control) and DTRT (drought-tolerant seedlings root treated). Note: (*) samples that significantly differ between control and stressed conditions ($P < 0.05$) according to a Student t-test statistical analysis.

Figure 3 - Differential relative expression of MIR4415 and the respective target genes *Glyma13g03650* and *Glyma20g12150* during asian soybean rust stress assay. For the rust-stress assay, the four samples are: RSLC (rust-susceptible seedlings leaves mock inoculation); RSLI (rust-susceptible seedlings leaves with rust-spore inoculation); RRLC (rust-resistant seedlings leaves with mock inoculation) and RRLT (rust-resistant seedlings leaves with rust-spore inoculation). Note: (*) samples that significantly differ between control and stressed conditions ($P < 0.05$) according to a Student t-test statistical analysis.

Supplementary file 1: Primer sequences and amplicon characteristics for all genes analyzed.

Figure 1

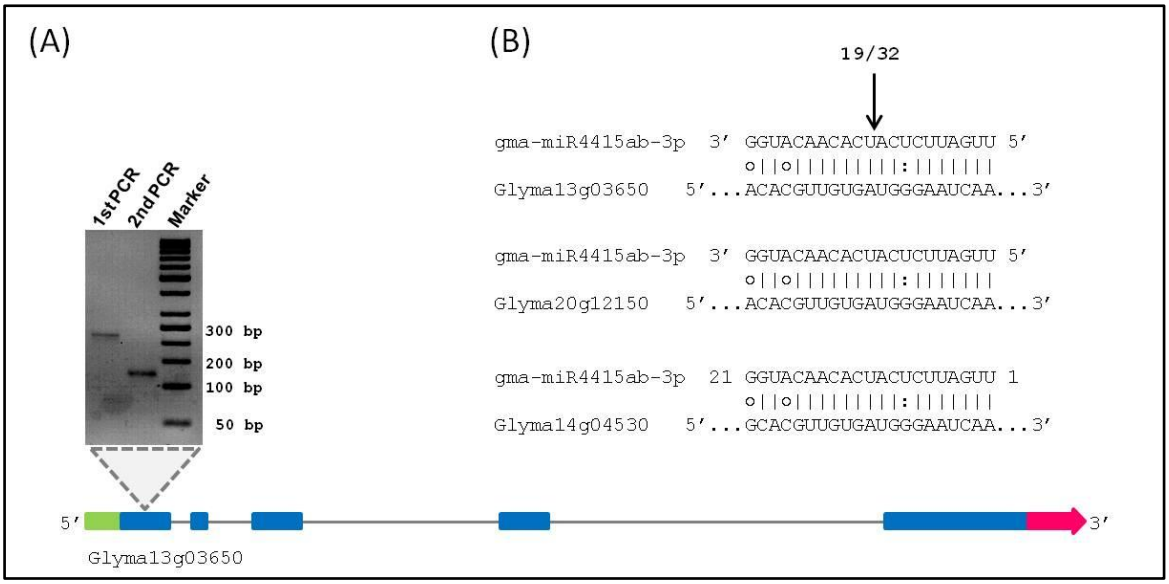


Figure 2

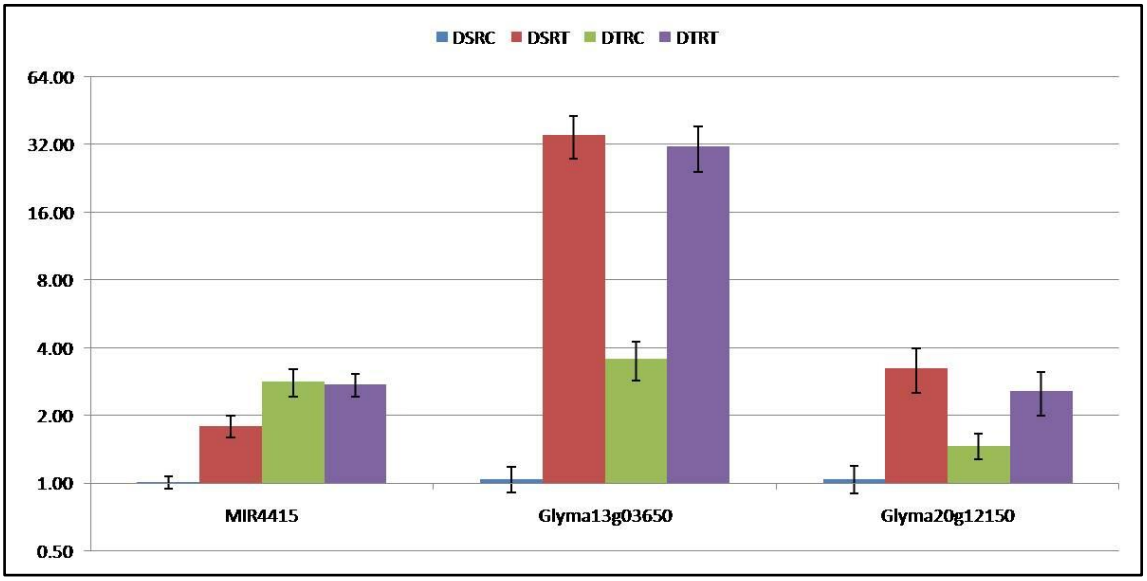
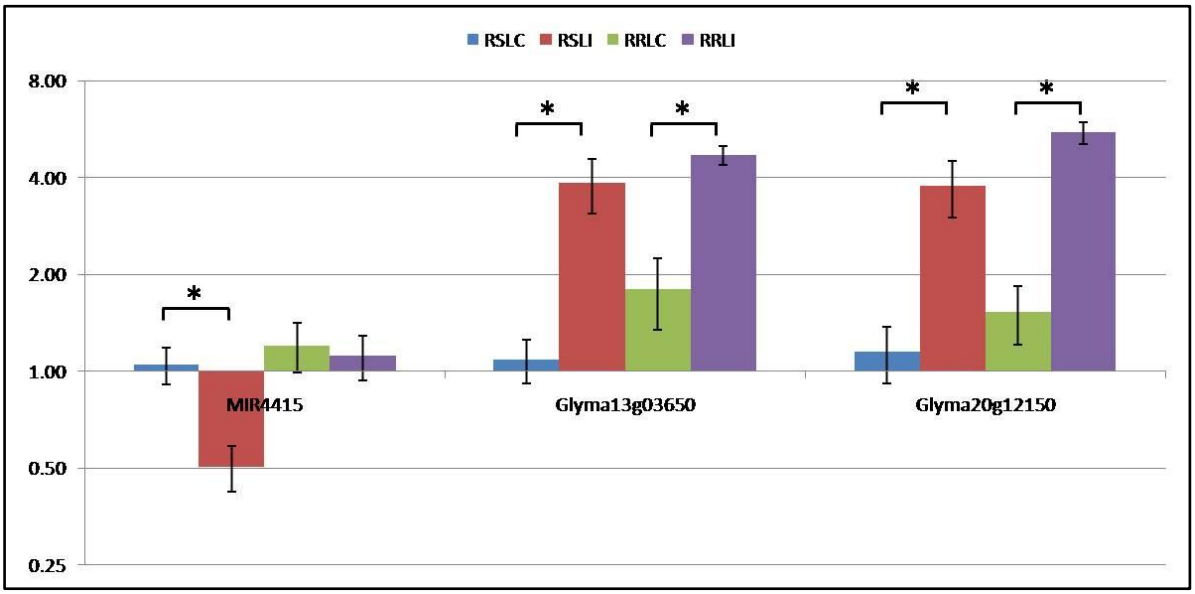


Figure 3



Supplementary file 1. Primer sequences and amplicon characteristics.

Primer ID	Forward Primer Sequence (5' - 3')	Reverse Primer Sequence (5' - 3')	Amplicon Size (bp)	Function
Gene-secific primer	-	TGCACACTGTGAGATGGCAGCA	233	5'RACE 1st PCR
nested primer	-	CCTCGGTGAAAAGCTTGTTGG	130	5'RACE 2nd PCR
Glyma13g03650	TGATCAGAAAGCCAGATTGC	TGAGATGGCAGCAGTACCA	200	RT-qPCR (AO)
Glyma20g12150	AGCACATTTGTCCCTTGGAG	GCCTCTCCTGGGTTTATAGCA	282	RT-qPCR (AO)
Glyma14g04530	GTTGCAACTATCTCTAGGAGCTG	GAAATGGCAGCAGTTCCATC	255	RT-qPCR (AO)
MIR4415ab	TTGATTCTCATCACAACATGG	GTGCAGGGTCCGAGGT	71	RT-qPCR (miRNA)
MIR156b	TGACAGAAGAGAGAGAGCACA	"	71	RT-qPCR (miRNA)
MIR1520d	ATCAGAACATGACACGTGACAA	"	72	RT-qPCR (miRNA)

CAPÍTULO VI

Considerações Finais

Como uma das principais *commodities* e em rápida expansão, a soja é a cultura agrícola brasileira que mais cresceu nas últimas três décadas e corresponde a 49% da área plantada em grãos do país. O aumento da produtividade está associado a inúmeros fatores como manejo e eficiência dos produtores. Porém, além dos fatores tradicionais, deve-se destacar o amplo avanço tecnológico que estudos envolvendo a genética da soja e a implementação de ferramentas biotecnológicas propiciaram para o franco desenvolvimento desta cultura.

Estresses abióticos e bióticos podem reduzir significativamente os rendimentos das lavouras e restringir as latitudes onde espécies comercialmente importantes podem ser cultivadas. As implicações são enormes, uma vez que prejuízos em grandes culturas acabam afetando o mercado econômico e consequentemente a sociedade em geral. Portanto, entender a tolerância e a resistência das plantas à seca e a patógenos respectivamente, devem ser julgadas não só como problemas de ordem genética ou fisiológica, mas também como importante meta de significância econômica e política.

A proposta ao desenvolvimento das atividades de pesquisa que culminaram nesta tese de doutorado foi identificar novos miRNAs de soja e caracterizar a expressão dos mesmos em relação aos processos de deficiência hídrica e FAS. Assim, os resultados obtidos e apresentados na presente tese foram estruturados de forma que assuntos específicos fossem abordados em diferentes artigos, os quais aqui estão disponibilizados nos diferentes capítulos, mas que de uma forma geral não deixaram de

estar conectados, uma vez que, exceto pelo capítulo III, os demais trabalhos foram dependentes daqueles previamente realizados.

Por exemplo, um primeiro aspecto e imprescindível, quando se trabalha com a expressão de genes, é o conhecimento de bons normalizadores. Desta forma o capítulo III relatou a busca e identificação de quais seriam os melhores genes normalizadores a serem utilizados nas análises de RT-qPCR envolvendo miRNAs em soja. Neste estudo foi observada uma clara diferença entre a estabilidade na expressão dos miRNAs e dos genes codificadores de proteínas comumente utilizados como normalizadores em trabalhos de expressão gênica. Utilizando os softwares GeNorm and NormFinder a expressão de miRNAs e genes de mRNA foi analisada em diferentes tecidos vegetais, diferentes genótipos, além de dois estresses distintos, deficiência hídrica e ferrugem asiática. Ao final desta análise foi observado que os miRNAs foram mais estáveis que os genes de mRNAs em todas as condições amostradas. Embora tenha ocorrido alguma discrepância entre os dois softwares, é importante salientar que os miRNAs ocuparam a primeira colocação no *ranking* com relação à maior estabilidade em ambas análises. Este foi o primeiro estudo a avaliar o potencial dos miRNAs como normalizadores em plantas (já haviam sido realizados trabalhos como este em humanos), e abriu portas para explorar uma nova possibilidade para o uso destes genes além das técnicas de silenciamento.

O quarto capítulo desta tese, abordou a identificação de novos miRNAs envolvidos em estresses abiótico e biótico, e relatou um número massivo de novos miRNAs descobertos em soja. O uso do sequenciamento de alto desempenho (Solexa) foi eficiente para permitir a identificação de um total de 256 miRNAs. Destas análises, foram encontradas 24 novas famílias, seis famílias que já haviam sido detectadas em coníferas, embriófitas e magnoliófitas, e 22 famílias que já eram conhecidas em soja.

Entretando, acredita-se que o número de miRNAs detectado poderia ter sido maior, uma vez que os filtros utilizados no processamento dos dados foram bastante restritivos, como em relação ao índice de qualidade dos reads obtidos no sequenciamento, além das buscas no genomas serem restringidas a *matchs* não maiores que cinco vezes. Outro ponto que não foi abordado no artigo publicado, mas que aqui cabe salientar, é que as buscas foram feitas no genoma sequenciado da soja, que pertence à cultivar americana Williams, e nossas bibliotecas foram construídas a partir de três cultivares brasileiras, EMBRAPA48, BR16 e PI561356. A questão aqui levantada é que durante nossas buscas, não foram permitidos *mismatch*es (erros) entre os *reads* e o genoma de referência, e desta forma podemos ter perdido algum polimorfismo que possa existir entre as diferentes cultivares. Esta ideia pode até ser refutada, devido o fato dos miRNAs serem altamente conservados entre diferentes espécies de planta, podendo alegar que seriam ainda mais conservados entre genótipos de uma mesma espécie, o que faz sentido. Entretanto, não há descoberta sem investigação, e este é um dos aspectos que poderia ser modificado ou agregado em novos experimentos que envolvam sequenciamento de última geração para detecção de novos miRNAs.

Outro aspecto importante a ser ressaltado no artigo apresentado no capítulo IV, foi o relato de isomiRNAs encontrados durante a análise das bibliotecas. Foram detectados um total de 121 isomiRNAs. Há algum tempo, nenhuma atenção especial era dada a estas variações da sequência de miRNA maduro, e por anos foi considerado ser resultado de clivagens errôneas da enzima DCL-1, envolvida no processamentos dos miRNAs maduros. Entretanto, atualmente mudou-se a visão sobre estas variantes de miRNAs, os quais se acredita serem preferencialmente expressos em determinadas condições. A prova desta mudança com relação aos isomiRNAs, é que o miRBase, principal banco de depósito de miRNAs, está catalogando a abundância das diferentes

formas de pequenos RNAs derivados de pre-miRNAs, os isomiRNAs, em algumas espécies. Em soja ainda não foram estruturadas estas isoformas com relação aos seus precursores, mas acredito que seja apenas uma questão de tempo até serem organizadas todas estas informações.

Ainda com relação ao capítulo IV, as análises por RT-qPCR de alguns miRNAs com relação à deficiência hídrica e à ferrugem da soja, forneceu novas informações quanto ao padrão de expressão de miRNAs de soja quando exposta a estes estresses. De uma forma geral, observou-se que os miRNAs variaram quanto à resposta ao estresse dependendo do genótipo de soja. Por exemplo, vários miRNAs foram induzidos durante o estresse hídrico em genótipos sensíveis, entretanto durante o mesmo estresse o padrão de expressão foi diferente no genótipo tolerante à seca. Esta mesma variação foi observada nos experimentos com o fungo *P. pachyrhyzy*, onde em geral os miRNAs foram reprimidos durante a infecção no genótipo suscetível, e não variaram entre a condição controle e infectada nas plantas resistentes à ferrugem. Particularmente acredito que este cenário aponte para um aspecto importante na via dos miRNAs: estar submetidos a um controle bastante refinado. Embora muitos aspectos com relação a regulação gênica por miRNAs já tenham sido caracterizados através da identificação e validação de seus alvos, pouco ainda se conhece com relação a regulação dos miRNAs, como promotores, cis-elementos responsáveis por sua transcrição, e ainda fatores e co-fatores envolvidos na biogênese destes pequenos RNAs.

No capítulo V está descrita a busca no entendimento de um miRNA espécie-específico que demonstrou ser diferencialmente expresso entre os diferentes genótipos submetidos à deficiência hídrica e à ferrugem da soja. O MIR4415, o qual foi apenas encontrado em soja até o momento, foi identificado no trabalho apresentado ainda no capítulo IV e sua validação via RT-qPCR mostrou claramente que este miRNA era

induzido ou reprimido conforme a condição de estresse em que os particulares genótipos se encontravam. Este comportamento, despertou o interesse de ir além e investigar o possível alvo do mesmo. Pela predição computacional chegou-se ao mRNA de uma enzima: a ascorbato oxidase (AO). O passo seguinte foi investigar se apenas um transcrito seria o alvo deste miRNA. Ao realizar-se BLAST no genoma da soja (disponível do banco de dados Phytozome), descobriu-se então que havia três transcritos de AO que apresentavam além do mesmo sítio de ligação do miRNA, uma sequência bastante conservada ao longo de todo transcrito. Este fator, dificultou muito na elaboração dos oligonucleotídeos que deveriam ser utilizados nas técnicas para validação do alvo. No caso do 5' RACE não houve como projetar primers diferentes, e assim os dois primers necessários para os distintos PCRs hibridizavam em uma região conservada nos três transcritos. Mesmo tendo a capacidade de hibridizar em todos os candidatos alvos, o sequenciamento dos clones demonstraram que apenas o transcrito Glyma13g03650 era clivado pelo MIR4415 (pelo menos na condição em que foram extraídas as amostras para o 5' RACE). Os resultados da clivagem foram bastante consistentes, um grande número de clones apresentou o mesmo sítio sendo clivado pelo miRNA. Mesmo com a comprovação de apenas Glyma13g03650 via 5'RACE, foram realizados os RT-qPCR para os demais transcritos. Glyma20g12150 apresentou um padrão de expressão semelhante à Glyma13g03650 em todas a condições testadas, entretando Glyma14g04530 falhou na amplificação. Como dito anteriormente, já era conhecida a expressão diferenciada de MIR4415 nos diferentes genótipos sob estresse hídrico e FAS e, por este motivo, reproduziu-se a investigação deste miRNA mais Glyma13g03650 e Glyma20g12150 sob as mesmas condições.

Durante o ensaio de déficit hídrico, não se observou qualquer correlação inversa entre o miRNA e ambos transcritos de AO (o que é esperado em padrões de expressão

de alvo e miRNA). Entretanto durante as análises de RT-qPCR com infecção pelo fungo *P. pachyrhizi*, foi observado o padrão de expressão típico de um miRNA e seu alvo. O MIR4415 foi reprimido durante a infecção pelo patógeno no genótipo suscetível, enquanto ambos os alvos, Glyma13g03650 e Glyma20g12150, tiveram uma indução na sua expressão. Entretanto esta mesma relação não foi observada no genótipo resistente, o que leva a pensar-se que a interação MIR4415 e AO-mRNA possa ser bastante específica, podendo até ser relacionada à suscetibilidade de cultivares de soja. Não existem muitas informações com relação à atividade de AO em relação à ferrugem da folha. Como relatado no capítulo V, muitos estudo já foram realizados com esta enzima, mas ainda não está totalmente claro seu papel nos casos de estresses em plantas. Tentando traçar um raciocínio lógico na relação AO *versus* ataque fúngico, pode-se direcionar o pensamento ao ácido ascórbico (AA). Sabe-se que o nível de AA é diminuído quando AO é induzido, sendo AA uma barreira contra o ataque de patógenos, plantas suscetíveis com redução no acúmulo de AA no apoplasto poderiam ter como consequência o estabelecimento do fungo sobre o tecido foliar facilitado. E por que esse cenário não seria visto no genótipo resistente? Uma possibilidade, seria devido a presença dos genes maiores de resistência (genes *R*), que contam com uma resposta específica e diferenciada à infecção pelo fungo. Antes de finalizar os comentários, é importante esclarecer, que quando afirma-se da possibilidade de interação MIR4415-AO ser um dos fatores envolvidos na resposta de suscetibilidade, não se quer de forma alguma simplificar a ampla complexidade existente nas interações planta-patógeno. Entretanto, este é mais um mecanismo a ser observado com maior atenção na busca de novas vias e fatores que colaborem com os estudos de resistência e suscetibilidade em plantas.

Em suma, espera-se que as descobertas que foram relatadas nesta Tese, bem como as perguntas que ainda permanecem em alguns pontos, sejam de alguma forma fonte para novas pesquisas que estão por vir. Espero que os resultados aqui apresentados possam colaborar com futuros estudos que, como este, são fontes de formação para pesquisadores brasileiros.

Bibliografia (CAPÍTULO I)

Addo-Quaye, C., Eshoo, T.W., Bartel, D.P., and Axtell, M.J. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. Curr Biol *18*, 758-762.

Addo-Quaye, C., Miller, W., and Axtell, M.J. (2009). CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. Bioinformatics *25*, 130-131.

Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell *121*, 207-221.

Arenas-Huertero, C., Perez, B., Rabanal, F., Blanco-Melo, D., De la Rosa, C., Estrada-Navarrete, G., Sanchez, F., Covarrubias, A.A., and Reyes, J.L. (2009). Conserved and novel miRNAs in the legume Phaseolus vulgaris in response to stress. Plant Mol Biol *70*, 385-401.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. Cell *136*, 215-233.

Bewley, J.D., and Black, M. (1994). Seeds: physiology of development and germination (New York, Plenum).

Bollman, K.M., Aukerman, M.J., Park, M.Y., Hunter, C., Berardini, T.Z., and Poethig, R.S. (2003). HASTY, the *Arabidopsis* ortholog of exportin 5/MSN5, regulates phase change and morphogenesis. Development *130*, 1493-1504.

Bonnet, E., Wuyts, J., Rouze, P., and Van de Peer, Y. (2004). Detection of 91 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. Proc Natl Acad Sci U S A *101*, 11511-11516.

Bromfield, K.R., and Hartwig, E.E. (1980). Resistance to soybean rust and mode of inheritance. Crop Science *20*, 254-255.

Cai, X., Davis, E.J., Ballif, J., Liang, M., Bushman, E., Haroldsen, V., Torabinejad, J., and Wu, Y. (2006). Mutant identification and characterization of the laccase gene family in *Arabidopsis*. J Exp Bot *57*, 2563-2569.

Cannon, S.B., and Shoemaker, R.C. (2012). Evolutionary and comparative analyses of the soybean genome. Breed Sci *61*, 437-444.

Casagrande, E.C., Farias, J.R.B., Neumaier, N., Oya, T., Pedroso, J., Martins, P.K., Breton, M.C., and Nepomuceno, A.L. (2001). Expressão gênica diferencial durante déficit hídrico em soja. R Bras Fisiol Veg *13*, 168-184.

Chen, R., Hu, Z., and Zhang, H. (2009). Identification of microRNAs in wild soybean (Glycine soja). J Integr Plant Biol *51*, 1071-1079.

Chen, X. (2005). MicroRNA biogenesis and function in plants. FEBS Lett *579*, 5923-5931.

Cheng, Y.W., and Chan, K.L. (1968). The breeding of rust resistant soybean Tainung 3. J Taiwan Agr Res *17*, 30-34.

Choi, J.J., Alkharouf, N.W., Schneider, K.T., Matthews, B.F., and Frederick, R.D. (2008). Expression patterns in soybean resistant to *Phakopsora pachyrhizi* reveal the importance of peroxidases and lipoxygenases. Funct Integr Genomics *8*, 341-359.

Chuck, G., Candela, H., and Hake, S. (2009). Big impacts by small RNAs in plant development. Curr Opin Plant Biol *12*, 81-86.

Clement, M., Lambert, A., Herouart, D., and Boncompagni, E. (2008). Identification of new up-regulated genes under drought stress in soybean nodules. Gene *426*, 15-22.

Clemente, T.E., and Cahoon, E.B. (2009). Soybean oil: genetic approaches for modification of functionality and total content. Plant Physiol *151*, 1030-1040.

CONAB Companhia Nacional de Abastecimento. Disponível em: <http://www.conab.gov.br>. Acesso em: dez 2012.

Crespi, M., and Frugier, F. (2008). De novo organ formation from differentiated cells: root nodule organogenesis. Sci Signal *1*, re11.

Dai, X., and Zhao, P.X. (2011). psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res *39*, W155-159.

Dalmay, T., Horsefield, R., Braunstein, T.H., and Baulcombe, D.C. (2001). *SDE3* encodes an RNA helicase required for post-transcriptional gene silencing in *Arabidopsis*. EMBO J *20*, 2069-2078.

Desclaux, D., Huynh, T.T., and Roumet, P. (2000). Identification of soybean plant characteristics that indicate the timing of drought stress. Crop Science *40*, 716-722.

Desclaux, D., and Roumet, P. (1996). Impact of drought stress on the phenology of two soybean cultivars. Field Crop Res *46*, 61-70.

Dey, P.M., and Harbone, J.B. (1997). Plant Biochemistry (San Diego, Academic Press).

EMBRAPA (2008). Tecnologias de produção de soja – região central do Brasil – 2009 e 2010. *In* Sistemas de Produção (Londrina, Embrapa Soja), pp. 262.

EMBRAPA (2011). Tecnologias de produção de soja – região central do Brasil 2012 e 2013. In Sistemas de Produção / Embrapa Soja (Londrina, Embrapa Soja), pp. 261.

EMBRAPA. Disponível em: <http://www.embrapa.br> . Acesso em: dez 2012.

Fagard, M., Boutet, S., Morel, J.B., Bellini, C., and Vaucheret, H. (2000). AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals. Proc Natl Acad Sci U S A *97*, 11650-11654.

Faller, M., and Guo, F. (2008). MicroRNA biogenesis: there's more than one way to skin a cat. Biochim Biophys Acta *1779*, 663-667.

Fang, Y., and Spector, D.L. (2007). Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living *Arabidopsis* plants. Curr Biol *17*, 818-823.

FAO. Disponível em: <www.fao.org>. Acesso em: dez 2012.

Fehr, W.R., and Caviness, C.E. (1977). Stages of soybean development, Vol 80 (Ames, State University of Science and Technology).

Franco-Zorrilla, J.M., Valli, A., Todesco, M., Mateos, I., Puga, M.I., Rubio-Somoza, I., Leyva, A., Weigel, D., Garcia, J.A., and Paz-Ares, J. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. Nat Genet *39*, 1033-1037.

Furtado, G.Q., Alves, A.M., Carneiro, L.C., Godoy, C.V., and Massola, N.S. (2009). Influência do estágio fenológico e da idade dos trifólios de soja na infecção de *Phakopsora pachyrhizi*. Tropical Plant Pathol *34*, 118-122.

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

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

Griffiths-Jones, S. (2004). The microRNA Registry. Nucleic Acids Res *32*, D109-111.

Griffiths-Jones, S. (2006). miRBase: the microRNA sequence database. Methods Mol Biol *342*, 129-138.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., and Enright, A.J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res *34*, D140-144.

Griffiths-Jones, S., Saini, H.K., van Dongen, S., and Enright, A.J. (2008). miRBase: tools for microRNA genomics. Nucleic Acids Res *36*, D154-158.

Guimaraes-Dias, F., Neves-Borges, A.C., Viana, A.A., Mesquita, R.O., Romano, E., de Fatima Grossi-de-Sa, M., Nepomuceno, A.L., Loureiro, M.E., and Alves-Ferreira, M. (2012). Expression analysis in response to drought stress in soybean: Shedding light on the regulation of metabolic pathway genes. Genet Mol Biol *35*, 222-232.

Guo, L., and Lu, Z. Global expression analysis of miRNA gene cluster and family based on isomiRs from deep sequencing data. Comput Biol Chem *34*, 165-171.

Guo, L., and Lu, Z. (2009). Global expression analysis of miRNA gene cluster and family based on isomiRs from deep sequencing data. Comput Biol Chem *34*, 165-171.

Guo, N., Ye, W.W., Wu, X.L., Shen, D.Y., Wang, Y.C., Xing, H., and Dou, D.L. (2011). Microarray profiling reveals microRNAs involving soybean resistance to *Phytophthora sojae*. Genome *54*, 954-958.

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

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

Hu, Z., Jiang, Q., Ni, Z., Chen, R., Xu, S., and Zhang, H. (2012). Analyses of a *Glycine max* degradome library identify microRNA targets and microRNAs that trigger secondary SiRNA biogenesis. J Integr Plant Biol.

Hymowitz, T. (1970). On the domestication of the soybean. Econ Bot *24*, 408-421.
Jones-Rhoades, M.W., and Bartel, D.P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell *14*, 787-799.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. (2006). MicroRNAS and their regulatory roles in plants. Annu Rev Plant Biol *57*, 19-53.

Joshi, T., Yan, Z., Libault, M., Jeong, D.H., Park, S., Green, P.J., Sherrier, D.J., Farmer, A., May, G., Meyers, B.C., *et al.* Prediction of novel miRNAs and associated target genes in *Glycine max*. BMC Bioinformatics *11 Suppl 1*, S14.

Joshi, T., Yan, Z., Libault, M., Jeong, D.H., Park, S., Green, P.J., Sherrier, D.J., Farmer, A., May, G., Meyers, B.C., *et al.* (2010). Prediction of novel miRNAs and associated target genes in *Glycine max*. BMC Bioinformatics *11 Suppl 1*, S14.

Kantar, M., Lucas, S.J., and Budak, H. (2010). miRNA expression patterns of *Triticum dicoccoides* in response to shock drought stress. Planta *233*, 471-484.

Karam, F., Masaad, R., Sfeir, T., Mounzer, O., and Rouphael, Y. (2005). Evapotranspiration and seed yield of field grown soybean under deficit irrigation conditions. Agr Water Manag *75*, 226-244.

Katiyar-Agarwal, S., and Jin, H. Role of small RNAs in host-microbe interactions. Annu Rev Phytopathol *48*, 225-246.

Kawashima, C.G., Yoshimoto, N., Maruyama-Nakashita, A., Tsuchiya, Y.N., Saito, K., Takahashi, H., and Dalmay, T. (2009). Sulphur starvation induces the expression of microRNA-395 and one of its target genes but in different cell types. Plant J *57*, 313-321.

Kim, S., Yang, J.Y., Xu, J., Jang, I.C., Prigge, M.J., and Chua, N.H. (2008). Two cap-binding proteins CBP20 and CBP80 are involved in processing primary MicroRNAs. Plant Cell Physiol *49*, 1634-1644.

Kozomara, A., and Griffiths-Jones, S. (2010). miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res *39*, D152-157.

Kulcheski, F.R., de Oliveira, L.F.V., Molina, L.G., Almerão, M.P., Rodrigues, F.A., Marcolino, J., Barbosa, B.J., Stolf-Moreira, R., Nepomuceno, A.L., Marcelino-Guimarães, F.C., *et al.* (2011). Identification of novel soybean microRNAs involved in abiotic and biotic stresses. BMC Genomics *12*, 307-322.

Kurihara, Y., Takashi, Y., and Watanabe, Y. (2006). The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. RNA *12*, 206-212.

Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. Science *294*, 862-864.

Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., and Kim, V.N. (2004). MicroRNA genes are transcribed by RNA polymerase II. EMBO J *23*, 4051-4060.

Lelandais-Briere, C., Sorin, C., Declerck, M., Benslimane, A., Crespi, M., and Hartmann, C. Small RNA diversity in plants and its impact in development. Curr Genomics *11*, 14-23.

Lelandais-Briere, C., Sorin, C., Declerck, M., Benslimane, A., Crespi, M., and Hartmann, C. (2010). Small RNA diversity in plants and its impact in development. Curr Genomics *11*, 14-23.

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15-20.

Li, H., Dong, Y., Yin, H., Wang, N., Yang, J., Liu, X., Wang, Y., Wu, J., and Li, X. (2011). Characterization of the stress associated microRNAs in *Glycine max* by deep sequencing. BMC Plant Biol *11*, 170.

Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. Curr Biol *15*, 1501-1507.

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

Li, X., Wang, X., Zhang, S., Liu, D., Duan, Y., and Dong, W. (2012b). Identification of soybean microRNAs involved in soybean cyst nematode infection by deep sequencing. PLoS One *7*, e39650.

Liang, G., Yang, F., and Yu, D. (2010). MicroRNA395 mediates regulation of sulfate accumulation and allocation in *Arabidopsis thaliana*. Plant J *62*, 1046-1057.

Liu, F., Andersen, M.N., and Jensen, C.R. (2003). Loss of pod set caused by drought stress is associated with water status and ABA content of reproductive structures in soybean. Funct Plant Biol *30* 271-280.

Liu, H.H., Tian, X., Li, Y.J., Wu, C.A., and Zheng, C.C. (2008). Microarray-based analysis of stress-regulated microRNAs in *Arabidopsis thaliana*. RNA *14*, 836-843.

Liu, Q., and Chen, Y.-Q. (2009). Insights into the mechanism of plant development: interactions of miRNAs pathway with phytormone response. Biochem Biophys Res Commun *384*, 1-5.

Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. Science *297*, 2053-2056.

Lu, C., and Fedoroff, N. (2000). A mutation in the *Arabidopsis HYL1* gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin. Plant Cell *12*, 2351-2366.

Lu, S., Sun, Y.H., and Chiang, V.L. (2008a). Stress-responsive microRNAs in *Populus*. Plant J *55*, 131-151.

Lu, X.Y., and Huang, X.L. (2008). Plant miRNAs and abiotic stress responses. Biochem Biophys Res Commun *368*, 458-462.

Lu, Y.D., Gan, Q.H., Chi, X.Y., and Qin, S. (2008b). Roles of microRNA in plant defense and virus offense interaction. Plant Cell Rep *27*, 1571-1579.

Mallory, A.C., and Vaucheret, H. (2006). Functions of microRNAs and related small RNAs in plants. Nat Genet *38 Suppl*, S31-36.

Manavella, P.A., Hagmann, J., Ott, F., Laubinger, S., Franz, M., Macek, B., and Weigel, D. Fast-Forward Genetics Identifies Plant CPL Phosphatases as Regulators of miRNA Processing Factor HYL1. Cell *151*, 859-870.

Markmann, K., and Parniske, M. (2009). Evolution of root endosymbiosis with bacteria: How novel are nodules? Trends Plant Sci *14*, 77-86.

Martin, R.C., Liu, P.P., Goloviznina, N.A., and Nonogaki, H. (2010). microRNA, seeds, and Darwin?: diverse function of miRNA in seed biology and plant responses to stress. J Exp Bot *61*, 2229-2234.

Martins, P.K., Jordão, B.Q., Yamanaka, N., Farias, J.R.B., Beneventi, M.A., Binneck, E., Fuganti, R., Stolf, R., and Nepomuceno, A.L. (2008). Differential gene expression and mitotic cell analysis of the drought tolerant soybean (*Glycine max* L. Merrill Fabales, Fabaceae) cultivar MG/BR46 (Conquista) under two water deficit induction systems. Genet Mol Biol *31*, 512-521.

Meyer, J.D., Silva, D.C., Yang, C., Pedley, K.F., Zhang, C., van de Mortel, M., Hill, J.H., Shoemaker, R.C., Abdelnoor, R.V., Whitham, S.A., *et al.* (2009). Identification and analyses of candidate genes for rpp4-mediated resistance to Asian soybean rust in soybean. Plant Physiol *150*, 295-307.

Meyers, B.C., Axtell, M.J., Bartel, B., Bartel, D.P., Baulcombe, D., Bowman, J.L., Cao, X., Carrington, J.C., Chen, X., Green, P.J., *et al.* (2008). Criteria for annotation of plant MicroRNAs. Plant Cell *20*, 3186-3190.

Ministério da Agricultura. Disponível em:<www.agricultura.gov.br>. Acesso em: dez 2012.

Morceli, T.G.S., Uneda-Trevisoli, S.H., Morceli Junior, A.A., Kiihl, R.A., Calvo, É.S., Di Mauro, A.O., and Garcia, A. (2008). Identificação e validação de marcadores microssatélites ligados ao gene *Rpp5* de resistência à ferrugem-asiática-da-soja. Pesq Agropec Bras *43*, 1533-1541.

Mourrain, P., Beclin, C., and Vaucheret, H. (2000). Are gene silencing mutants good tools for reliable transgene expression or reliable silencing of endogenous genes in plants? Genet Eng (N Y) *22*, 155-170.

Naqvi, A.R., Sarwat, M., Hasan, S., and Roychodhury, N. (2012). Biogenesis, functions and fate of plant microRNAs. J Cell Physiol *227*, 3163-3168.

Neumaier, N., Nepomuceno, A.L., and Farias, J.R.B. (2000). Estádios de desenvolvimento da culturade soja. In Estresses em soja, E.R. Bonato, ed. (Passo Fundo, EMBRAPA-CNPT), pp. 19-44.

Ni, Z., Hu, Z., Jiang, Q., and Zhang, H. (2012). Overexpression of gma-MIR394a confers tolerance to drought in transgenic *Arabidopsis thaliana*. Biochem Biophys Res Commun *427*, 330-335.

Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. Curr Biol *12*, 1484-1495.

Radwan, O., Liu, Y., and Clough, S.J. (2011). Transcriptional analysis of soybean root response to *Fusarium virguliforme*, the causal agent of sudden death syndrome. Mol Plant Microbe Interact *24*, 958-972.

Ramachandran, V., and Chen, X. (2008). Degradation of microRNAs by a family of exoribonucleases in *Arabidopsis*. Science *321*, 1490-1492.
Ramanjulu, S., and Bartels, D. (2002). Drought- and desiccation-induced modulation of gene expression in plants. Plant Cell Environ *25*, 141-151.

Rayapati, P.J., and Stewart, C.R. (1991). Solubilization of a Proline Dehydrogenase from maize (*Zea mays* L.) Mitochondria. Plant Physiol *95*, 787-791.

Reinhart, B.J., and Bartel, D.P. (2002). Small RNAs correspond to centromere heterochromatic repeats. Science *297*, 1831.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., *et al.* (2010). Genome sequence of the palaeopolyploid soybean. Nature *463*, 178-183.

Shamimuzzaman, M., and Vodkin, L. (2012). Identification of soybean seed developmental stage-specific and tissue-specific miRNA targets by degradome sequencing. BMC Genomics *13*, 310.

Shukla, L.I., Chinnusamy, V., and Sunkar, R. (2008). The role of microRNAs and other endogenous small RNAs in plant stress responses. Biochim Biophys Acta *1779*, 743-748.

Sinclair, J.B., and Hartman, G.L. (1999). Soybean rust. In Compendium of soybean diseases, e.a. Hartman, ed. (St. Paul, American Phytopathological Society), pp. 25-26.

CISoja - Centro de inteligência da soja. Disponível em:<www.cisoja.com.br>. Acesso em: dez 2012.

Song, L., Han, M.H., Lesicka, J., and Fedoroff, N. (2007). *Arabidopsis* primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. Proc Natl Acad Sci U S A *104*, 5437-5442.

Song, Q.X., Liu, Y.F., Hu, X.Y., Zhang, W.K., Ma, B., Chen, S.Y., and Zhang, J.S. (2011). Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. BMC Plant Biol *11*, 5.

Subramanian, S., Fu, Y., Sunkar, R., Barbazuk, W.B., Zhu, J.K., and Yu, O. (2008). Novel and nodulation-regulated microRNAs in soybean roots. BMC Genomics *9*, 160.

Sunkar, R., and Zhu, J.K. (2004). Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. Plant Cell *16*, 2001-2019.

Taji, T., Ohsumi, C., Iuchi, S., Seki, M., Kasuga, M., Kobayashi, M., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2002). Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana*. Plant J *29*, 417-426.

Trindade, I., Capitao, C., Dalmay, T., Fevereiro, M.P., and Santos, D.M. (2010). miR398 and miR408 are up-regulated in response to water deficit in *Medicago truncatula*. Planta *231*, 705-716.

Turner, M., Yu, O., and Subramanian, S. (2012). Genome organization and characteristics of soybean microRNAs. BMC Genomics *13*, 169.

van de Mortel, M., Recknor, J.C., Graham, M.A., Nettleton, D., Dittman, J.D., Nelson, R.T., Godoy, C.V., Abdelnoor, R.V., Almeida, A.M., Baum, T.J., *et al.* (2007). Distinct biphasic mRNA changes in response to Asian soybean rust infection. Mol Plant Microbe Interact *20*, 887-899.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P., and Crete, P. (2004). Endogenous trans-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. Mol Cell *16*, 69-79.

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. Cell *136*, 669-687.

Wang, J.F., Zhou, H., Chen, Y.Q., Luo, Q.J., and Qu, L.H. (2004a). Identification of 20 microRNAs from *Oryza sativa*. Nucleic Acids Res *32*, 1688-1695.

Wang, X.J., Reyes, J.L., Chua, N.H., and Gaasterland, T. (2004b). Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. Genome Biol *5*, R65.

Wang, Y., Li, P., Cao, X., Wang, X., Zhang, A., and Li, X. (2009). Identification and expression analysis of miRNAs from nitrogen-fixing soybean nodules. Biochem Biophys Res Commun *378*, 799-803.

Wei, L., Zhang, D., Xiang, F., and Zhang, Z. (2009). Differentially expressed miRNAs potentially involved in the regulation of defense mechanism to drought stress in maize seedlings. Int J Plant Sci *170*, 979–989.

Xie, Z., Kasschau, K.D., and Carrington, J.C. (2003). Negative feedback regulation of Dicer-Like1 in *Arabidopsis* by microRNA-guided mRNA degradation. Curr Biol *13*, 784-789.

Yamaguchi-Shinozaki, K., and Shinozaki, K. (1994). A novel cis-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress. Plant Cell *6*, 251-264.

Yang, T., Xue, L., and An, L. (2007). Functional diversity of miRNA in plants. Plant Science *172*, 423-432.

Yorinori, J.T., Paiva, W.M., Frederick, R.D., Costa Milan, L.M., Bertagnolli, P.F., Hartman, G.L., Godoy, C.V., and Nunes Junior, J. (2005). Epidemics of soybean rust

(*Phakopsora pachyrhizi*) in Brazil and Paraguay from 2001 to 2003. Plant Disease *89*, 675-677.
Yu, B., Bi, L., Zheng, B., Ji, L., Chevalier, D., Agarwal, M., Ramachandran, V., Li, W., Lagrange, T., Walker, J.C., *et al.* (2008). The FHA domain proteins DAWDLE in *Arabidopsis* and SNIP1 in humans act in small RNA biogenesis. Proc Natl Acad Sci U S A *105*, 10073-10078.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X. (2005). Methylation as a crucial step in plant microRNA biogenesis. Science *307*, 932-935.

Zabala, G., Campos, E., Varala, K.K., Bloomfield, S., Jones, S.I., Win, H., Tuteja, J.H., Calla, B., Clough, S.J., Hudson, M., *et al.* (2012). Divergent patterns of endogenous small RNA populations from seed and vegetative tissues of *Glycine max*. BMC Plant Biol *12*, 177.

Zeng, Q.Y., Yang, C.Y., Ma, Q.B., Li, X.P., Dong, W.W., and Nian, H. (2012). Identification of wild soybean miRNAs and their target genes responsive to aluminum stress. BMC Plant Biol *12*, 182.

Zhang, B., Pan, X., Cannon, C.H., Cobb, G.P., and Anderson, T.A. (2006). Conservation and divergence of plant microRNA genes. Plant J *46*, 243-259.

Zhang, B., Pan, X., and Stellwag, E.J. (2008). Identification of soybean microRNAs and their targets. Planta *229*, 161-182.

Zhang, X., Zou, Z., Gong, P., Zhang, J., Ziaf, K., Li, H., Xiao, F., and Ye, Z. (2011). Over-expression of microRNA169 confers enhanced drought tolerance to tomato. Biotechnol Lett *33*, 403-409.

Zhao, B., Liang, R., Ge, L., Li, W., Xiao, H., Lin, H., Ruan, K., and Jin, Y. (2007). Identification of drought-induced microRNAs in rice. Biochem Biophys Res Commun *354*, 585-590.

Zhou, L., Liu, Y., Liu, Z., Kong, D., Duan, M., and Luo, L. (2010). Genome-wide identification and analysis of drought-responsive microRNAs in *Oryza sativa*. J Exp Bot. *61*, 4157-4168.

Anexo

Anexo 1 - Nomenclatura segundo o banco de dados miRBase para os novos miRNAs descritos no capítulo II

Código - miRBase	Código - artigo	Sequência do miRNA maduro
gma-MIR5368	gma-MIR-Seq01	GGACAGUCUCAGGUAGACA
gma-MIR5369	gma-MIR-Seq03	UGAGAAAAGGAGGAUGUCA
gma-MIR826a	gma-MIR-Seq04a	GCUGGAUGUCUUUGAAGGA
gma-MIR826b	gma-MIR-Seq04b	GCUGGAUGUCUUUGAAGGA
gma-MIR5037b	gma-MIR-Seq05	AACCCUCAAAGGCUUCCUAG
gma-MIR5037c	gma-MIR-Seq06	AGUGGAACUUUGAGGCCUGC
não classificado	gma-MIR-Seq07	AAAUGACUUGAGAGGUGUAG
gma-MIR5370	gma-MIR-Seq08	CUAAAGAUUGUCCAAAAGGAA
gma-MIR5044	gma-MIR-Seq09	GUAGUGGAUGCCUAGAGGUCC
gma-MIR5371-5p	gma-MIR-Seq10-5p	UAGGAAUUAGUCACUCAGAUC
gma-MIR5371-3p	gma-MIR-Seq10-3p	AUCUCAGUGACUAAUUUCUAG
gma-MIR5372	gma-MIR-Seq11	UUGUUCGAUAAAACUGUUGUG
gma-MIR5373	gma-MIR-Seq12	UCUCUUGAUUCUAGAUGAUGU
gma-MIR4412-5p	gma-MIR-Seq13	UGUUGCGGGUAUCUUUGCCUC
gma-MIR5038a	gma-MIR-Seq14a	UGAGAAUUUGGCCUCUGUCCA
gma-MIR5038b	gma-MIR-Seq14b	UGAGAAUUUGGCCUCUGUCCA
gma-MIR403a	gma-MIR-Seq15a	UUAGAUUCACGCACAAACUUG
gma-MIR403b	gma-MIR-Seq15b	UUAGAUUCACGCACAAACUUG
gma-MIR5374	gma-MIR-Seq16	UUAUAGUCUGACAUCUGGAAU
gma-MIR5375	gma-MIR-Seq17	ACUAUAGAAGUACUUGUGGAGC
gma-MIR1507c	gma-MIR-Seq18	CCUCAUUCCAAACAUCAUCUAA
gma-MIR5376	gma-MIR-Seq19	UGAAGAUUUGAAGAAUUUGGGA
gma-MIR4397-5p	gma-MIR-Seq20	CAUCGUUGACGCUGACUGUACG
gma-MIR5377	gma-MIR-Seq21	CUGAAGGAUCGAUGUAGAAUGCU
gma-MIR5378	gma-MIR-Seq22	CAUCUGAAGGAUAGAACACAUA
gma-MIR4998	gma-MIR-Seq23	AGUUUCGUGACUACAACUUCUGAA
gma-MIR5379	gma-MIR-Seq24	AUGAAAAUCAUUCAUUAUGAUAUC
gma-MIR5380a	gma-MIR-Seq25a	GAAAAUGAAUGAUGAGGAUGGGGA

Anexo 2

Franceli Rodrigues Kulcheski
Curriculum Vitae

Dados pessoais

Nome Franceli Rodrigues Kulcheski
Filiação Sílvio Miguel Kulcheski e Ivone Rodrigues Kulcheski
Nascimento 14/08/1978 - Bento Gonçalves/RS - Brasil
Carteira de Identidade 4056237987 SJS - RS - 19/07/2000

Formação acadêmica/titulação

2009-2013 Doutorado em Biologia Celular e Molecular.
Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre, Brasil
Com período sanduíche no Max Planck Institute for Developmental Biology (Orientador : Detlef Weigel)
Título: Identificação e análise de expressão de miRNAs em soja sob extresse biótico e abiótico
Orientador: Rogério Margis
Bolsista do(a): Conselho Nacional de Desenvolvimento Científico e Tecnológico

2005 - 2007 Mestrado em Fitotecnia.
Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre, Brasil
Título: Potencial da resistência genética à ferrugem da folha em aveia para o controle da moléstia no Sul do Brasil, Ano de obtenção: 2007
Orientador: Carla Andrea Delatorre
Bolsista do(a): Conselho Nacional de Desenvolvimento Científico e Tecnológico

2000 - 2005 Graduação em Ciências Biológicas.
Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre, Brasil
Título: Filogenia molecular do Gênero Petunia Juss. (Solanaceae)
Orientador: Loreta Brandão de Freitas
Bolsista do(a): Conselho Nacional de Desenvolvimento Científico e Tecnológico

---

Formação complementar

2009 - 2009 Curso de curta duração em Bioinformática: Genômica e Transcriptômica.
Embrapa Soja, EMBRAPA SOJA, Brasil

2007 - 2007 Curso de curta duração em Uso do Software Genes.
Sociedade Brasileira de Melhoramento de Plantas, SBMP, Campos Dos Goytacazes, Brasil

2004 - 2004 Extensão universitária em Genômica Funcional de Plantas.
Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre, Brasil

2003 - 2003 Extensão universitária em Filogenias Moleculares e Conservação da Biodiversidade
Sociedade Brasileira de Genética, SBG(BR), Brasil

1999 - 1999 Extensão universitária em Biologia Genética e Comportamento de Abelhas.
Universidade Estadual de Campinas, UNICAMP, Campinas, Brasil

1997 - 1997 Extensão universitária em Capacitação de Orientadores do Telecurso 2000.
Serviço Nacional de Aprendizagem Industrial, SENAI, Brasil

Atuação profissional

1. Diversity Arrays Technology Pty Ltd - DART

Vínculo institucional

2008 - 2008 Vínculo: Research Technical Officer , Enquadramento funcional: Research Technical Officer , Carga horária: 40, Regime: Dedicação exclusiva

2. Universidade Federal do Rio Grande do Sul - UFRGS

Vínculo institucional

2005 - 2007 Vínculo: Bolsista , Enquadramento funcional: Mestrado , Carga horária: 40, Regime: Dedicação exclusiva

2002 - 2005 Vínculo: Bolsista , Enquadramento funcional: Iniciação Científica , Carga horária: 20, Regime: Dedicação exclusiva

3. Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA

Vínculo institucional

1998 - 2000 Vínculo: Bolsista , Enquadramento funcional: Iniciação Científica , Carga horária: 20, Regime: Dedicação exclusiva

4. Serviço Nacional de Aprendizagem Industrial - SENAI

Vínculo institucional

1997 - 1998 Vínculo: Orientador TeleCurso 2000 , Enquadramento funcional: Prestação de Serviço , Carga horária: 15, Regime: Parcial

---

Prêmios e títulos

2011 Premiação por publicação de artigo científico, Programa de Pós-Graduação em Biologia Celular e Molecular - UFRGS

2010 Premiação por publicação de artigo científico, Programa de Pós-Graduação em Biologia Celular e Molecular - UFRGS

2009 Prêmio pós-graduação - Menção honrosa para apresentação oral, Sociedade Brasileira de Genética

2006 Prêmio Jovem Cientista, Comissão Brasileira de Pesquisa de Aveia

2004 Destaque XVI Salão de Iniciação Científica - Sessão de Genética Vegetal, UFRGS

2004 Prêmio Iniciação Científica- Menção Honrosa para Painel, Sociedade Brasileira de Genética

---

Produção bibliográfica

Artigos completos publicados em periódicos

1. Molina, L. G., Da Fonseca, G. C., de Morais, G. L., de Oliveira, L. F. V., Carvalho, J. B. de, **Kulcheski**, **F. R**., Margis, R. Metatranscriptomic analysis of small RNAs present in soybean deep sequencing libraries. Genetics and Molecular Biology (Impresso). , v.35, p.xxx - , 2012.

2. **Kulcheski**, **F. R.**, de Oliveira, L. F. V. Molina, L. G., Almerão, M. P., Rodrigues, F. A., Marcolino, J., Barbosa, J. F., Stolf-Moreira, R., Nepomuceno, A. L., Marcelino-Guimarães, F. C., Abdelnoor, R. V., Nascimento, L. C., Carazzolle, M. F, Pereira, G. A. G., Margis, R. Identification of novel soybean microRNAs involved in abiotic and biotic stresses. BMC Genomics, v.12, p.307 - , 2011.

3. **Kulcheski, F. R.**, Graichen, F. A. S., Martinelli, J. A., Locatelli, A. B., Federizzi, L. C., Delatorre, C. A. Molecular mapping of *Pc68*, a crown rust resistance gene in *Avena sativa*. Euphytica (Wageningen) , v.175, p.423 - 432, 2010.

4. **Kulcheski, F. R.**, Marcelino-Guimarães, F. C., Nepomuceno, A. L., Abdelnoor, R. V., Margis, R. The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean. Analytical Biochemistry (Print), v.406, p.185 - 192, 2010.

5. **Kulcheski, F. R.**, Muschner, V. C., Lemke, A. P. L., Stehmann, J. R., Bonatto, S. L., Salzano, F. M., Freitas, L. B. Molecular Phylogenetic Analysis of *Petunia* Juss. (Solanaceae). Genetica ('s-Gravenhage), v.126, p.3 - 14, 2006.

6. Botton, M., **Kulcheski, F. R.**, Colletta, V. D., Arioli, C. J., Pastori, P. L. Avalição do uso do feromônio de confundimento no controle de *Grapholita molesta* (Lepidoptera: Torticidae) em pomares de pessegueiro. Idesia (Arica), v.23, p.43 - 50, 2005.

Trabalhos publicados em anais de eventos (completo)

1. Kulcheski, F. R., Delatorre, C. A. Análise de Ligação entre o Gene de resitência *Pc68* e marcadores do tipo AFLP em *Avena sativa* In: XXVII Reunião da Comissão Brasileira de Pesquisa da Aveia, 2007, Passo Fundo. **XXVII Reunião da Comissão Brasileira de Pesquisa da Aveia**. 2007. v.27.

2. Kulcheski, F. R., Graichen, F. A. S., Delatorre, C. A., Martinelli, J. A., Pacheco, M. Evolução da resistência parcial à ferrugem da folha na população de aveia branca UFRGS 7 X UFRGS 910906 In: XXVI Reunião da Comissão Brasileira de Pesquisa de Aveia, 2006, Guarapuava. **Resultados Experimentais**. 2006. v.26. p.61 - 64

3. Kulcheski, F. R., Delatorre, C. A. Identificação de marcadores SSR associados ao gene de resistência à ferrugem da folha (*Pc68*) em aveia (Avena sativa L.) In: XXVI Reunião da Comissão Brasileira de Pesquisa de Aveia, 2006, Guarapuava. **Resultados Experimentais**. 2006. v.26. p.193 - 194

4. Kulcheski, F. R., Botton, M., Kovaleski, A., Braghini, L. C. Avaliação de inseticidas visando ao controle da Pérola-da-terra *Eurhizococcus brasiliensis* (Hemiptera:Margarodidae) na cultura da videira. In: VII Reunião Sul-brasileira de insetos do solo, 1999, Piracicaba. **Anais e Ata Piracicaba:Esalq-USP**. 1999. p.102 - 104

Trabalhos publicados em anais de eventos (resumo)

1. **Kulcheski, F. R.**, de Oliveira, L. F. V., Molina, L. G., Almerão, M. P., Rodrigues, F. A., Marcolino, J., Barbosa, J. F., Stolf-Moreira, R., Nepomuceno, A. L., Marcelino-Guimarães, F. C., Abdelnoor, R. V., Nascimento, L. C., Carazzolle, M. F., Pereira, Gonçalo, A. G., Margis, R. Identification and expression analyses of soybean microRNAs under biotic and abiotic stresses In: III Simpósio Brasileiro de Genética Molecular de Plantas, 2011, Ilhéus. **III Simpósio Brasileiro de Genética Molecular de Plantas**. SBGMP, 2011.

2. Margis, R., Molina, L. G., Cordenonsi, G., Loss, G., de Oliveira, L. F. V., Carvalho, J. B., **Kulcheski, F. R.** Metranscriptomic analysis of small RNAs present in soybean deep sequencing libraries In: III Simpósio Brasileiro de Genética Molecular de Plantas, 2011, Ilhéus. **III Simpósio Brasileiro de Genética Molecular de Plantas**. SBGMP, 2011.

3. **Kulcheski, F. R.,** Rodrigues, F., Nepomunceno, A. L., Margis, R. Análise da expressão de microRNAs em raízes de soja sob condição de deficiência hídrica. In: XXXVII Reunião de Pesquisa de Soja da Região Sul, 2009, Porto Alegre. **Livro de Resumos da XXXVII Reunião de Pesquisa de Soja da Região Sul**. Porto Alegre: Editora da UFRGS, 2009.

4. **Kulcheski, F. R.**, Margis, R. Effect of drought stress on microRNA expression in soybean roots In: 55 Congresso Brasileiro de Genética, 2009, Águas de Lindóia. **Anais de resumos do 55 Congresso Brasileiro de Genética**. , 2009.

5. Longo, D, Muschner, V. C., **Kulcheski, F. R.**, Bonatto, S. L., Salzano, F. M., Freitas, L. B. Formas diplóides e hexaplóides de Passiflora misera HBR (Passifloraceae) podem ser espécies distintas? In: 56º Congresso Nacional de Botânica, 2005, Curitiba. **Anais do 56º Congresso Nacional de**. São Paulo: Sociedade Brasileira de Botânica, 2005. v.1.

6. Togni, PD, Lemke, A. P. L., **Kulcheski, F. R.**, Muschner, V. C., Stehmann, J. R., Bonatto, S. L., Salzano, F. M., Freitas, L. B. Relacionamento filogenético entre as espécies do gênero Petunia Juss (Solanaceae) In: 51° Congresso Nacional de Genética, 2005, Águas de Lindóia. **Anais do 51° Congresso Nacional de Genética**. Ribeirão Preto: Sociedade Brasileira de Genética, 2005. v.1.

7. **Kulcheski, F. R.**, Lemke, A. P. L., Muschner, V. C., Longo, D., Stehmann, J. R., Bonatto, S. L., Salzano, F. M., Freitas, L. B. Baixa variabilidade no gênero Petunia Juss. (Solanaceae) In: XVI **Salão de Iniciação Científica e XIII Feira de Iniciação Científica**, 2004, Porto Alegre. 2004.

8. **Kulcheski, F. R.**, Lorenz, A. P., Longo, D., Stehmann, J. R., Salzano, F. M., Bonatto, S. L., Freitas, L. B. Radiação Adaptativa no Gênero Petunia Juss. ss. (Solanaceae) In: 50º Congresso Brasileiro de Genética, 2004, Florianópolis. **Anais do 50º Congresso Brasileiro de Genética.** 2004.

9. **Kulcheski, F. R.**, Lemke, A. P. L., Muschner, V. C., Longo, D., Stehmann, J.

R., Bonatto, S. L., Salzano, F. M., Freitas, L. B. Relações Evolutivas no Gênero Petunia Juss. (Solanaceae) In: XV Encontro de geneticistas do Rio Grande do Sul, 2004, Canoas. **Anais do XV Encontro de geneticistas do Rio Grande do Sul**. 2004.

10. **Kulcheski, F. R.**, Lorenz, A. P., Mushner, V. C., Stehmann, J. R., Bonatto, S. L., Salzano, F. M., Freitas, L. B. Análise filogenética do Gênero Petunia Juss. s.s. (Solanaceae) In: 49º Congresso Nacional de Genética, 2003, Águas de Lindóia. **Anais do 49º Congresso Nacional de Genética**. , 2003.

11. **Kulcheski, F. R.**, Lorenz, A. P., Mushner, V. C., Stehmann, J. R., Bonatto, S. L., Salzano, F. M., Freitas, L. B. Filogenia Molecular do Gênero Petunia Juss. (Solanaceae) In: **XV Salão de Iniciação Científica**, 2003, Porto Alegre. 2003.

12. Contini, V, Lorenz, A. P., **Kulcheski, F. R.**, Mushner, V. C., Stehmann, J. R., Bonatto, S. L., Salzano, F. M., Freitas, L. B. Filogenia Molecular no Gênero Calibrachoa (Solanaceae) In: 49º Congresso Nacional de Genética, 2003, Águas de Lindóia. **Anais do 49º Congresso Nacional de Genética**. Ribeirão Preto: Sociedade Brasileira de Genética, 2003. v.1.

13. Contini, V, Lorenz, A. P., **Kulcheski, F. R.**, Mushner, V. C., Stehmann, J. R., Bonatto, S. L., Salzano, F. M., Freitas, L. B. Variabilidade Genética e Evolução do Gênero Calibrachoa (Solanaceae) In: XV Salão de Iniciação Científica e XII Feira de Iniciação Científica, 2003, Porto Alegre. **Anais do XV Salão de Iniciação Científica da UFRGS**. Porto Alegre: UFRGS, 2003. v.1.

14. **Kulcheski, F. R.**, Freitas, L. B., Lorenz, A. P., Leipnitz, A., Giacomet, C., Longo, D., Salzano, F. M., Stehmann, J. R., Lima, M. F., Mega, N. O., Santos, P., Bonatto, S., Souzachies, T., Muschner, V., Santos, V. Relações Filogenéticas em plantas:uma abordagem molecular nos generos Passiflora L. (Passifloraceae) e Petunia Juss. (s.s.) (Solanaceae) In: **XIII Encontro de geneticistas do Rio Grande do Sul**, 2002, Porto Alegre. 2002.

15. **Kulcheski, F. R.**, Freitas, L. B. Variabilidade Genética entre dois citótipos de Passiflora misera avaliada por RAPD-PCR In: **XIV Salão de Iniciação Científica da UFRGS**, 2002, Porto Alegre. 2002.

16. **Kulcheski, F. R.**, Botton, M. Oriental fruit control throught mating disruption with microencapsulate pheromone in peach orcherds. In: **XXI Internacional Congress of Entomology**, 2000, Foz do Iguaçu. 2000.

17. **Kulcheski, F. R.**, Colletta, VD, Botton, M., Braghini, L. C. Avaliação de germoplasma da videira visando a resistência da Pérola-da-terra *Eurhizococcus brasiliensis* (Hemiptera: Margarodidae). In: **IX Congresso Aberto de Viticultura e Enologia**, 1999, Bento Gonçalves. 1999.

18. **Kulcheski, F. R.**, Colletta, VD, Botton, M., Kovaleski, A. Avaliação do sistema Atrai e Mata visando controle de *Grapholita molesta* (Lepidoptera: Tortricidae) na cultura do pessegueiro. In: **IV Congresso Aberto aos Estudantes de Biologia**, 1999, Campinas. 1999.

Trabalhos publicados em anais de eventos (resumo expandido)

1. **Kulcheski, F. R.**, Delatorre, C. A., Pacheco, M., Martinelli, J. A. Avaliação da Estabilidade da Resistência Quantitativa à Ferrugem da Folha em Linhagens de Aveia Branca In: 4º Congresso Brasileiro de Melhoramento de Plantas, 2007, São Lourenço. **4º Congresso Brasileiro de Melhoramento de Plantas**. 2007. v.4.

Apresentação oral de trabalho

1. **Kulcheski, F. R.** Identification and expression analyses of soybean microRNAs under biotic and abiotic stresses, 2012. (Seminário, Apresentação de Trabalho)

2. **Kulcheski, F. R.**, Rodrigues, F., Nepomuceno, A. L., Margis, R. Análise da expressão de microRNAs em raízes de soja sob condição de deficiência hídrica, 2009. (Outra, Apresentação de Trabalho)

3. **Kulcheski, F. R.,** Margis, R. Effect of drought stress on microRNA expression in soybean roots, 2009. (Congresso, Apresentação de Trabalho)

4. **Kulcheski, F. R.**, Lemke, A. P. L., Muschner, V. C., Longo, D., Stehmann, J. R., Bonatto, S. L., Salzano, F. M., Freitas, L. B. Baixa Variabiliadde no Gênero Petunia Juss. (Solanaceae), 2004. (Outra, Apresentação de Trabalho)

5. **Kulcheski, F. R.**, Lorenz, A. P., Mushner, V. C., Stehmann, J. R., Bonatto, S. L., Salzano, F. M., Freitas, L. B. Filogenia Molecular do Gênero Petunia Juss. (Solanaceae), 2003. (Outra, Apresentação de Trabalho)

6. **Kulcheski, F. R.**, Freitas, L. B. Variabilidade Genética entre dois citótipos de Passiflora misera avaliada por RAPD-PCR, 2002. (Outra, Apresentação de Trabalho)

Organização de evento

1. **Kulcheski, Franceli Rodrigues**
Curso de Férias - As Células, 2010. (Outro, Organização de evento)

Totais de produção

Produção bibliográfica

Artigos completos publicados em periódico.....	6
Trabalhos publicados em anais de eventos.....	21
Apresentações de trabalhos (Congresso).....	1
Apresentações de trabalhos (Seminário).....	1
Apresentações de trabalhos (Outra).....	4