



UNIVERSIDADE
ESTADUAL DE LONDRINA

GESIELE ALMEIDA BARROS DE CARVALHO

**ANÁLISE TRANSCRICIONAL DOS GENES ENVOLVIDOS
NA NODULAÇÃO E PREDIÇÃO *IN SILICO* DE MICRORNAS
EM RAÍZES DE SOJA INOCULADAS COM A ESTIRPE CPAC
15 DE *BRADYRHIZOBIUM JAPONICUM***

GESIELE ALMEIDA BARROS DE CARVALHO

**ANÁLISE TRANSCRICIONAL DOS GENES ENVOLVIDOS
NA NODULAÇÃO E PREDIÇÃO *IN SILICO* DE MICRORNAS
EM RAÍZES DE SOJA INOCULADAS COM A ESTIRPE CPAC
15 DE *BRADYRHIZOBIUM JAPONICUM***

Dissertação apresentada ao Programa de Pós-Graduação em Biotecnologia, da Universidade Estadual de Londrina – UEL, como requisito parcial à obtenção do título de Mestre em Biotecnologia.

Orientadora: Dra. Mariangela Hungria.
Co-Orientadoras: Dra. Francismar C. M. Guimarães.
Dra. Jesiane S. S. Batista.

Londrina
2012

**Catálogo elaborado pela Divisão de Processos Técnicos da Biblioteca Central da
Universidade Estadual de Londrina**

Dados Internacionais de Catalogação-na-Publicação (CIP)

C331a Carvalho, Gesiele Almeida Barros de.

Análise transcricional dos genes envolvidos na nodulação e predição *in silico* de microRNAs em raízes de soja inoculadas com a estirpe CPAC 15 de *Bradyrhizobium japonicum* / Gesiele Almeida Barros de Carvalho. – Londrina, 2012.

93 f. : il.

Orientador: Mariangela Hungria.

Co-orientador: Francismar Corrêa Marcelino-Guimarães.

Co-orientador: Jesiane Stefânia da Silva Batista.

Dissertação (Mestrado em Biotecnologia) – Universidade Estadual de Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em Biotecnologia, 2012.

Inclui bibliografia.

1. Biotecnologia agrícola – Teses. 2. Genética microbiana – Teses. 3. Genética bacteriana – Teses. 4. Microorganismos do solo – Teses. 5. Biotecnologia vegetal – Teses. I. Hungria, Mariangela. II. Marcelino-Guimarães, Francismar Corrêa. III. Batista, Jesiane Stefânia da Silva. IV. Universidade Estadual de Londrina. Centro de Ciências Exatas. Programa de Pós-Graduação em Biotecnologia. V. Título.

CDU 663.1

GESIELE ALMEIDA BARROS DE CARVALHO

**ANÁLISE TRANSCRICIONAL DOS GENES ENVOLVIDOS NA
NODULAÇÃO E PREDIÇÃO *IN SILICO* DE MICRORNAS EM RAÍZES
DE SOJA INOCULADAS COM A ESTIRPE CPAC 15 DE
*BRADYRHIZOBIUM JAPONICUM***

Dissertação apresentada ao Programa de Pós-Graduação em Biotecnologia, da Universidade Estadual de Londrina – UEL, como requisito parcial à obtenção do título de Mestre em Biotecnologia, considerada aprovada pela banca examinadora.

BANCA EXAMINADORA

Dra. Mariangela Hungria
Embrapa Soja/UEL – Londrina - PR

Dra. Francismar Corrêa Marcelino-Guimarães
Embrapa Soja/UEL – Londrina - PR

Dra. Jesiane Stefania da Silva Batista
Embrapa Soja/UEL – Londrina - PR

Londrina, 23 de março de 2012.

“Não confunda derrotas com fracasso, nem vitórias com sucesso.

Na vida de um campeão sempre haverá algumas derrotas,
assim como na vida de um perdedor sempre haverá vitórias.

A diferença é que, enquanto os campeões crescem nas derrotas,
os perdedores se acomodam nas vitórias.”

(Roberto Shinyashiki)

AGRADECIMENTOS

Agradecer é admitir que houve um momento em que se precisou de alguém, agradecer é reconhecer que o homem jamais poderá lograr para si o dom de ser autossuficiente. Por isso, em primeiro lugar, sempre, agradeço a Deus, por estar ao meu lado em todos os momentos da minha vida, dando-me forças e guiando-me, para que eu possa atravessar os caminhos trilhados pelos desafios da vida.

Agradeço também à minha família, pelo apoio, em especial ao meu irmão, Jefferson, pela compreensão, e à minha mãe, Élia, pelo amor incondicional e por estar presente nas minhas maiores dificuldades, mãe você é simplesmente meu porto seguro, muito obrigada por muitas vezes ter abdicado dos seus sonhos para que eu pudesse realizar os meus.

Muito obrigado aos meus melhores amigos, Amanda, Francielle, Biba, Alex e Cássia, vocês são a família que Deus me permitiu escolher. Amo muito vocês, sou uma pessoa de muita sorte por poder contar com o seu apoio absolutamente todos os momentos da minha vida. Além disso, fica aqui meu agradecimento aos demais amigos e companheiros, que de alguma forma contribuíram nessa conquista.

À Mariangela, muito mais do que orientadora, nenhum “obrigado” seria o bastante para agradecer tudo o que ela faz por cada orientado. Mesmo assim, Chefa obrigada pela oportunidade e por fazer esse momento se tornar possível.

Aos amigos do laboratório Biotecnologia do Solo, aos que estão presentes e àqueles que passaram por aqui, pessoas muito especiais, que participaram dia após dia dessa caminhada, e que me proporcionaram momentos muito especiais. Em especial, agradeço à Ligia pelo grande carinho e à Jesi pela ajuda e incentivo, muito obrigada.

Um agradecimento especial ao Leandro Nascimento e ao Alexandre Paschoal, pelo auxílio nas análises de bioinformática, fundamentais na realização deste trabalho.

Também quero agradecer à Dra. Francismar pela disponibilidade de participar da banca da qualificação e agora da defesa da dissertação, e ainda, pelas sugestões que sem dúvida enriqueceram muito o trabalho, e sem mesmo que ela pudesse perceber, pelas conversas que me passaram tanta confiança.

Agradeço à Aline e à Dra. Elisete pela ajuda na realização deste estudo.

A todos os professores do programa de mestrado em Biotecnologia da UEL que contribuíram na minha formação.

Por último, mas não menos importante, agradeço à Embrapa Soja pelo espaço físico concedido para a realização das pesquisas, à CAPES pelo auxílio financeiro e ao CNPq pelo financiamento parcial do projeto GenoSoja.

CARVALHO, Gesiele Almeida Barros de. **Análise transcricional dos genes envolvidos na nodulação e predição *in silico* de miRNAs em raízes de soja inoculadas com a estirpe CPAC 15 de *Bradyrhizobium japonicum***. 2012. 93 f. Dissertação (Mestrado em Biotecnologia) – Universidade Estadual de Londrina, Londrina, 2012.

RESUMO

A fixação biológica de nitrogênio é um processo de grande importância para a cultura da soja, capaz de suprir toda sua demanda de nitrogênio, necessário para a síntese de biomoléculas. Entretanto, para que os nódulos sejam formados todas as etapas devem ser estreitamente coordenadas por uma série de sinais moleculares entre a planta e a bactéria. Conseqüentemente, o processo envolve várias etapas complexas e, embora venha sendo estudado há várias décadas, ainda há muito para ser entendido. Este trabalho objetivou analisar a expressão global de genes expressos diferencialmente nas raízes de soja, cultivar Conquista, inoculadas com *Bradyrhizobium japonicum* estirpe CPAC 15 (=SEMIA 5079), ambas amplamente utilizadas no cultivo dessa leguminosa no Brasil. O estudo visou, ainda, prever computacionalmente novos miRNAs que possam estar envolvidos na regulação da simbiose. Para atingir tal objetivo foi utilizada a técnica de hibridização subtrativa supressiva, combinada com o sequenciamento de nova geração e análises de bioinformática, o que resultou em uma biblioteca subtrativa (inoculadas X não-inoculadas) de raízes de soja com 3.210 transcritos diferencialmente expressos. Os dados foram analisados de acordo com as ontologias de função molecular e processo biológico. Em seguida, foram analisadas as vias metabólicas mais ativas nessa etapa da nodulação, e foram enfatizados genes relacionados ao metabolismo primário, às modificações da parede celular e ao sistema de defesa antioxidante. Funções putativas na simbiose de alguns desses genes foram atribuídas pela primeira vez na simbiose soja-*Bradyrhizobium*. Foi também realizada uma minuciosa predição *in silico* de miRNAs e seus alvos que, possivelmente, estão envolvidos na regulação da nodulação. Foram identificados nove miRNAs potenciais, bem como seus possíveis genes alvos. Os resultados obtidos permitiram obter um melhor entendimento sobre a regulação da expressão gênica em soja dez dias pós-inoculação, bem como fortaleceram a ideia de que a predição computacional de miRNAs pode auxiliar na compreensão das etapas iniciais da simbiose. O estudo é pioneiro com uma cultivar de soja e estirpe brasileiras.

Palavras-chave: Biblioteca subtrativa. Expressão diferencial de genes. Nodulação. Predição de miRNA.

CARVALHO, Gesiele Almeida Barros de. **Transcriptional analysis of genes involved in nodulation and *in silico* prediction of microRNAs in soybean roots inoculated with *Bradyrhizobium japonicum* strain CPAC 15**. 2012. 93 f. Dissertation (Master's degree in Biotechnology) – Universidade Estadual de Londrina, Londrina, 2012.

ABSTRACT

Biological nitrogen fixation is a process of great importance to the soybean crop, capable of supplying all plant's needs on nitrogen, necessary for the synthesis of biomolecules. However, for nodule formation all steps should be closely coordinated by a series of molecular signals between the plant and bacterium. Therefore, the biological process involves several complex steps, and although it has been studied for several decades, much remains to be understood. This study aimed to analyze the global expression of genes differentially expressed in soybean roots of cultivar Conquista inoculated with *Bradyrhizobium japonicum* strain CPAC 15 (=SEMIA 5079), both broadly used in Brazil. In addition, we performed a computationally prediction of new miRNAs that may be involved in regulating the process. To achieve this we used the suppressive subtractive hybridization technique combined with the new generation sequencing and bioinformatics analyses, which resulted in a subtractive library (non-inoculated X inoculated) of soybean roots with 3,210 differentially expressed transcripts. The data were grouped according to the ontology of molecular function and biological process. Next, we analyzed the most active metabolic pathways at this stage of nodulation, and emphasis was given to genes related to the primary metabolism, cell wall modifications and antioxidant defense system. Putative functions were attributed to some of these genes for the first time. We have also performed a detailed *in silico* prediction of miRNAs and their targets, which are possibly involved in regulation of nodulation. Nine potential miRNAs were identified, as well as the possible target genes. The results obtained have allowed to get a better understanding about the regulation of gene expression in soybean ten days post-inoculation, as well as have reinforced the idea that the computational prediction of miRNAs may help to understand the initial steps of the symbiosis. Moreover, the study is pioneer with a Brazilian soybean cultivar and a Brazilian *Bradyrhizobium* strain.

Keywords: Subtractive library. Differential expression of genes. Nodulation. Prediction of miRNA.

LISTA DE TABELAS

Tabela 1 – Classificação dos processos biológicos desencadeados nas raízes da soja em resposta à inoculação da estirpe CPAC 15 de <i>B. japonicum</i>	37
Tabela 2 – Genes que codificam enzimas presentes na via glicolítica (A) e no ciclo de Krebs (B) diferencialmente expresso durante a nodulação.....	39
Tabela 3 – Genes da soja que codificam as enzimas que estão atuando na proteção celular contra os danos oxidativos.	40
Tabela 4 – Genes selecionados, com base no RPKM, a partir dos processos biológicos induzidos nas raízes de soja inoculadas com <i>Bradyrhizobium</i>	42
Tabela 5 – Genes da soja que codificam as enzimas ativas identificadas a partir da via de metabolismo de carboidratos.....	48
Tabela 6 – Relatório do pipeline de anotação geral de RNAs não codificantes.	49
Tabela 7 – Sequências precursoras de miRNAs evidenciadas através do pipeline de anotação de ncRNA.....	50
Tabela 8 – Resultado da análise das sequências candidatas no miRBase.	52
Tabela 9 – Resultado da predição in silico do miRNA maduro.	55
Tabela 10 – Transcritos alvos de <i>Glycine max</i> para os miRNAs preditos.....	59

LISTA DE FIGURAS

Figura 1 – Estágios de desenvolvimento de nódulos de soja.....	19
Figura 2 – Esquema simplificado da biogênese de miRNA em planta.....	23
Figura 3 – Esquema simplificado da hibridização subtrativa supressiva.....	26
Figura 4 – Representação esquemática das etapas de sequenciamento de nova geração da plataforma Illumina/Solexa.....	27
Figura 5 – Plântulas de soja utilizadas no experimento.....	29
Figura 6 – Porcentagem de sequências (%) alocadas em (A) funções moleculares e (B) processos biológicos (segundo classificação do Gene Ontology) de genes diferencialmente expressos em raízes de soja na presença da bactéria fixadora de nitrogênio <i>Bradyrhizobium japonicum</i> , estirpe CPAC15 (=SEMIA 5079).....	36
Figura 7 – Estruturas secundárias dos pré-miRNAs e dos miRNAs maduros de <i>Glycine max</i> preditos in silico.....	56

SUMÁRIO

1	INTRODUÇÃO	12
2	REVISÃO BIBLIOGRÁFICA	14
2.1	A FIXAÇÃO BIOLÓGICA DE NITROGÊNIO NA CULTURA DA SOJA.....	14
2.2	INTERAÇÃO MOLECULAR NO ESTÁGIO INICIAL DO ESTABELECIMENTO SIMBIÓTICO.....	15
2.3	ALTERAÇÕES FISIOLÓGICAS QUE OCORREM NA SOJA EM RESPOSTA À PRESENÇA DE BRADYRHIZOBIUM	17
2.4	EXPRESSÃO GÊNICA NA NODULAÇÃO.....	20
2.5	MICRORNAs: IMPORTANTES REGULADORES DA EXPRESSÃO GÊNICA.....	22
2.6	FERRAMENTAS PARA ESTUDO TRANSCRICIONAL	25
2.6.1	Hibridização Subtrativa Supressiva	25
2.6.2	Sequenciamento de Nova Geração em Estudos de Transcriptoma	26
3	OBJETIVO	28
3.1	OBJETIVOS ESPECÍFICOS	28
4	MATERIAL E MÉTODOS	29
4.1	MATERIAL VEGETAL.....	29
4.2	TRATAMENTOS	29
4.3	PREPARO DO INÓCULO E INOCULAÇÃO	30
4.4	EXTRAÇÃO DE RNA E ISOLAMENTO DO MRNA.....	30
4.5	CONSTRUÇÃO DA BIBLIOTECA SUBTRATIVA	31
4.6	SEQUENCIAMENTO, MONTAGEM DAS SEQUÊNCIAS E ANÁLISES DE BIOINFORMÁTICA.....	31
4.7	PREDIÇÃO DE MICRORNAs ENVOLVIDOS NA NODULAÇÃO.....	32
4.7.1	Predição de Alvos dos miRNAs.....	34
5	RESULTADOS E DISCUSSÃO	35
5.1	CATEGORIZAÇÃO FUNCIONAL DOS GENES DIFERENCIALMENTE EXPRESSOS	35
5.1.1	Processo Metabólico	38
5.1.2	Sinalização	43

5.1.3	Resposta a Estímulos.....	44
5.1.4	Transporte	45
5.1.5	Ciclo Celular e Regulação de Processos Biológicos	46
5.1.6	Modificação da parede celular	47
5.2	PREDIÇÃO IN SILICO DE MI RNAs ENVOLVIDOS NA NODULAÇÃO	49
CONCLUSÃO		62
REFERÊNCIAS.....		63
ANEXOS.....		76
ANEXO A –	Via glicolítica ativa durante a nodulação.	77
ANEXO B –	Ciclo de Krebs ativo durante a nodulação.....	78
ANEXO C –	Via metabólica da glutathione – sistema de defesa antioxidante presente na nodulação.....	79
ANEXO D –	Via metabólica de carboidratos que regulam a formação dos nódulos.....	80
ANEXO E –	Resultado detalhado do alinhamento das sequências dos candidatos precursores de miRNAs segundo o miRBase.....	81
ANEXO F –	Resultado da predição das estruturas secundárias dos candidatos precursores de miRNAs e a localização dos miRNAs maduros	84

1 INTRODUÇÃO

A soja (*Glycine max* (L.) Merrill) é uma leguminosa com importância em nível mundial, sendo que o interesse por essa cultura é devido, sobretudo, ao teor proteico elevado nos grãos, principalmente nos países que apresentam carências nutricionais. Por esse motivo, a soja apresenta uma demanda elevada de nitrogênio (N) para seu desenvolvimento e, conseqüentemente, para sua produtividade. Isso faz com que esse nutriente seja, frequentemente, limitante para a produtividade da cultura, pois apesar de estar presente na atmosfera, nenhum organismo eucariótico consegue assimilá-lo em sua forma molecular (N₂). Entretanto, o N que a planta necessita pode ser totalmente suprido pelo processo de fixação biológica de nitrogênio (FBN), através da associação simbiótica com bactérias pertencentes, principalmente, às espécies *Bradyrhizobium japonicum* e *Bradyrhizobium elkanii*.

A FBN é essencial para se atingir alta produtividade a baixo custo, sendo esse processo um dos responsáveis pelo sucesso da cultura no Brasil, hoje posicionado como segundo maior produtor de soja. A FBN é capaz de suprir toda demanda de N dessa leguminosa, além de ser mais favorável ao meio ambiente. Em termos econômicos, viabiliza o cultivo para o agricultor, pela diminuição dos custos de produção, em relação à suplementação com fertilizantes nitrogenados.

Uma maior compreensão da simbiose rizóbio-leguminosa pode ser alcançada através da análise da expressão gênica em determinadas etapas do processo de interação. Apesar do processo de nodulação já ser bastante estudado há várias décadas, não só em soja, mas em outras culturas de plantas, por ser um processo extremamente complexo, ainda existem vários mecanismos de regulação que precisam ser esclarecidos.

O estudo aprofundado de cada fase da nodulação é, portanto de grande relevância, não só por proporcionar avanços no conhecimento, como também pelas perspectivas de identificação de mecanismos que permitam a maximização do processo de FBN com a soja, contribuindo para o sucesso agrícola da cultura.

Nesse sentido, este estudo objetivou avaliar os mecanismos moleculares e reguladores, ativados na soja durante a nodulação, através de ferramentas para a análise global dos genes diferencialmente expressos após a

inoculação das raízes da soja com uma estirpe brasileira de *B. japonicum*. Além disso, realizou-se ainda a predição de microRNAs potencialmente envolvidos neste processo biológico.

2 REVISÃO BIBLIOGRÁFICA

2.1 A FIXAÇÃO BIOLÓGICA DE NITROGÊNIO NA CULTURA DA SOJA

A soja (*Glycine max* (L). Merrill) é uma planta da família *Fabaceae* (= *Leguminosae*), também conhecida como leguminosa, de grande importância nutricional e econômica no mundo. Atualmente são produzidos, por ano, aproximadamente 200 milhões de toneladas de soja no mundo, sendo que o Brasil aparece como o segundo maior produtor, com uma produção de 75 milhões de toneladas na safra 2010/2011 (EMBRAPA, 2012).

Essa leguminosa é, hoje, uma das mais importantes plantas cultivadas, sendo empregada na produção de alimentos, de óleo, na medicina, na produção de biocombustível, entre outros (GRAHAM; VANCE, 2003). O interesse por essa cultura pode ser explicado, principalmente, pelo elevado teor de proteína (cerca de 40%) presente nos grãos, sendo de particular importância nos países que apresentam carências nutricionais (ALVES *et al.*, 2003). Por apresentar esse elevado teor proteico, apresenta uma grande demanda em nitrogênio (N), sendo esse nutriente, em geral, um dos principais fatores limitantes à produtividade da cultura (HUNGRIA *et al.*, 2006).

O N é um nutriente indispensável para o desenvolvimento das plantas, requerido para a síntese de aminoácidos, proteínas, bases nitrogenadas, ácidos nucleicos, hormônios, clorofila, entre outros (KUSANO *et al.*, 2011). Apesar desse elemento representar cerca de 80% dos gases presentes na atmosfera, na forma de N_2 , organismos eucariontes são incapazes de absorver o N_2 e convertê-lo a uma forma assimilável, devido à tripla ligação existente entre os átomos de N_2 ($N\equiv N$), uma das mais fortes que se tem conhecimento na natureza (BRECHENMACHER *et al.*, 2008).

Contudo, uma classe de organismos procariotos, denominados diazotróficos, consegue utilizar o N_2 através do processo de fixação biológica de nitrogênio (FBN). Esses microrganismos possuem um complexo enzimático, denominado nitrogenase, capaz de quebrar a tríplice ligação do N_2 , resultando na síntese de amônia (NH_3), que será disponibilizada, inicialmente, na forma de amônio (NH_4^+). Estes são imediatamente incorporados em esqueletos de carbono para ser

utilizado por determinadas plantas, suprimindo o N necessário ao desenvolvimento e à produtividade (FERGUSON *et al.*, 2010).

Os microrganismos diazotróficos se associam a diversas espécies vegetais em diferentes graus de especificidade, podendo ser classificados como de vida livre, associativos ou simbióticos (PEOPLES; CRASWELL, 1992; JAMES, 2000; HUNGRIA *et al.*, 2006). No caso da soja e de outras leguminosas, essa associação simbiótica pode ser facilmente identificada, pois estruturas especializadas para o processo biológico, chamadas de nódulos são formadas nas raízes, característica da simbiose (van RHIJN; VANDERLEYDEN, 1995; STOUGAARD, 2000; FERGUSON *et al.*, 2010).

Dentre as vantagens obtidas pela FBN, destaca-se a economia no uso de fertilizantes nitrogenados. A inoculação das sementes de soja com bactérias específicas e comprovadamente eficientes é capaz de dispensar a aplicação de fertilizantes nitrogenados, proporcionando uma economia estimada em quase U\$ 7 bilhões por ano para o país (HUNGRIA *et al.*, 2007). Além disso, é uma prática ambientalmente segura, visto que uma parcela significativa dos fertilizantes nitrogenados aplicados é perdida através dos processos de lixiviação e desnitrificação (JENSEN; HAUGGAARD-NIELSEN, 2003).

2.2 INTERAÇÃO MOLECULAR NO ESTÁGIO INICIAL DO ESTABELECIMENTO SIMBIÓTICO

A soja consegue suprir sua necessidade de N através da simbiose com bactérias pertencentes, principalmente, às espécies *Bradyrhizobium japonicum* e *Bradyrhizobium elkanii*, sendo que, entre e dentro dessas duas espécies existe grande variabilidade, com diferentes propriedades morfológicas, fisiológicas e genéticas, que passam a caracterizar as estirpes (ou cepas) distintas (HUNGRIA *et al.*, 1998).

O termo simbiose é atribuído pela relação estreita entre os parceiros: o N fixado pelos microrganismos é disponibilizado para as plantas hospedeiras, enquanto estas disponibilizam fontes de C necessárias para a nutrição e energia do microssimbionte (PRELL *et al.*, 2009).

No estabelecimento da simbiose, o microssimbionte e a planta hospedeira iniciam uma comunicação através de uma sinalização molecular altamente específica (STOUGAARD, 2000; OLDROYD *et al.*, 2011). A princípio, a

planta hospedeira sintetiza e libera metabólitos secundários específicos, que atuam como quimio-atraentes das bactérias fixadoras de nitrogênio e que, na etapa seguinte, atuam como indutores dos genes de nodulação dos rizóbios (COOPER, 2007; OLDROYD *et al.*, 2011). As leguminosas hospedeiras liberam diferentes grupos de metabólitos secundários indutores (flavonoides e betainas), havendo grande especificidade com o microssimbionte. A daidzeína e a genisteína, isoflavonoides produzidos pela soja, são fortes indutores de *B. japonicum* (SUBRAMANIAN *et al.*, 2006; DOLATABADIAN *et al.*, 2012).

Em resposta aos indutores liberados pela planta hospedeira, a transcrição de uma classe especial de genes bacterianos, os genes de nodulação, é ativada. Esses genes codificam a síntese de moléculas de oligossacarídeos lipoquitínicos, denominados fatores de nodulação ou fatores Nod (LIMPENS; BISSELING, 2003; STACEY *et al.*, 2006, FERGUSON *et al.*, 2010). As células vegetais detectam os fatores Nod e a presença dos rizóbios através de receptores presentes na superfície celular das raízes e, em seguida, ocorre ativação de vias de sinalização, que acionam as respostas adaptativas necessárias para a nodulação. É um processo complexo, e controlado por múltiplos receptores e vias de sinalização (LIMPENS; BISSELING, 2003; OLDROYD *et al.*, 2011).

A percepção dos fatores Nod pelo hospedeiro desencadeia uma cascata de transdução de sinais, envolvendo canais iônicos proteicos localizados na membrana nuclear, as nucleoporinas (FERGUSON *et al.*, 2010), e mudanças na concentração intracelular de cálcio (Ca^{2+}), que são detectadas pelas calmodulinas (CaM) (BRECHENMACHER *et al.*, 2008). As calmodulinas são proteínas que, quando vinculadas ao Ca^{2+} , sofrem uma mudança conformacional, ativando a calmodulina dependente quinase (CaMK), responsável pela indução dos fatores transcricionais da nodulação, permitindo a expressão dos genes da planta envolvidos na nodulação, denominadas nodulinas (OLDROYD; DOWNIE, 2004; OLDROYD *et al.*, 2011). A oscilação na concentração citosólica de Ca^{2+} , conhecida como Ca^{2+} *spiking*, é induzida nas células alguns minutos após a indução de fluxos de Ca^{2+} (WAIS *et al.*, 2000), e tem um papel central no processo de nodulação, pois pode induzir a deformação do pelo da raiz e as mudanças no citoesqueleto que são requeridas para o encurvamento e invasão da raiz pelos rizóbios (CÁRDENAS *et al.*, 1998; OLDROYD *et al.*, 2011).

2.3 ALTERAÇÕES FISIOLÓGICAS QUE OCORREM NA SOJA EM RESPOSTA À PRESENÇA DE *BRADYRHIZOBIUM*

Os fatores Nod e os flavonoides são os principais determinantes da especificidade entre o hospedeiro e o simbiote (OLDROYD; DOWNIE, 2008). Assim, a presença da bactéria compatível e seu fator Nod correspondente são suficientes para iniciar o desenvolvimento do nódulo (FERGUSON *et al.*, 2010). Os genes da planta cuja expressão é induzida por esses fatores Nod são denominados de nodulinas, que codificam proteínas requeridas para o desenvolvimento nodular e para a manutenção da FBN. De acordo com seu padrão de expressão e função, as nodulinas são classificadas em nodulinas precoces ou nodulinas tardias (STOUGAARD, 2000).

A infecção da planta pela bactéria específica resulta no encurvamento dos pelos radiculares, provocando a degradação da parede celular. A penetração do simbiote na raiz provoca a expansão da parede celular e, o início da migração nuclear para o local de infecção (STAIGER, 2000; OLDROYD *et al.*, 2011). A entrada da bactéria redireciona a membrana plasmática e a deposição interna da parede celular, formando uma invaginação. Esse processo leva à formação dos cordões de infecção, estruturas tubulares compostas por células da parede celular vegetal, que conduzem os rizóbios para o interior das células corticais da planta (LIMPENS; BISSELING, 2003; MINIC, 2008; OLDROYD *et al.*, 2011).

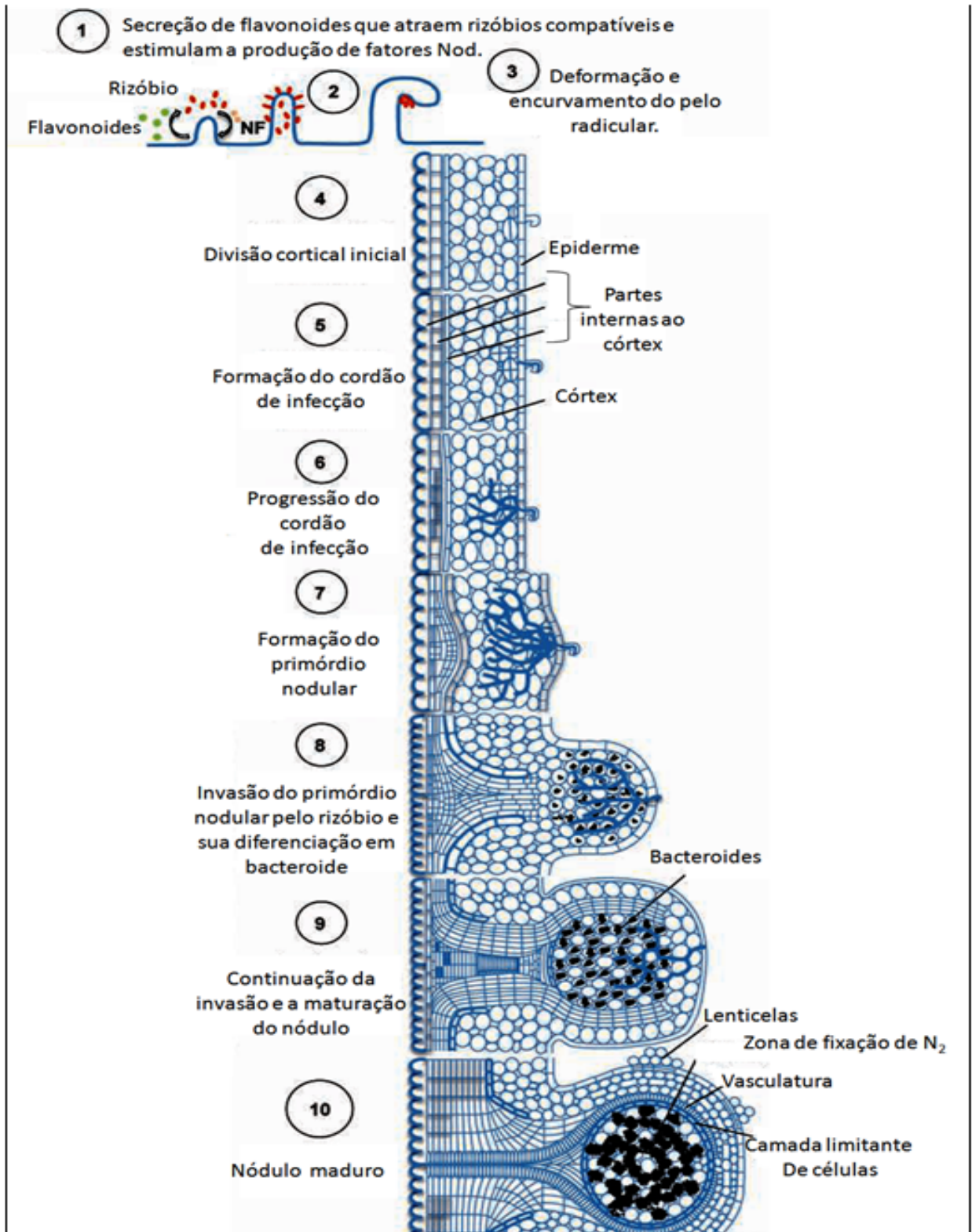
As bactérias são então liberadas no citoplasma da célula hospedeira, e através de um processo que se assemelha à endocitose, os rizóbios são envolvidos por uma membrana derivada da planta, chamada membrana peribacteroidal, formando o simbiossomo (STOUGAARD, 2001; OLDROYD *et al.*, 2011). A bactéria envolvida por essa membrana continua se dividindo dentro da célula hospedeira antes de se diferenciar em bacteroide e começar o processo de FBN propriamente dito (FERGUSON *et al.*, 2010).

Essa etapa de infecção pelos rizóbios provoca alterações no ciclo celular, desencadeia a reorganização do citoesqueleto e induz à divisão celular (BLANCAFLOR *et al.*, 2006; OLDROYD *et al.*, 2011). Todos esses acontecimentos são controlados pela cascata de transdução de sinais, envolvendo a regulação do nível hormonal da planta, a oscilação na concentração de cálcio dentro da célula e alterações no nível redox celular, que resultam na indução ou repressão de fatores

transcricionais responsáveis pelas alterações moleculares necessárias ao estabelecimento da simbiose e ao processo de FBN (FERGUSON; MATHESIUS, 2003; MARINO *et al.*, 2009; FERGUSON *et al.*, 2010; OLDROYD *et al.*, 2011). Essas alterações na morfologia da planta que ocorrem durante a formação dos nódulos é apresentada na Figura 1.

É interessante observar ainda, que, no estágio de infecção, mais especificamente na colonização da rizosfera, os rizóbios geralmente não provocam reações de defesa na planta, ao contrário do que ocorre pela invasão por patógenos, embora os processos compartilhem várias semelhanças (BARON; ZAMBRYSKI, 1995). Conseqüentemente, para um estabelecimento simbiótico bem sucedido, os rizóbios devem suprimir as respostas de defesa do hospedeiro (BARON; ZAMBRYSKI, 1995; MITHÖFER, 2002; BRECHENMACHER *et al.*, 2008). Um exemplo de como isso pode ocorrer, é pela presença de compostos específicos derivados dos rizóbios, que podem inibir a resposta da defesa vegetal, permitindo que o hospedeiro seja colonizado. Esses compostos podem estar presentes na superfície celular do rizóbio, como os exopolissacarídeos (EPS), lipopolissacarídeos (LPS) e β -glicanos cíclicos, ou podem ser derivados dessas moléculas, que atuam como compostos de sinalização e suprimem respostas de defesa da planta (MITHÖFER, 2002).

Figura 1 – Estágios de desenvolvimento de nódulos de soja.



Fonte: Modificado de Ferguson et al. (2010).

2.4 EXPRESSÃO GÊNICA NA NODULAÇÃO

Na nodulação, sabe-se que há interação entre os genótipos da planta e da bactéria, sendo que os primeiros exercem influência seletiva sobre as estirpes que formarão nódulos nas suas raízes. Portanto, diante da complexidade dessa interação, o estudo de expressão gênica na nodulação é de extrema importância. Assim, estratégias para análises funcionais de padrões de expressão gênica em células específicas podem fornecer informações importantes, permitindo uma maior compreensão do processo de nodulação (STACEY *et al.*, 2006).

Uma das abordagens mais adotadas no estudo de expressão gênica de eucariotos, denominado estudo de transcriptômica, representa a análise do conteúdo global de transcritos de diferentes tecidos em resposta a uma dada condição biótica ou abiótica específica. Existem várias metodologias que permitem o acesso a tal informação, como o SAGE (*Serial Analysis of Gene Expression*) (VELCULESCU *et al.*, 1995), os microarranjos (ou *microarrays*) (SCHENA *et al.*, 1995), a hibridização subtrativa supressiva (DIATCHENKO *et al.*, 1996) e o RNA Seq (WANG *et al.*, 2009). Além disso, as técnicas *Northern blot* (ALWINE *et al.*, 1977) e PCR em tempo real (RT qPCR) (HEID *et al.*, 1996) permitem analisar a expressão de genes pontuais.

Existem estudos importantes de transcriptômica em larga escala, que analisaram as alterações desencadeadas nas raízes de soja em resposta à inoculação com *B. japonicum*. Em um estudo conduzido por Brechenmacher e colaboradores (2008), o perfil de expressão gênica de raízes de soja inoculadas com *B. japonicum* (estirpe USDA 110) foi avaliado aos 4, 8 e 16 dias após a inoculação (DAI), utilizando as técnicas de microarranjo e PCR em tempo real, na qual foram identificados vários genes diferencialmente expressos, além disso, observou-se que *B. japonicum* reduz as respostas de defesa da planta durante a nodulação. E ainda, demonstraram que existe um complexo mecanismo regulatório na planta, que permite sua adaptação às alterações que ocorrem no seu estado nutricional, essencial ao desenvolvimento nodular.

Já o estudo de Libault *et al.* (2010) empregou três diferentes metodologias para analisar o transcriptoma das raízes de soja também inoculadas com a estirpe USDA 110: microarranjo, sequenciamento e PCR em tempo real, onde as raízes foram analisadas algumas horas após a inoculação (12, 24 e 48 horas).

Sendo que, um total de 1.973 genes de soja foram diferencialmente expressos durante a infecção das raízes, esses dados apontaram que ocorre uma rápida transcrição de genes nas células radiculares em resposta à inoculação, sugerindo que o reconhecimento dos fatores Nod desencadeia a cascata de sinalização. Mostraram ainda, que a expressão dos genes da soja é modificada de acordo com os diferentes estágios da nodulação, que levarão às alterações morfofisiológicas necessárias durante o processo.

Além desses, estudos de perfil transcricional com outras leguminosas durante a nodulação também foram realizados, principalmente com *Lotus japonicus* e *Medicago truncatula*, consideradas plantas modelos para esse tipo de estudo.

Em *L. japonicus* observou-se que os genes envolvidos na quebra da sucrose, glicólise, reciclagem de CO₂ e síntese de aminoácidos são diferencialmente expressos, pois esses processos são acelerados durante a nodulação. Além disso, genes envolvidos no transporte transmembrana, metabolismo hormonal, parede celular e transdução de sinal também foram induzidos pela presença da bactéria (COLEBATCH *et al.*, 2002; ASAMIZU *et al.*, 2005). Kouchi *et al.* (2004) mostraram que a expressão de um total de 1.076 genes foi significativamente acelerada em *L. japonicus* durante as sucessivas fases de desenvolvimento nodular até a fixação de N. Em meio a esses, incluem os genes envolvidos no catabolismo de fotossintatos, transporte transmembrana, síntese hormonal, modificação da parede celular e regulação da transcrição. Além disso, observaram que no início da infecção, genes envolvidos com a resposta de defesa foram induzidos, mas nos estágios subsequentes da nodulação sua expressão foi suprimida.

Os resultados obtidos nos estudos de transcriptoma de *M. truncatula* também mostraram que ocorre uma série de eventos moleculares coordenados durante a nodulação, onde muitos processos fisiológicos são acelerados pela presença da bactéria fixadora de nitrogênio. Isso provoca um aumento na expressão dos genes envolvidos com os processos de sinalização, regulação transcricional, oscilações de cálcio intracelular, explosão oxidativa provocada pela intensa atividade metabólica, modificações da parede celular vegetal, proliferação celular e alterações no citoesqueleto, necessários para o desenvolvimento e funcionamento dos nódulos (EL YAHYAOUÏ *et al.*, 2004; LOHAR *et al.*, 2006; STARKER *et al.*, 2006).

Entretanto, ainda não existe nenhum estudo transcricional na nodulação em soja com cultivares de soja brasileira, ou com estirpes selecionadas de *B. japonicum* quanto às habilidades simbióticas nas condições de solos brasileiros. Estudos dessa natureza merecem particular atenção, visto que o Brasil ocupa posição de liderança em benefícios com a FBN.

2.5 MICRORNAS: IMPORTANTES REGULADORES DA EXPRESSÃO GÊNICA

Sabe-se que o processo de nodulação envolve uma regulação coordenada da expressão dos genes, onde atuam inúmeras moléculas regulatórias, dentre elas, os microRNAs (miRNAs). Esses são pequenas moléculas de RNA de fita simples, não codificantes de proteínas, que apresentam um tamanho aproximado de 21-24 nucleotídeos (nt), e regulam negativamente genes alvos em nível pós-transcricional, em eucariotos (BARTEL, 2004; JUNG *et al.*, 2009; VOINNET, 2009). MiRNAs apresentam grande importância nos sistemas biológicos, em especial nas plantas, onde regulam inúmeros processos, como crescimento e desenvolvimento; também atuam em resposta ao estresse biótico e abiótico, além de participarem da regulação gênica durante a nodulação (JUNG *et al.*, 2009; SIMON *et al.*, 2009).

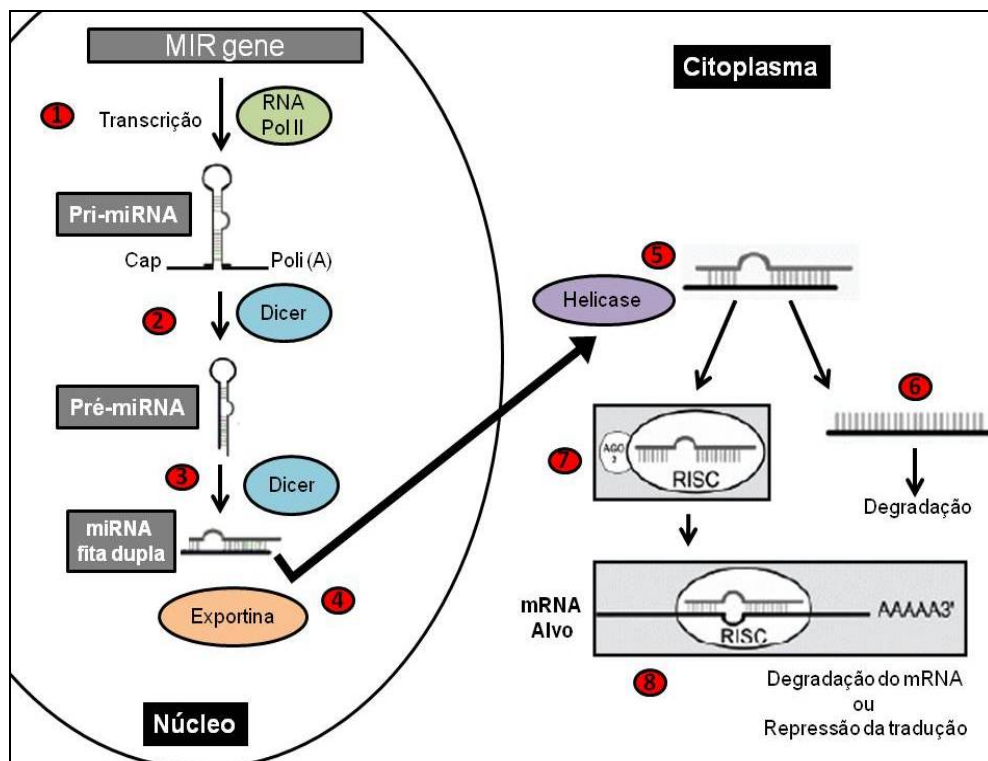
Os miRNAs são transcritos pela ação da RNA polimerase II em longos transcritos primários (pri-miRNAs) possuindo 5' capa e 3' cauda poli (A+) (BARTEL, 2004; CHEN *et al.*, 2009). Os pri-miRNAs são processados pela ação da proteína RNase III, conhecida como *Dicer*, resultando nos precursores de miRNA (pré-miRNAs), que formam uma estrutura distinta auto-complementar (*hairpin*), conhecida como *stem-loop* (MILLAR; WATERHOUSE, 2005; JUNG *et al.*, 2009).

A *Dicer*, em plantas, também é responsável pelo processamento do pré-miRNA em miRNA de fita dupla (miRNA:miRNA*), sendo que todo esse processo, em plantas, ocorre no interior do núcleo (BARTEL, 2004; JUNG *et al.*, 2009; VOINNET, 2009). No próximo estágio, os miRNAs maduros de fita dupla são exportados para o citoplasma pela ação da proteína exportina, onde a helicase desfaz a dupla fita do miRNA. Desse modo, o miRNA* é degradado e o miRNA maduro, agora fita simples, é seletivamente incorporado a um complexo ribonucleoproteico contendo uma proteína Argonaute (AGO), conhecida como complexo de RNA que induz o silenciamento dos genes (RISC- *RNA-induced*

silencing complex), que a partir de então estão prontos para exercerem sua função (BARTEL, 2004; JUNG *et al.*, 2009; VOINNET, 2009).

Após a formação do complexo miRNA-RISC, este pode regular a expressão gênica através de dois mecanismos pós-transcricionais: clivagem ou repressão da tradução (BARTEL, 2004; JUNG *et al.*, 2009). O que determina esse mecanismo é o alto grau de complementaridade entre miRNA-mRNA, onde a regra geral diz que se ocorrer um pareamento perfeito ou praticamente perfeito a regulação da expressão gênica se dá através da clivagem. No entanto, caso a complementaridade não seja específica, ocorre a repressão da tradução (CARRINGTON; AMBROS, 2003; JONES-RHOADES; BARTEL, 2004; VOINNET, 2009). Diferente da maioria dos animais, as plantas possuem uma complementaridade praticamente perfeita, senão perfeita, entre os miRNAs e seus alvos, portanto, há predomínio do mecanismo de regulação da expressão gênica por clivagem da região codificante dos mRNAs em plantas (MILLAR; WATERHOUSE, 2005; JUNG *et al.*, 2009). O processo de biogênese do miRNA em planta pode ser visualizado na Figura 2.

Figura 2 – Esquema simplificado da biogênese de miRNA em planta.



Os miRNAs são altamente conservados nas plantas, mas o grau de conservação é baixo entre plantas e animais (JUNG *et al.*, 2009). Além disso, existem diferenças entre os miRNAs de plantas e animais, principalmente em termos de biogênese e no mecanismo funcional (JUNG *et al.*, 2009). Diferem, ainda, no comprimento da sequência do precursor do miRNA, que em animais é de, aproximadamente, 70-80 nt, enquanto que em plantas pode variar de 50 a 350 nt (JUNG *et al.*, 2009). O genes que codificam miRNAs são chamados genes MIR (VOINNET, 2009).

Ao contrário da maioria dos genes MIR de metazoários que são encontrados principalmente no interior de introns ou exons, a maioria dos genes MIR de plantas advém de regiões intergênicas (VOINNET, 2009), seguido por regiões intrônicas e região UTR (*untranslated region*). Além disso, representando uma minoria, genes MIR também podem ser encontrados em regiões que codificam proteínas (CDS – *coding sequence*) (FAHLGREN *et al.*, 2010; NOZAWA *et al.*, 2012). Finalmente, diferem dos animais pelo fato dos genes MIR de plantas serem geralmente monocitrônicos, e com poucas famílias transcritas como unidades policitrônicas (NAQVI *et al.*, 2012).

O primeiro trabalho a identificar miRNAs em soja foi realizado por Subramanian e colaboradores (2008), que estudaram o papel dos miRNAs na simbiose soja-*Bradyrhizobium*, identificando 35 novas família de miRNA. O segundo estudo de miRNAs envolvidos na nodulação em soja foi realizado no ano seguinte por Wang *et al.* (2009). Apesar do conhecimento atual de que miRNAs estão envolvidos na regulação da nodulação, participando tanto dos estágios iniciais, compreendendo a regulação da homeostase de auxina e o processo de sinalização, como na maturação do nódulo e na própria fixação de nitrogênio (SUBRAMANIAN *et al.*, 2008; WANG *et al.*, 2009; LI *et al.*, 2010), suas funções ainda não estão totalmente esclarecidas.

Os miRNAs parecem regular os estágios iniciais da nodulação, como mostrado em *Medicago truncatula*, onde o fator transcricional MtHAP2 é regulado pelo mir-169 (COMBIER *et al.*, 2006). Outro trabalho importante realizado por Joshi e colaboradores (2010) ampliou o conhecimento sobre miRNAs em soja, pela identificação de 87 novos miRNAs, em diferentes tecidos dessa leguminosa (semente, flor, raiz e nódulo). Mais recentemente, o estudo de Kulcheski *et al.* (2011)

identificou novos miRNAs em soja que respondem aos estresses bióticos e abióticos.

Atualmente existem 362 miRNAs identificados em soja (miRBase – v.18 – Fev. 2012) (GRIFFITHS-JONES *et al.*, 2006). Diante da complexidade do processo da nodulação, identificar novos miRNAs que participam da nodulação e do processo de FBN pode auxiliar na elucidação sobre como esse processo é coordenado. Na simbiose com leguminosas, vários genes que codificam nodulinas já foram identificados, mas a chave para um maior entendimento sobre a regulação desses genes pode residir nos miRNAs ainda desconhecidos.

2.6 FERRAMENTAS PARA ESTUDO TRANSCRICIONAL

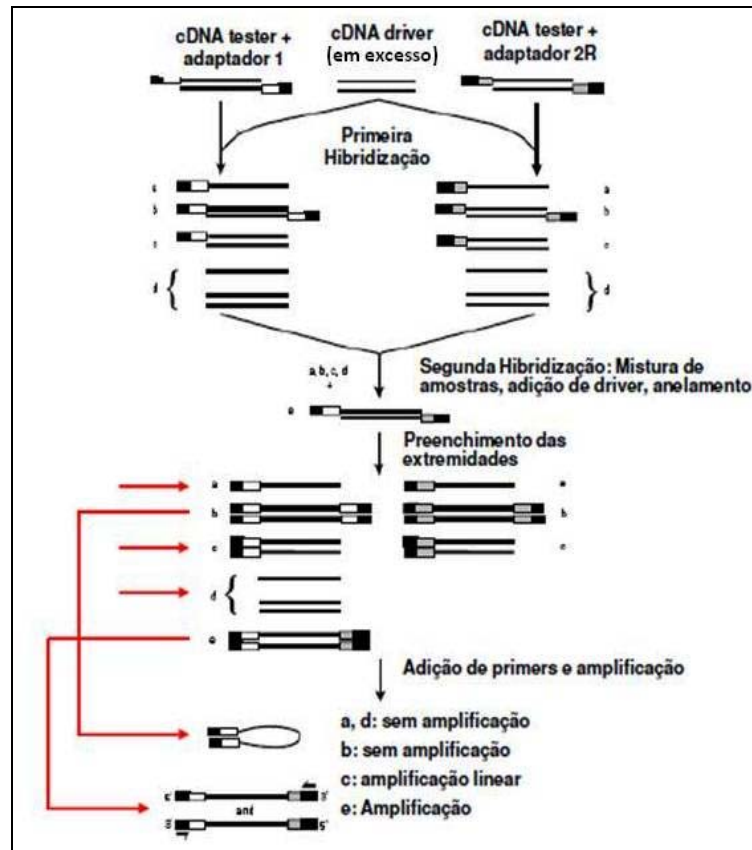
2.6.1 Hibridização Subtrativa Supressiva

Devido ao interesse em entender as alterações moleculares e fisiológicas que ocorrem na soja em resposta à inoculação com *B. japonicum* e à possibilidade de prospecção de novos genes envolvidos no processo, o uso da técnica hibridização subtrativa supressiva, que analisa a expressão diferencial de genes, representa uma boa alternativa.

Essa técnica foi desenvolvida por Diatchenko e colaboradores (1996) e se baseia na amplificação por PCR de sequências que são expressas diferencialmente, sob determinada condição específica, enquanto que a amplificação das outras sequências é suprimida. Uma das vantagens dessa técnica é a sua alta eficiência no isolamento de transcritos raros, tais como moléculas reguladoras (HINDERHOFER; ZENTGRAF, 2001).

Analisando o esquema apresentado na Figura 3, duas etapas dessa metodologia podem ser visualizadas. A primeira consiste no processo de hibridização subtrativa, que se caracteriza pela normalização da população de cDNAs correspondentes às moléculas de mRNAs (DIATCHENKO *et al.*, 1996). A segunda é denominada reação supressiva de polimerização em cadeia (*suppressive PCR*), onde apenas as moléculas diferencialmente expressas serão exponencialmente amplificadas, enquanto as demais são suprimidas (DIATCHENKO *et al.*, 1996).

Figura 3 – Esquema simplificado da hibridização subtrativa supressiva.



Fonte: Modificado de Diatchenko et al. (1996).

2.6.2 Sequenciamento de Nova Geração em Estudos de Transcriptoma

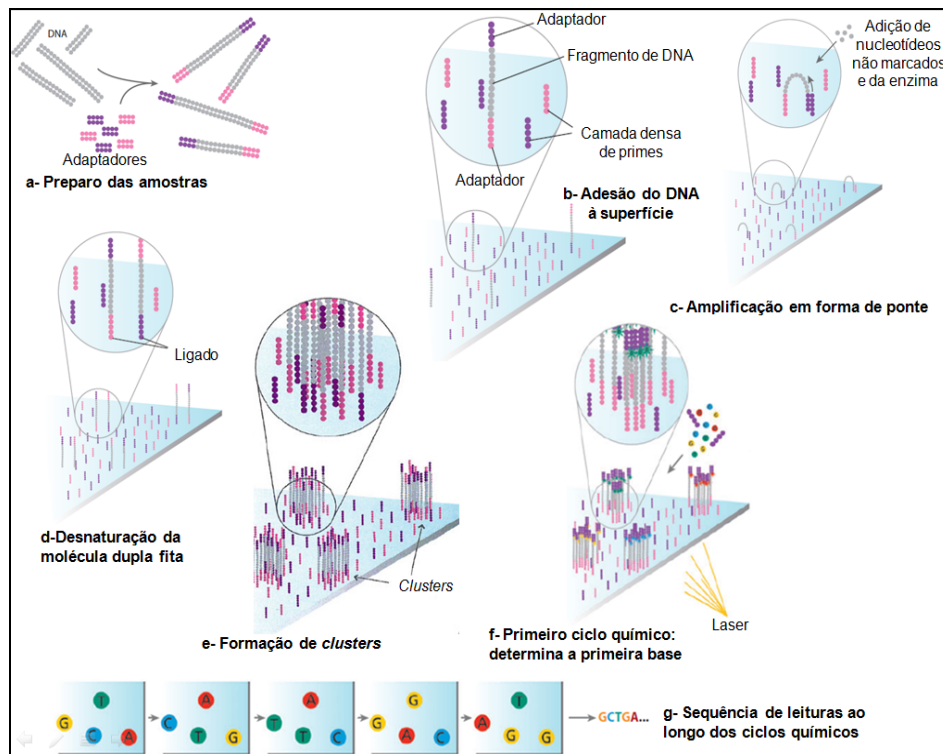
Estudos transcricionais são muitas vezes limitados pelo número de genes que podem ser pesquisados simultaneamente. Como alternativa, medidas diretas de expressão gênica podem ser obtidas por meio das plataformas de sequenciamento de nova geração. Através dessas novas tecnologias, é possível obter perfis transcricionais que levam a uma melhor compreensão dos elementos funcionais dentro do genoma em um dado momento, com maior capacidade de fornecer uma visão mais detalhada do transcriptoma (WANG *et al.*, 2010).

Entre as plataformas de sequenciamento disponíveis que apresentam ampla utilização destacam-se o 454/pirosequenciamento (RONAGHI *et al.*, 1998), o SOLID (*Sequencing by Oligonucleotide Ligation and Detection*) (McKERNAN *et al.*, 2008), o Illumina/Solexa Genome Analyzer (FEDURCO *et al.*, 2006; TURCATTI *et al.*, 2008), o Ion Torrent (ROTHBERG *et al.*, 2011), entre outras. Essas plataformas podem gerar grande quantidade de informação, permitindo que milhares de leituras (*reads*) sejam produzidas de uma só vez, com maior economia

de custo e de tempo, sendo que essa grande eficiência advém, principalmente, da capacidade de realizar clonagem *in vitro*, descartando o trabalho laboratorial intensivo na produção de clones bacterianos (MARDIS, 2008).

Na Figura 4 está representada a plataforma Illumina com sequenciamento similar ao método de Sanger, onde a síntese é realizada utilizando nucleotídeos terminadores marcados com diferentes fluoróforos.

Figura 4 – Representação esquemática das etapas de sequenciamento de nova geração da plataforma Illumina/Solexa.



Fonte: Modificado de Mardis (2008).

Uma inferência do nível de expressão gênica utilizando a plataforma Illumina é o valor de RPKM (*reads per kilobase of exon per million mapped reads*), onde a quantificação de transcritos é obtida através do número de leituras (concentração molar), que são mapeadas para as regiões éxon de um gene específico, e pelo comprimento da sequência (MORTAZAVI *et al.*, 2008; WANG *et al.*, 2010). Através do RPKM, torna-se mais fácil comparar o nível de expressão em diferentes tratamentos e entre diferentes genes (JIANG; WONG, 2009). Além disso, o uso de plataforma de sequenciamento de nova geração, combinado com análises de bioinformática, pode fornecer dados importantes na predição de novos genes.

3 OBJETIVO

Analisar o perfil de expressão gênica e prever potenciais microRNAs de raízes de soja (*Glycine max* (L.) Merrill), cultivar MG/BR46 – Conquista, inoculadas com *Bradyrhizobium japonicum* estirpe CPAC 15 (=SEMIA 5079), utilizada em inoculantes comerciais no Brasil.

3.1 OBJETIVOS ESPECÍFICOS

- Construir uma biblioteca subtrativa dos genes expressos diferencialmente em raízes da cultivar brasileira de soja, Conquista, inoculada com a estirpe CPAC 15 (=SEMIA 5079) de *Bradyrhizobium japonicum*, utilizada em inoculantes comerciais;
- Categorizar os genes diferencialmente expressos de acordo com sua funcionalidade, a fim de compreender as alterações moleculares que ocorrem na planta na presença de *Bradyrhizobium japonicum*;
- Fazer a predição *in silico* de novos miRNAs envolvidos na regulação da nodulação;
- Identificar computacionalmente genes alvos que possam ser regulados pelos miRNAs preditos.

4 MATERIAL E MÉTODOS

4.1 MATERIAL VEGETAL

Sementes de soja, cultivar MG/46-Conquista, foram cuidadosamente selecionadas e homogeneizadas. A seguir, foram desinfetadas superficialmente com uma solução de etanol 70% (1 min., para quebrar a tensão superficial) e, em seguida, com solução de hipoclorito de sódio 5% (3 min.), e lavadas cinco vezes com água destilada estéril (VINCENT, 1970). As sementes foram colocadas entre duas camadas de papel de germinação, umedecidas com água destilada estéril, e incubadas a $22 \pm 2^{\circ}\text{C}$ (no escuro), por três dias. As sementes germinadas foram selecionadas e cuidadosamente transferidas para sacos plásticos estéreis, contendo 200 mL de solução nutritiva isenta de nitrogênio (BROUGHTON; DILWORTH, 1970).

4.2 TRATAMENTOS

O ensaio consistiu de dois tratamentos: soja inoculada com *B. japonicum* estirpe CPAC 15 (=SEMIA 5079) e soja não inoculada (*mock-inoculated*, com água destilada estéril). O delineamento experimental foi inteiramente casualizado, com três repetições, cada uma contendo vinte plantas. A disposição das plantas no experimento pode ser visualizada na Figura 5.

Figura 5 – Plântulas de soja utilizadas no experimento.



4.3 PREPARO DO INÓCULO E INOCULAÇÃO

Para o crescimento da estirpe CPAC 15 de *B. japonicum* foi utilizado o meio de cultivo YMB (*yeast-mannitol broth*) (VINCENT, 1970) até a fase exponencial de crescimento. As células foram então centrifugadas e lavadas com solução salina (0,85% NaCl). Posteriormente, foi efetuada a contagem da suspensão de células lavadas em meio YMA (*yeast-manitol agar*) (VINCENT, 1970), revelando uma concentração de $2,27 \times 10^7$ células.mL⁻¹.

Para o tratamento de inoculação, alíquotas de 1 mL de células bacterianas foram inoculadas na base das radículas de soja pré-germinadas por 3 dias. As plantas cresceram em condições de casa de vegetação, com ciclo aproximado de 12 h/12 h (dia/noite) e a uma temperatura média de 28°C/18°C (dia/noite), por dez dias. As raízes foram então separadas do caule na altura do nó cotiledonar aos 10 dias após a inoculação (DAI), imediatamente colocadas em nitrogênio líquido para evitar degradação do material e estocadas à -80°C, até o momento de extração do RNA.

4.4 EXTRAÇÃO DE RNA E ISOLAMENTO DO MRNA

Para a construção da biblioteca subtrativa, as raízes de ambos os tratamentos, inoculadas e não inoculadas, foram maceradas separadamente em nitrogênio líquido, procedimento necessário para romper a parede celular vegetal e permitir a extração do RNA. O RNA total foi extraído utilizando o reagente Trizol[®] (Invitrogen), de acordo com as especificações do fabricante, com a aplicação de 1 mL para cada 100 mg de tecido. O procedimento foi realizado em ambiente desinfectado e em temperatura ambiente de $25 \pm 2^\circ\text{C}$. A avaliação da qualidade do RNA e sua quantificação foram verificadas em espectrofotômetro Nanodrop 1000 (Agilent).

O mRNA foi obtido a partir de 2 µg de RNA total, com o kit *Fast Track MAG mRNA Isolation* (Invitrogen), que usa partículas magnéticas conjugadas com oligos (dT) para capturar mRNA através da cauda poli (A+), de acordo com as especificações do fabricante.

4.5 CONSTRUÇÃO DA BIBLIOTECA SUBTRATIVA

A biblioteca resultante da hibridização subtrativa supressiva (HSS, *suppression subtractive hybridization*) foi construída utilizando o kit *PCR-Select cDNA Subtraction* (Clontech), de acordo com as instruções do fabricante. A biblioteca de cDNA foi construída utilizando a amostra de cDNA *tester* de plantas inoculadas, que foi subtraída do cDNA *driver* de plantas não inoculadas.

As etapas de construção da biblioteca subtrativa consistem, basicamente, de: (i) síntese da primeira e da segunda fita de cDNA; (ii) digestão dos fragmentos de cDNA com a endonuclease de restrição *RsaI*; (iii) ligação de adaptadores nas amostras *testers* de cDNA; (iv) primeira e segunda reação de hibridização; (v) amplificação do primeiro produto da hibridização por PCR (*Nested PCR*); e (vi) segunda amplificação por PCR. Esta última etapa permitiu um enriquecimento das sequências de cDNA diferencialmente expressas. Os produtos da biblioteca resultantes da reação de PCR foram encaminhados para sequenciamento em plataforma de nova geração.

4.6 SEQUENCIAMENTO, MONTAGEM DAS SEQUÊNCIAS E ANÁLISES DE BIOINFORMÁTICA

Os transcritos foram submetidos a 36 ciclos de sequenciamento através da tecnologia *Solexa Genome Analyzer IIe* (Illumina), que foi realizada pela empresa *Fasteris S.A.* (Suíça), produzindo “*single reads*” com cerca 76 pb.

As sequências geradas foram analisadas e montadas em colaboração com o LGE (Laboratório de Genômica e Expressão – UNICAMP, São Paulo). Primeiramente, as milhares de leituras (*reads*) obtidas no sequenciamento foram alinhadas no genoma de referência da soja (DOE Joint Genome Institute – *Phytozome*- <http://www.phytozome.net/soybean>) (SCHMUTZ *et al.*, 2010), utilizando o programa SOAP (Short Oligonucleotide Alignment Program, LI *et al.*, 2008), que permite no máximo dois *mismatches*. A seguir, procedeu-se à montagem dos *contigs* com o auxílio do programa EDENA (Exact DE Novo Assembler, HERNANDEZ *et al.*, 2008).

Além disso, para cada sequência presente na biblioteca subtrativa foi atribuída uma inferência de nível de expressão gênica (RPKM), que representa uma normalização do número de leituras, geradas no sequenciamento, que

alinham em um determinado gene. Assim, tornou-se possível comparar a expressão dos genes dentro da biblioteca.

Subsequentemente, o programa AutoFACT (KOSKI *et al.*, 2005) foi utilizado para a obtenção da anotação automática. Esse programa efetua buscas por similaridade de sequências contra diferentes bancos de dados, dentre eles o NCBI nrdb (non-redundant database *National Center of Biotechnology and Information*), o UniRef90 e o UniRef100 (WEIZHONG *et al.*, 2001), KEGG (*Kyoto Encyclopedia of Genes and Genomes*) (KANEHISA; GOTO, 2000) e Pfam (BATEMAN *et al.*, 2002). O uso do AutoFACT é de extrema importância, pois ao contrário do processo usual de anotação, que envolve busca manual de uma determinada sequência contra vários bancos de dados (KOSKI *et al.*, 2005), esta ferramenta permite diminuir o árduo trabalho da anotação manual, ainda mais difícil no caso de grandes conjuntos de sequências.

Posteriormente, as sequências resultantes da biblioteca subtrativa de raízes de soja inoculadas com *B. japonicum* foram depositadas no banco de dados do projeto GenoSoja, disponível no *website*: <http://bioinfo03.ibi.unicamp.br/soja>.

Finalmente, as sequências foram submetidas à categorização funcional, que foi realizada baseada no banco de dados do *Gene Ontology* (<http://www.geneontology.org>), através do programa Blast2GO (CONESA; GÖTZ, 2008), que permitiu agrupar os genes de acordo com a ontologia processo biológico (nível 2) e função molecular (nível 3), caracterizando os eventos moleculares desencadeados na planta pela presença da bactéria, além de permitir a identificação das principais rotas metabólicas (baseado no banco de dados do KEGG) ativadas nesta etapa inicial da nodulação.

4.7 PREDIÇÃO DE MICRORNAS ENVOLVIDOS NA NODULAÇÃO

Para a busca por miRNAs potenciais, os transcritos diferencialmente expressos foram submetidos a um *pipeline* de anotação geral de RNAs não codificantes (ncRNA) (PASCHOAL *et al.*, 2009), focando especificamente apenas nos resultados de miRNA. A anotação *in silico* analisou as sequências contra os principais bancos de dados de ncRNAs (ex.: miRBase, Rfam, NONCODE, RNAdb, ncRNAdb, fRNAdb, ASRP, piRNABank, TarBase, plant snoRNA database – veja

Paschoal *et al.* (2012) (<http://www.ncrnadatabases.org/>) para detalhes sobre os bancos de ncRNA); contra banco de dados de proteínas (Swiss-Prot - <http://www.uniprot.org/>); programas *ab initio* tRNAs [tRNAscan-SE - LOWE; EDDY (1997); ARAGORN - LASLETT; CANBACK (2004)]; snoRNA [*small nucleolar* RNA, como: SnoScan 1.0 - LOWE; EDDY (1999) e SnoReport - HERTEL *et al.* (2008)] e miRNA [HHMMIR - KADRI *et al.* (2009)]. Finalmente, a busca estrutural de famílias de RNAs foi realizada por *script cmsearch* do pacote de *software* INFERNAL (EDDY, 2002; NAWROCKI *et al.*, 2009), em conjunto com os dados do banco Rfam (versão 9.1) (GRIFFITHS-JONES *et al.*, 2003).

Com base no relatório final fornecido da anotação obtida pelo *pipeline*, o passo seguinte foi realizar uma anotação manual das sequências da biblioteca anotadas apenas como miRNA. Nesta etapa, o objetivo foi focar apenas nos miRNAs, deixando para uma posterior análise outras classes e sequências da biblioteca subtrativa.

No primeiro passo da anotação manual foi realizada uma busca contra os dados de miRNA do miRBase (versão 18), para identificar possíveis parálogos ou ortólogos. O relatório indicou, ainda, em qual região dos transcritos estariam os prováveis miRNAs, bem como o sentido da fita (senso/ antisenso) em que eles foram detectados. Essas informações auxiliaram na análise manual mais refinada das sequências.

Dessa forma, as sequências dos transcritos originais foram cortadas na região indicada pela anotação do *pipeline*, no caso, o resultado pela busca estrutural dos miRNAs com o programa INFERNAL no banco de dados Rfam (EDDY, 2002; NAWROCKI *et al.*, 2009). Em seguida, realizou-se uma comparação por similaridade (BLASTn) das sequências candidatas de pré-miRNA, com as sequências (*stem-loop*) depositadas no miRBase (GRIFFITHS-JONES *et al.*, 2006; KOZOMARA; GRIFFITHS-JONES, 2011).

Para fortalecer a hipótese de que as sequências identificadas nessa biblioteca poderiam ser processadas em miRNA maduro, os candidatos foram examinados de acordo com os seguintes critérios: *i-* a sequência do pré-miRNA deve formar uma estrutura secundária (*hairpin*); *ii-* baixa energia livre (a partir de -20 kcal/mol, segundo Thakur *et al.*, 2011); *iii-* o miRNA:miRNA* (duplex) deve estar inserido no “braço” da estrutura do pré-miRNA; *iv-* o pareamento entre o miRNA

senso e o miRNA antisense deve estar restrito a ≤ 4 *mismatches* (AMBROS *et al.*, 2003; JOSHI *et al.*, 2010; KULCHESKI *et al.*, 2011).

A predição das estruturas secundárias (*hairpin*) dos precursores de miRNAs foi realizada pelo *software* Mfold (ZUKER, 2003), utilizando os parâmetros padrões, uma análise baseada na formação de estrutura secundária e na energia livre mínima (ΔG) obtida. Outro aspecto analisado dos miRNAs foi a localização genômica (ex.: exon, intron, intergênica ou CDS) de cada sequência dentro do genoma da soja (Phytozome - SCHMUTZ *et al.*, 2010). Finalmente, a última etapa da anotação manual refinada foi por meio da análise de predição de possíveis novos miRNAs de soja envolvidos na nodulação, através da identificação computacional da sequência e da posição do miRNA maduro dentro do precursor predito, dado pelo programa MaturePred (XUAN *et al.*, 2011).

4.7.1 Predição de Alvos dos miRNAs

A predição de genes alvos de miRNAs em plantas é diferente daquela realizada para animais, pois apresentam uma complementaridade praticamente perfeita, senão perfeita. Dessa forma, os possíveis alvos regulados pelos miRNAs identificados e selecionados, neste estudo, foram preditos através do *software* psRNA Target (DAI; ZHAO, 2011). Foram utilizadas as sequências dos miRNAs maduros preditos para identificar os possíveis alvos presentes no banco de dados de *Glycine max* (DFCI gene index release 16). As sequências foram consideradas alvos quando o valor da pontuação foi inferior a 3,0, utilizando os parâmetros padrões do algoritmo. Os resultados dessas análises foram verificados manualmente no Phytozome, para identificar os genes e sua respectiva anotação.

5 RESULTADOS E DISCUSSÃO

5.1 CATEGORIZAÇÃO FUNCIONAL DOS GENES DIFERENCIALMENTE EXPRESSOS

Diante do interesse em analisar a resposta de uma cultivar brasileira de soja dez dias após a inoculação com a estirpe CPAC 15 (=SEMIA 5079), usada em inoculantes comerciais no Brasil, foi construída uma biblioteca subtrativa de cDNA (inoculadas com *B. japonicum* X não-inoculadas) das raízes de soja, onde o sequenciamento do *pool* de cDNA resultou em 4.621.072 leituras (*reads*). A montagem das *reads* gerou um total de 3.776 sequências diferencialmente expressas, que estão depositadas no banco de dados do GenoSoja (<http://bioinfo03.ibi.unicamp.br/soja>).

Do total de sequências apresentadas, 3.210 foram anotadas automaticamente utilizando o programa AutoFACT. Essas sequências foram agrupadas em categorias funcionais, de acordo com as ontologias “função molecular” (1.764 sequências) e “processo biológico” (1.728 sequências), através do *software* Blast2GO. É importante ressaltar que um gene pode estar incluso em mais de uma categoria, já que pode estar participando de mais de um processo biológico e com diferentes funções durante a nodulação. Na Figura 6.A, estão representados os papéis que esses genes diferencialmente expressos estão exercendo na planta, após a inoculação com *B. japonicum*. Basicamente, esses genes estão agrupados nas categorias moleculares de transporte, de ligação e atividade catalítica, o que demonstra a intensa atividade metabólica da planta durante a nodulação.

A Figura 6.B permite um maior entendimento quanto aos processos biológicos desencadeados no hospedeiro, 10 DAI com os rizóbios. Dentre esses, a categoria mais representativa, que agrupou as sequências diferencialmente expressas pela presença de *B. japonicum* nas raízes da soja foi a de “processo metabólico” (41%), seguido pela categoria de “resposta a estímulos”. Na Tabela 1 encontra-se uma descrição desses processos biológicos com maiores detalhes, indicando o número de sequência presentes em cada subcategoria.

Para uma discussão mais pontual dos resultados, algumas categorias foram abordadas e, dentre os diversos genes agrupados nos processos biológicos, alguns foram selecionados, com base no elevado nível de expressão gênica (RPKM), para uma abordagem mais detalhada.

Figura 6 – Porcentagem de sequências (%) alocadas em **(A)** funções moleculares e **(B)** processos biológicos (segundo classificação do *Gene Ontology*) de genes diferencialmente expressos em raízes de soja na presença da bactéria fixadora de nitrogênio *Bradyrhizobium japonicum*, estirpe CPAC15 (=SEMIA 5079).

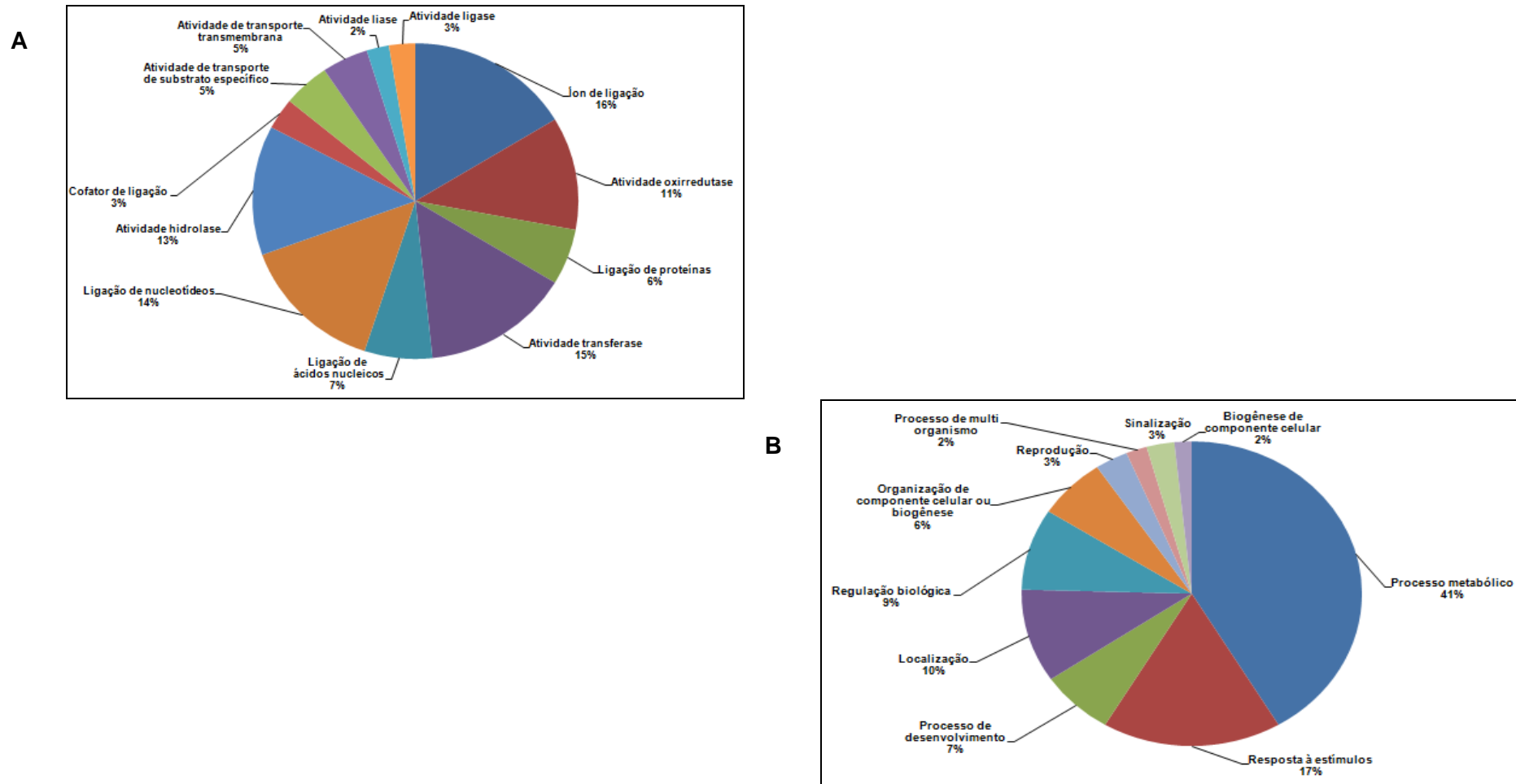


Tabela 1 – Classificação dos processos biológicos desencadeados nas raízes da soja em resposta à inoculação da estirpe CPAC 15 de *B. japonicum*.

ID GO	Descrição dos Processos Biológicos	Nº. Seq.
GO:0008150	Processo biológico	1.728
GO:0008152	I- Processo metabólico	1.290
GO:0044237	Processo celular metabólico	964
GO:0044238	Processo metabólico primário	756
GO:0019748	Processo metabólico secundário	62
GO:0006807	Processo metabólico de compostos nitrogenados	269
GO:0009058	Processo biossintético	373
GO:0009056	Processo catabólico	195
GO:0044281	Processo metabólico de pequenas moléculas	324
GO:0043170	I.a- Processo metabólico de macromolécula	625
GO:0019538	Processo metabólico de proteína	337
GO:0010467	Expressão gênica	69
GO:0050896	II- Resposta a estímulos	529
GO:0070887	Resposta a estímulos químicos	246
GO:0009628	Resposta a estímulos abióticos	162
GO:0009607	Resposta a estímulos bióticos	56
GO:0006950	II.a- Resposta a estresse	289
GO:0006952	Resposta de defesa	68
GO:0009719	II.b- Resposta a estímulos endógenos	113
GO:0009725	Resposta a estímulos hormonais	106
GO:0032502	III- Processo de desenvolvimento	218
GO:0007275	Desenvolvimento de organismo multicelular	124
GO:0048856	III.a- Desenvolvimento de estrutura anatômica	114
GO:0048513	Desenvolvimento de órgão	73
GO:0051179	IV- Localização	308
GO:0051641	Localização celular	56
GO:0033036	Localização de macromolécula	60
GO:0051234	IV.a- Estabelecimento de localização	306
GO:0006810	Transporte	265
GO:0065007	V- Regulação biológica	269
GO:0050789	Regulação de processo biológico	208
GO:0065008	Regulação da qualidade biológica	80
GO:0071840	VI- Organização de componente celular ou biogênese	203
GO:0071841	Organização de componente celular ou biogênese em nível celular	136
GO:0016043	Organização de componente celular	171
GO:0044085	VII- Biogênese de componente celular	51
GO:0000003	VIII- Reprodução	97
GO:0051704	IX- Processo de multi organismo	62
GO:0051707	Resposta a outro organismo	51
GO:0023052	X- Sinalização	82
GO:0007165	Transdução de sinal	82

5.1.1 Processo Metabólico

Conforme relatado em alguns trabalhos conduzidos sobre o efeito da inoculação de *B. japonicum* em soja, nos diferentes estágios da nodulação, variando entre horas (12, 24 e 48), dias (4, 8, 16) e semanas (2, 5, 10) após a inoculação (LEE *et al.*, 2004; BRECHENMACHER *et al.*, 2008; LIBAULT *et al.*, 2010), a simbiose rizóbio-leguminosa provoca profundas alterações no metabolismo do vegetal, variável, conforme o estágio analisado (DESBROSSES; STOUGAARD, 2011).

Este trabalho aponta as respostas gerais da planta durante a nodulação, sendo pertinente apresentar as vias metabólicas, que se mostraram mais ativas na soja na presença do rizóbio: a via glicolítica (Anexo A) e o ciclo de Krebs (Anexo B). Com o desenvolvimento e funcionamento dos nódulos é necessária a realocação de fontes de C da planta para o novo órgão (COLEBATCH *et al.*, 2004). El Yahyaoui *et al.* (2004) constataram que os genes envolvidos no processo de suprimento de energia mostram-se *up-regulated* durante o estágio inicial da nodulação em *Medicago truncatula*.

A via glicolítica e o ciclo de Krebs estão intimamente ligados, pois são as principais vias para a obtenção de energia. Para que os nódulos sejam formados, muitos processos são induzidos, como o crescimento da raiz, indução da divisão celular, entre outros, provocando um alto gasto energético. Por isso, os genes envolvidos no processo metabólico têm seu nível de expressão incrementado e, conforme observado por Brechenmacher *et al.* (2008), isso ocorre a partir do oitavo dia após a inoculação, uma constatação que se confirma com este trabalho.

No presente estudo foram identificados os genes de soja que correspondem a algumas enzimas que participam da glicólise (Tabela 2.A). Dentre eles, o que apresentou maior nível de expressão, aos 10 DAI, foi o Glyma14g36850.1 (frutose-bifosfato aldolase), apontando a suma importância desta enzima no decorrer da nodulação. Esta enzima participa da produção de ácidos dicarboxílicos para os rizóbios e esqueletos de carbono para a assimilação do N (EL YAHYAOUÏ *et al.*, 2004).

Da mesma forma, foram identificados os genes da soja que codificam enzimas envolvidas no ciclo de Krebs (Tabela 2.B), diferencialmente expressos após a inoculação com a bactéria fixadora de nitrogênio. Esses genes

são essenciais para atender à demanda energética elevada para a formação dos nódulos e, ainda, para a obtenção de esqueletos de carbono necessários à assimilação do N.

Tabela 2 – Genes que codificam enzimas presentes na via glicolítica (A) e no ciclo de Krebs (B) diferencialmente expresso durante a nodulação.

A	Número EC	Descrição	Gene	*RPKM	Nº. reads
	1.1.1.1	Álcool desidrogenase	Glyma06g12780.3	1.582,95	3.624
			Glyma05g35340.2	1.511,68	1.303
			Glyma04g41990.1	972,077	2.716
	1.2.1.3	Aldeído desidrogenase (NAD ⁺)	Glyma18g18910.1	637,658	1.741
	1.2.1.5	Aldeído desidrogenase (NAD(P) ⁺)	Glyma04g42740.1	452,563	554
	1.2.1.12	Gliceraldeído 3-fosfato desidrogenase	Glyma06g18110.6	1.444,8	335
			Glyma06g18110.6	706,073	2.376
			Glyma06g18120.1	564,17	2.823
	1.2.4.1	Piruvato desidrogenase	Glyma14g36540.3	1.228,86	2.347
	1.8.1.4	Didrolipoamida desidrogenase	Glyma07g36040.1	1.059,63	442
			Glyma17g04210.1	379,647	1.218
	2.7.1.1	Hexoquinase	Glyma11g01820.1	118,426	34
	2.7.1.11	6-fosfofrutoquinase	Glyma20g01010.1	305,958	195
	2.7.1.40	Piruvato quinase	Glyma10g07480.2	539,465	751
	2.7.2.3	Fosfoglicerato quinase	Glyma15g41550.1	1.145,85	3.074
	3.1.3.11	Frutose-1,6-bifosfatase I	Glyma07g17180.1	162,054	90
	4.1.1.1	Piruvato descarboxilase	Glyma13g30490.1	241,564	616
	4.1.1.49	Fosfoenolpiruvato carboxiquinase	Glyma01g02330.1	365,184	421
	4.1.2.13	Frutose-bifosfato aldolase	Glyma14g36850.1	2.293,45	4.469
	4.2.1.11	Enolase	Glyma19g37520.1	1.027,5	3.759
			Glyma03g34830.1	341,075	2.017
	5.3.1.1	Triosefosfato isomerase	Glyma15g04290.2	1.544,91	1.812
	5.4.2.2	Fosfoglucomutase	Glyma05g34790.1	1.159,25	2.329
	6.2.1.1	Acetil-CoA sintetase	Glyma15g25170.1	1.174,42	655
B	Número EC	Descrição	Gene	*RPKM	Nº. reads
	1.1.1.37	Malato desidrogenase	Glyma10g00920.2	280,322	723
	1.1.1.41	Isocitrato desidrogenase (NAD ⁺)	Glyma10g06590.2	240,644	217
	1.1.1.42	Isocitrato desidrogenase	Glyma14g39160.3	525,736	1.323
	1.2.4.1	Piruvato desidrogenase	Glyma14g36540.3	1.228,86	2.347
	1.8.1.4	Didrolipoamida desidrogenase	Glyma07g36040.1	1.059,63	442
			Glyma17g04210.1	379,647	1.218
	2.3.3.8	ATP citrato liase	Glyma09g04000.1	1.412,31	988
			Glyma15g42140.2	969,477	1.575
			Glyma09g04000.1	818,764	2.373
	4.1.1.49	Fosfoenolpiruvato carboxiquinase	Glyma01g02330.1	365,184	421
	4.2.1.3	Aconitato hidratase	Glyma12g32000.1	971,002	1.523
	6.2.1.5	Succinil-CoA sintetase	Glyma09g29460.1	348,288	792

* Reads per kilobase of exon per million mapped reads

Portanto, a identificação dos genes que estão participando do metabolismo primário durante a nodulação é relevante, pois podem auxiliar na seleção de cultivares de soja, eficiente na produção dos fotoassimilados, podendo, conseqüentemente, incrementar a FBN, visto que, o metabolismo do C é a base para o processo de FBN (VANCE; GANTT, 1992; CURIONI *et al.*, 1999).

Dentre os processos metabólicos, também foi possível observar que a nodulação provoca alterações no estado redox nas células das raízes, pois entre as principais vias metabólicas ativas na leguminosa, está a via da glutathione (Anexo C), que protege as células contra a ação de espécies reativas de oxigênio (ROS) (FRENDO *et al.*, 2005). ROS são moléculas que apresentam radicais livres, uma propriedade instável que leva a uma alta reatividade. Essas moléculas são geradas devido à intensa atividade metabólica, que causa modificação na distribuição de elétrons (BECANA *et al.*, 2010).

Conforme apresentado por Chang e colaboradores (2009), essas alterações no estado redox são observadas nos diferentes estágios da nodulação. Isso também foi observado durante o início da interação simbiótica em *Medicago sativa* (SANTOS *et al.*, 2001) e em *Phaseolus vulgaris* (CÁRDENAS *et al.*, 2008). Os transcritos presentes na biblioteca subtrativa, que participam do sistema de defesa antioxidante (via da glutathione), são apresentados na Tabela 3.

Tabela 3 – Genes da soja que codificam as enzimas que estão atuando na proteção celular contra os danos oxidativos.

Número EC	Descrição	Gene	*RPKM	Nº. reads
1.1.1.42	Isocitrate desidrogenase	Glyma14g39160.3	525,736	1.323
1.1.1.44	6-fosfogluconato desidrogenase	Glyma08g28230.1	366,492	616
1.1.1.49	Glucose-6-fosphate 1-desidrogenase	Glyma19g24250.1	393,621	477
1.11.1.9	Glutathione peroxidase	Glyma05g37900.3	3.425,42	2.444
		Glyma05g34490.4	699,904	1.713
1.8.1.7	Glutathione redutase	Glyma16g27210.1	590,971	658
		Glyma10g03740.3	561,325	391
2.3.2.2	Gama-glutamyltransferase	Glyma11g35950.1	1.631,78	1.058
2.5.1.18	Glutathione S-transferase	Glyma01g04690.1	1.392,19	344
		Glyma06g20730.1	1.102,13	1.774
		Glyma07g16910.1	806,678	1.954
		Glyma02g02860.1	1.013,09	1.019
		Glyma08g41960.1	693,606	1.347
6.3.2.2	Gama-glutamylcisteína sintetase	Glyma05g37850.3	326,118	410
6.3.2.3	Glutathione sintase	Glyma19g42600.1	363,132	148
		Glyma03g40050.1	285,275	213

* Reads per kilobase of exon per million mapped reads

Estes genes são importantes por evitar os danos celulares que as moléculas de ROS possam causar. Dentre estes, destaca-se os genes que codificam a glutathione peroxidase e a glutathione S-transferase, que parecem apresentar um papel central nesse processo, pois tiveram níveis de expressão bastante elevado, comparado aos outros genes presentes nessa via metabólica.

A glutathione, além de controlar a alteração redox, pode participar da regulação da divisão celular nas raízes, conforme demonstrado em *Arabidopsis* (SANCHEZ-FERNANDEZ *et al.*, 1997); conseqüentemente, é fundamental para o desenvolvimento dos nódulos. Em *Medicago truncatula* também foi comprovado o papel crítico da glutathione na nodulação, pois a inibição na síntese da mesma resultou na redução do número de nódulos formados nas raízes inoculadas (FRENDON *et al.*, 2005). Mais recentemente, Pucciarello e colaboradores (2009) mostraram que a diminuição da expressão da (homo) glutathione, durante a interação simbiótica entre *M. truncatula* e *S. meliloti*, é capaz de modular a expressão dos genes da planta envolvidos na nodulação, como no caso da expressão de aquaporinas, que na presença da glutathione, apresentam-se *up-regulated*, e na falta da glutathione passaram a ser *down-regulated*.

Com isso, este estudo confirma a importância da glutathione na nodulação e, ainda, aponta os genes que codificam as enzimas glutathione peroxidase, gama-glutamyltransferase e glutathione S-transferase como enzimas reguladoras essenciais desse sistema antioxidante das raízes da soja.

Entre os genes agrupados nessa categoria, também tiveram destaque, com base no elevado nível de expressão (RPKM), os genes que codificam a metionina sintase, a S-adenosylmetionina sintetase (SAMS) e a ácido 1-carboxílico-1 aminociclopropano oxidase (ACC) (Tabela 4), que participam do metabolismo de aminoácidos e da via de biossíntese do etileno (WANG *et al.*, 2002), apontando a presença do etileno na regulação da nodulação. Esses genes apresentaram nível de expressão elevado na presença de *B. japonicum*, principalmente o Glyma16g04240.3 (metionina sintase) e Glyma17g04340.1 (SAMS) (Tabela 4). Oehrle *et al.* (2008) identificaram a presença da metionina sintase nos nódulos de soja, após 4 semanas de inoculação, codificada pelo mesmo gene identificado neste estudo aos 10 DAI, demonstrando que esse gene é expresso tanto durante períodos iniciais da nodulação, quanto em nódulos já formados. Além disso, a importância da

SAMS na nodulação deve ser ainda salientada, visto que três diferentes genes SAMS com alto nível de expressão foram identificados (Tabela 4).

Tabela 4 – Genes selecionados, com base no RPKM, a partir dos processos biológicos induzidos nas raízes de soja inoculadas com *Bradyrhizobium*.

Processos biológicos	ID genes	Nome	*RPKM	N°. reads
Biossíntese	Glyma16g04240.3	Metionina sintase	1.752,17	10.747
	Glyma07g36150.1	S-adenosilmetionina sintetase	1.648,07	5.433
	Glyma15g21890.2	S-adenosilmetionina sintetase	2.376,08	5.565
	Glyma17g04340.1	S-adenosilmetionina sintetase	5.050,72	14.038
	Glyma14g05390.1	Ácido 1-carboxílico-1 aminociclopropano	1.710,30	611
Sinalização	Glyma16g29220.1	LRR (<i>Leucine rich repeat</i>)	210.781	158.479
	Glyma03g32700.2	LRR (<i>Leucine rich repeat</i>)	1.943,56	1.149
	Glyma13g03910.1	Calmodulina	2.424,67	1.940
	Glyma05g13900.1	Calmodulina	1.743	2.510
	Glyma20g10820.1	Calmodulina	1.470,50	2.522
Resposta a estímulos	Glyma20g23080.1	Calreticulina	1.823,09	5.531
	Glyma15g02220.1	PDR (<i>Pleiotropic drug resistance</i>)	1.735,93	384
	Glyma08g29090.1	Glicosídeo hidrolase 19	1.177,12	1.319
	Glyma18g51980.1	Glicosídeo hidrolase 19	1.027,26	1.122
Localização	Glyma20g32000.1	Aquaporina	1.377,30	3.533
	Glyma10g35520.2	Aquaporina	1.613,69	3.245
	Glyma14g06680.5	Aquaporina	2.989,98	9.195
	Glyma12g29510.2	Aquaporina	1.744,05	4.347
	Glyma12g08040.1	Aquaporina	5.022,48	5.497
	Glyma05g32620.1	ABC transporte	1.834,50	531
	Glyma08g00280.1	ABC transporte	1.053,41	949
	Glyma08g28190.1	Transportador de aminoácido	1.734,53	4.473
	Glyma18g51220.1	Transportador de aminoácido	2.511,83	5.376
	Glyma20g33000.1	Transportador de aminoácido	1.273,41	1.325
	Glyma10g32750.1	Transportador de oligopeptídeo	2.462,59	5.684
		Transportador de potássio	6.052,79	4.615
	Glyma19g45260.1	Transportador de potássio	1.865,16	1.468
		Transportador de potássio	1.636,90	2.268
	Transportador de potássio	2.088,80	3.890	
	Glyma03g31310.1	Transportador de sulfato	1.421,28	1.314
Ciclo celular	Glyma05g32220.2	Actina	2.417	1.069,31
	Glyma04g39380.2	Actina	2.410	1.038,68
	Glyma12g08760.1	Ciclina dependente quinase	1.601	1.612,16
	Glyma10g40150.1	α -Tubulina	1.951	3.090,34
	Glyma20g27280.1	α -Tubulina	1.461	1.620,73
Regulação	Glyma20g27950.1	Ubiquitina	1.345,02	18.694
	Glyma01g38470.3	Ubiquitina	1.396,14	2.254

* Reads per kilobase of exon per million mapped reads

5.1.2 Sinalização

Os transcritos diferencialmente expressos agrupados nas categorias de “processo de multi organismo” (interação entre organismos) e “sinalização” (Figura 6) desempenham um papel relevante na interação molecular entre o microssimbionte e a planta hospedeira (COOPER, 2007; SUBRAMANIAN *et al.*, 2007). Essas categorias incluem genes envolvidos na cascata de transdução de sinais devido a essa interação (LIMPENS; BISSELING, 2003), levando ao desenvolvimento coordenado da epiderme e ao desenvolvimento cortical, necessários para permitir a infecção bacteriana e o desenvolvimento nodular (DING; OLDROYD, 2009).

Sabe-se que os processos de reconhecimento e sinalização, desencadeados na planta pela presença do rizóbio, provocam ativação de genes que participam da nodulação, muitos dos quais são bem caracterizados em soja, como por exemplo, os receptores quinases NARK (*nodule autoregulation receptor kinase*) (SEARLE *et al.*, 2003) e NORK (*nodulation receptor kinase*) (ZHU *et al.*, 2006), proteínas quinases do tipo LRR (*Leucine Rich Repeat*), transcritas, respectivamente, pelos genes Glyma12g04390 e Glyma09g33510. Libault e colaboradores (2010) também identificaram outros genes referentes às LRRs diferencialmente reguladas durante a nodulação.

Além da soja, proteínas LRRs já foram identificadas nas raízes de *Lotus japonicus* (STRACKE *et al.*, 2002), de *Medicago truncatula* e de *Sesbania rostrata* inoculadas com seus respectivos simbiontes específicos (CAPOEN *et al.*, 2005). Dentre os genes identificados neste estudo, pelo maior nível de expressão, encontram-se dois genes que codificam LRRs (Tabela 4), sendo destacado o Glyma16g29220.1, por ter apresentado o maior nível de expressão (RPKM), entre todos os genes presentes na biblioteca subtrativa. Com isso, esse estudo apresenta um novo gene, que pode ser tão importante quanto os receptores citados anteriormente, indicando ser importante para a regulação da nodulação.

Além dos genes já mencionados, é pertinente apontar aqueles que codificam as calmodulinas (Tabela 4), proteínas muito importantes, que participam da transdução de sinais desencadeada pela interação dos fatores Nod com receptores específicos da superfície dos pelos radiculares, onde durante as oscilações de cálcio intracelular (Ca^{2+} *spiking*), vinculam-se ao Ca^{2+} , sofrendo uma

mudança conformacional e induzem os fatores transcricionais da nodulação, permitindo a expressão das nodulinas (OLDROYD; DOWNIE, 2004). O nível de expressão desses genes foi bastante elevado. Os genes das calmodulinas mais comuns, identificados durante a nodulação em soja são os Glyma15g35070.1 e Glyma08g24360.1 (SCHMUTZ *et al.*, 2010), e aqui são apresentados três genes (Glyma05g13900.1, Glyma13g03910.1, Glyma20g10820.1), cujas participações na nodulação não havia sido identificada até então.

5.1.3 Resposta a Estímulos

A segunda categoria com maior número de transcritos foi a de “resposta a estímulos”. Compreende os genes referentes a estímulos abióticos, bióticos, genes de resposta a estresse, resposta de defesa e de resposta a estímulos endógenos. Esses genes estão envolvidos nas mudanças da atividade celular, alterações morfológicas, bioquímicas e fisiológicas estimuladas pela inoculação com *B. japonicum*. Em meio a tantos genes importantes para a nodulação, encontram-se genes que codificam a calreticulina, a proteína PDR (*pleiotropic drug resistance*) e as glicosídeo hidrolases 19 (Tabela 4).

As calreticulinas são proteínas localizadas, predominantemente, no retículo endoplasmático e estão presentes na maioria dos tecidos. Elas apresentam grande diversidade funcional, pois participam do dobramento de proteínas (chaperonas), da homeostase intracelular de Ca^{2+} , da sinalização e também são cruciais para o crescimento e desenvolvimento das plantas. Além disso, são potentes reguladoras na resposta da planta a vários tipos estresses, como exemplo, participam da ativação dos mecanismos de resistência contra infecção de patógenos (CHRISTENSEN *et al.*, 2008; JIA *et al.*, 2009; GUPTA; TUTEJA, 2011).

Calreticulinas são bastante estudadas em animais, em contrapartida, nos vegetais, conhecimentos específicos sobre funções dessas proteínas ainda são limitados (JIA *et al.*, 2009). O presente trabalho mostra que essas proteínas são requeridas para a nodulação, parecendo fazer parte de alguma forma a resposta de defesa da planta, pois de acordo com a categorização realizada pelo Blast2GO, estas estão presentes nas subcategorias de “resposta a estímulos bióticos” e “resposta de defesa”. Genes que codificam calreticulinas foram diferencialmente expressos na presença do rizóbio, dentre eles, destacou-se o Glyma20g23080.1,

com nível elevado de expressão (Tabela 4). Não foram encontrados trabalhos que relatam a função e a importância destas proteínas, especificamente, durante a nodulação, por isso, estudos posteriores com o gene apresentado aqui poderiam levar a um maior esclarecimento.

Outra proteína induzida durante a nodulação soja-*Bradyrhizobium* identificada, neste trabalho, foi a PDR (Tabela 4), pertencente à família ABC transporte, presente em fungos e plantas, envolvidas na resposta de estresse biótico e abiótico (CROUZET *et al.*, 2006). Essas proteínas são apontadas como transportadores envolvidos na secreção de genisteína nas raízes de soja em resposta a *B. japonicum* (SUGIYAMA *et al.*, 2008), uma informação que pode ser fortalecida com este estudo, pois genes que codificam tais proteínas apresentaram-se diferencialmente expressos na presença dessa bactéria fixadora de N.

Os genes que codificam glicosídeo hidrolases da família 19 são importantes no processo de nodulação e foram identificados dentre os genes diferencialmente expressos com alto nível de expressão (Tabela 4). Essas enzimas são conhecidas como quitinases (EC:3.2.1.14) e, dentre outras funções na simbiose, acredita-se que desempenhem um papel de controle durante a nodulação, pois parecem participar da percepção e degradação dos fatores Nod, moléculas de oligossacarídeos lipoquitínicos enviadas à planta pelos rizóbios (KASPRZEWSKA, 2003; SANTOS *et al.*, 2008). Xie *et al.* (1999) mostraram um aumento na atividade das quitinases na presença dos fatores Nod de rizóbios, nas raízes de soja, sugerindo, que estas enzimas possam atuar num processo de *feedback* da nodulação.

5.1.4 Transporte

Este estudo também ressalta a importância do processo biológico “localização”, que corresponde a cerca de 10% do total de genes expressos diferencialmente nessa biblioteca. É um processo que engloba, dentre outros, os genes que codificam proteínas transportadoras de membranas. Na Tabela 4 estão apresentados os genes representantes dessa categoria que tiveram os maiores níveis de expressão. Dentre estes, estão os genes que codificam aquaporinas, proteínas de membrana responsáveis pelo transporte, principalmente, de água, que têm um elevado nível de expressão nos tecidos em crescimento, neste caso, as

raízes, devido à necessidade de influxo de água para dentro das células em expansão (KJELLBOM *et al.*, 1999; JOHANSSON *et al.*, 2000; UEHLEIN *et al.*, 2007). Os genes de soja codificadores de aquaporinas mais encontrados na nodulação são Glyma08g12650 e Glyma19g22210, Nodulina-26, encontrada no simbiossomo (LIBAULT *et al.*, 2010; SCHMUTZ *et al.* 2010). Entretanto, neste estudo, foram apontados outros genes que codificam aquaporinas, induzidos nesse processo (Tabela 4). Durante a nodulação em *Medicago truncatula*, as aquaporinas mostraram-se como as proteínas de membrana mais expressas, desde os estágios iniciais, até os estágios mais tardios (EL YAHYAOUÏ *et al.*, 2004), o que comprova a importância destas no decorrer da formação dos nódulos.

As demais proteínas transportadoras diferencialmente reguladas (Tabela 4) também apresentam papéis importantes na nodulação, pois podem estar participando do transporte de nutrientes entre as células das raízes e, ainda, entre as células da planta com o simbiossomo, além do processo de sinalização em resposta ao estímulo imposto às raízes, entre outros. Em meio aos genes que codificam essas proteínas, os que apresentaram os maiores níveis de expressão foram os correspondentes aos transportadores ABC, transporte de aminoácido, de oligopeptídeo, de potássio e transporte de sulfato. Benedito *et al.* (2006) apresentaram a importância desses transportadores presentes nos nódulos de leguminosas. Além disso, as famílias de proteínas transportadoras abordadas neste estudo com soja, como diferencialmente expressas, também estão entre as proteínas transmembrana mais abundantes na nodulação em *M. truncatula* (BENEDITO *et al.*, 2010). Observa-se que a presença dessas proteínas de membrana são muito requeridas, assim como previamente mostrado por Brechenmacher *et al.* (2008) e Libault *et al.* (2010).

5.1.5 Ciclo Celular e Regulação de Processos Biológicos

O processo de nodulação envolve a indução da divisão celular e o rearranjo do citoesqueleto, principalmente no córtex da raiz, causando alterações no ciclo celular, processos necessários para acomodar os rizóbios (BLANCAFLOR *et al.*, 2006; SCHMIDT; PANSTRUGA, 2007; OLDROYD *et al.*, 2011). Essa regulação que ocorre no ciclo celular durante a organogênese do nódulo foi detalhada em *Medicago* (FOUCHER; KONDOROSI, 2000). Em concordância com estes trabalhos,

no presente estudo, foram identificados genes que codificam actinas, α -tubulinas e ciclina dependente quinase (CDK), diferencialmente expressos em raízes de soja em resposta ao *B. japonicum* (Tabela 4).

As CDKs controlam a progressão do ciclo celular e os genes correspondentes às actinas e às tubulinas estão relacionados ao citoesqueleto celular. Estes representam suma importância para a formação dos nódulos, pois estão coordenando a divisão e a diferenciação celular em resposta ao microssimbionte. Um estudo com *Lupinus albus* apresentou a mudança que ocorre na conformação do citoesqueleto para acomodar *Bradyrhizobium* sp. durante a nodulação (FEDOROVA *et al.*, 2007), confirmando a importância dos transcritos identificados neste estudo.

Em dez dias pós-inoculação com rizóbios, os primórdios nodulares das raízes estão formados, e no processo de “regulação biológica” (Figura 6), estão os genes que participam, principalmente, da regulação das biomoléculas envolvidas no estabelecimento da simbiose e, conseqüentemente, na formação dos nódulos. Dentre esses, neste estudo, foram selecionados os genes que codificam as ubiquitinas (Tabela 4), proteínas que participam do processo de regulação pós-traducional, como a degradação seletiva de muitas proteínas que atuam na atividade celular, tais como progressão do ciclo celular, resposta ao estresse, na transdução de sinais, regulação da transcrição, entre outros. Portanto, apresentam extrema importância, pois auxiliam nos processos regulatórios, necessários na nodulação (WEISSMAN, 2001; KONDOROSI *et al.*, 2005).

5.1.6 Modificação da Parede Celular

Diante da necessidade de reorganização estrutural da raiz para permitir a infecção dos rizóbios, os genes relacionados ao processo de “organização de componentes celulares” (Figura 6) participam, dentre outros processos, da modificação da parede celular vegetal e da reorganização do citoesqueleto das células da raiz. Apesar do número elevado de genes envolvidos com a modificação da parede celular da planta durante a nodulação, foi dado destaque aqueles presentes na via metabólica de carboidratos (Anexo D).

O metabolismo de carboidratos representa especial importância na nodulação, pois os transcritos que codificam tais enzimas, ativas nessa via

metabólica, podem estar atuando especificamente na reorganização da raiz e na formação da estrutura nodular, estes estão apresentados na Tabela 5.

Tabela 5 – Genes da soja que codificam as enzimas ativas identificadas a partir da via de metabolismo de carboidratos.

Número EC	Descrição	Gene	*RPKM	Nº. reads
1.1.1.22	UDP-glucose 6-desidrogenase	Glyma01g06970.1	1.118,45	2.934
2.4.1.12	Celulose sintase	Glyma05g32100.1	1.225,43	1.688
		Glyma06g07320.2	555,591	4.585
2.4.1.13	Sucrose sintase	Glyma15g20180.3	1.552,25	5.024
2.4.1.25	4-alfa-glucanotransferase	Glyma03g27600.1	160,692	105
2.4.1.34	Calose sintetase	Glyma06g44770.1	441,939	201
2.7.1.1	Hexoquinase	Glyma11g01820.1	118,426	34
3.1.1.11	Pectinesterase	Glyma03g03460.1	2.433,21	2.497
		Glyma01g45110.1	874,835	1.157
3.2.1.15	Poligalacturonase	Glyma05g37490.1	761,534	1.405
3.2.1.2	Beta-amilase	Glyma06g45700.1	189,028	391
3.2.1.26	Invertase	Glyma20g31730.1	677,598	1.994
3.2.1.39	Endo-1,3-beta-D-glucosidase	Glyma03g28850.1	2.612,62	6.736
3.2.1.4	Celulase	Glyma05g36930.1	608,64	417
4.1.1.35	UDP-glucuronate descarboxilase	Glyma12g06980.3	193,024	268
5.4.2.2	Fosfoglucomutase	Glyma05g34790.1	1.159,25	2.329

* Reads per kilobase of exon per million mapped reads

Além disso, o estudo de Salah *et al.* (2009) mostrou que, em *Medicago ciliaris*, o metabolismo de carboidratos é ainda mais relevante na manutenção da fixação de nitrogênio sob condições de estresse. De um modo geral, os transcritos que codificam as enzimas pertencentes a esse processo, identificados neste estudo, atuam promovendo a reorganização da parede celular, considerando que a entrada do simbiote provoca a expansão da parede celular para a formação do cordão de infecção e dos nódulos (OLDROYD *et al.*, 2011), razão pela qual tal processo metabólico é induzido durante a nodulação (EL YAHYAOUÏ *et al.*, 2004; LOHAR *et al.*, 2006; BRECHENMACHER *et al.*, 2008; LIBAULT *et al.*, 2010).

Em um estudo conduzido por Kaewsuralikhit *et al.* (2005), uma pectinesterase (*putative*) apresentou maior nível de expressão, em soja, aos 12 DAI, sendo uma das enzimas responsáveis pela degradação da parede celular durante a fase de formação dos nódulos, o que também ocorreu em *Sesbania rostrata* (LIEVENS *et al.*, 2002). Em *Medicago truncatula*, a pectinesterase também se apresentou *up-regulated* e a celulase foi induzida no terceiro e quarto dia pós-inoculação (EL YAHYAOUÏ *et al.*, 2004).

Outra importante enzima, com alto nível de expressão neste estudo, foi a sucrose sintase, que dentre outras funções conhecidas na nodulação, contribui significativamente para o desenvolvimento da parede celular (KOCH *et al.*, 2004). O presente experimento confirma a importância de tais enzimas na nodulação, indicando que, aos 10 DAI, as enzimas, com maior nível de expressão foram as endo-1,3-beta-D-glucosidases e as pectinesterases.

5.2 PREDIÇÃO *IN SILICO* DE MIRNAS ENVOLVIDOS NA NODULAÇÃO

Embora existam trabalhos em nível de genômica estrutural e funcional sobre os mecanismos moleculares envolvidos no processo de nodulação, tanto na planta quanto na bactéria, alguns mecanismos de regulação da expressão dos genes envolvidos na simbiose ainda são pouco conhecidos e estudados. Desse modo, além de analisar as vias metabólicas mais ativas na planta em resposta à nodulação, objetivou-se também a busca *in silico* de novos possíveis microRNAs (miRNAs), moléculas não codificantes que regulam a expressão pós-transcricional dos genes. Esses miRNAs participam de numerosos processos de desenvolvimento da planta, inclusive, do estabelecimento simbiótico e da formação dos nódulos (SIMON *et al.*, 2009).

As análises de bioinformática visando a predição dos miRNAs na biblioteca subtrativa tiveram como ponto de partida as 3.210 sequências que passaram por um *pipeline* de anotação de RNAs não codificantes (*non-coding* RNA ou ncRNA), a fim de identificar aquelas que apresentavam evidências de precursores de miRNA. O resultado dessa anotação geral de RNAs é mostrado na Tabela 6.

Tabela 6 – Relatório do *pipeline* de anotação geral de RNAs não codificantes.

Evidência	Nº sequências	%
Sem Evidência	2.024	63,1
<i>small nucleolar</i> RNA	721	22,4
microRNA	161	5,02
<i>non-coding</i> RNA	155	4,8
<i>small</i> RNA	52	1,62
CRISP	39	1,22
<i>small nuclear</i> RNA	28	0,9
Cis-regulatório	16	0,5
Outros	14	0,44

A maioria das sequências não apresentou evidências de representar ncRNA. Por outro lado, a presença de possíveis *small nucleolar* RNAs, que estão envolvidos basicamente na modificação de RNAs ribossomais (MATTICK; MAKUNIN, 2006) foi bastante representativa. Apesar de 155 sequências terem apresentado evidência de *non-coding* RNA, não foi possível identificar a quais classes as mesmas pertencem. Além disso, a anotação evidenciou a presença de outras moléculas reguladoras (ex. CRISP). O estudo foi, então, concentrado nas 161 sequências hipotéticas de precursores de microRNAs e, com base nas evidências do *software* INFERNAL (EDDY, 2002; NAWROCKI *et al.*, 2009), 19 transcritos foram selecionados (Tabela 7), baseando-se nas melhores pontuações obtidas.

A Tabela 7 apresenta a identificação das 19 sequências originais da biblioteca subtrativa, conforme depositadas no banco de dados do GenoSoja (<http://bioinfo03.ibi.unicamp.br/soja>). O tamanho das sequências onde os miRNAs preditos estão contidos variaram de 523 até 2.064 nucleotídeos e apresentaram pontuações entre 19,42 e 34,98 (Tabela 7). Essa pontuação, segundo o programa INFERNAL, é denominada *bit score*, e baseia-se na diferença entre uma sequência de miRNA verdadeira e uma sequência sem evidência (NAWROCKI *et al.*, 2009). Além disso, o trabalho de Paschoal *et al.* (2009) mostrou que uma pontuação a partir de 19 já é considerada como de confiança.

Tabela 7 – Sequências precursoras de miRNAs evidenciadas através do *pipeline* de anotação de ncRNA.

ID sequência	^a Tam.	Candidato miRNA	Bit score	^b Fita	Início	Fim	^c Tam.
Contig5263	1624	gma-MIR-Cand01	34,98	-	312	217	96
Contig1834	523	gma-MIR-Cand02	28,99	-	414	253	162
Contig24974	601	gma-MIR-Cand03	27,53	+	464	544	81
SJ10-E1-R05-043-H05-UC.F	569	gma-MIR-Cand04	27,39	+	235	378	144
Contig16659	707	gma-MIR-Cand05	24,85	-	630	488	143
Contig13869	540	gma-MIR-Cand06	23,04	-	372	279	94
Contig26154	1100	gma-MIR-Cand07	22,02	+	9	86	78
Contig8113	1092	gma-MIR-Cand08	21,79	-	211	51	161
Contig24959	925	gma-MIR-Cand09	21,51	-	918	823	96
Contig6190	668	gma-MIR-Cand10	20,98	+	551	626	76
Contig29200	1466	gma-MIR-Cand11	20,95	-	95	14	82
SJ16-E1-L08-029-E05-UC.F	590	gma-MIR-Cand12	20,64	+	8	129	122
Contig7552	1058	gma-MIR-Cand13	20,58	+	130	257	128
Contig27883	2064	gma-MIR-Cand14	20,57	-	1059	1007	53
Contig24513	1248	gma-MIR-Cand15	20,37	+	45	158	114
Contig21304	1224	gma-MIR-Cand16	20,23	+	242	335	94
Glyma08g08970.1	859	gma-MIR-Cand17	20,20	-	92	6	87
Contig19125	964	gma-MIR-Cand18	19,86	-	869	800	70
Contig8009	1370	gma-MIR-Cand19	19,42	-	1355	1229	127

^aTamanho da sequência original da biblioteca/ ^bSentido da fita de DNA onde encontra-se o pré-miRNA/ ^cTamanho da sequência do precursor predito.

Dentre as sequências candidatas, onze foram identificadas na fita negativa (antisense) e oito na fita positiva (sense). Essa predição indicou, ainda, o início e o fim da sequência do pré-miRNA predito com o seu respectivo tamanho, uma informação necessária para o corte dessa região na sequência original do transcrito. Com esse resultado foram realizadas as análises posteriores, agora com a sequência e a sentido da fita preditos, sendo que para as fitas de sentido negativo foi utilizado seu reverso complementar. Todos os candidatos passaram a ser identificados, aqui, como *gma-MIR-Cand(xx)*.

Partindo para uma anotação manual, foi realizada uma pesquisa no miRBase contra as sequências precursoras de miRNAs já depositadas. O resultado do alinhamento mostrou seis candidatos apresentando similaridade com pré-miRNAs das famílias de miRNAs de *Glycine max*, indicando que, possivelmente, representam sequências parálogas (Tabela 8). Os candidatos *gma-MIR-Cand04*, *gma-MIR-Cand05* e *gma-MIR-Cand12* alinharam com o precursor de MIR1530; enquanto que o candidato *gma-MIR-Cand10* alinhou com MIR1520 e o *gma-MIR-Cand15* alinhou com a sequência da família MIR1522. Essas famílias estão envolvidas nos estágios iniciais da nodulação, conforme identificado por Subramanian *et al.* (2008) em raízes de soja após 3 h de inoculação com *B. japonicum*. Outra família de miRNA de soja que apresentou similaridade com um candidato precursor foi a MIR4393 (*gma-MIR-Cand13*), identificada por Joshi *et al.* (2010), em um trabalho envolvendo quatro tecidos diferentes (semente, flor, raiz e nódulo).

Tabela 8 – Resultado da análise das sequências candidatas no miRBase.

Candidato miRNA	Organismo/Família miRNA	E-valor	Evidência
gma-MIR-Cand01	<i>Arabidopsis thaliana</i> (MIR414)	0.001	Possível Ortólogo
gma-MIR-Cand02	<i>Oryza sativa</i> (MIR812)	3E-05	Possível Ortólogo
gma-MIR-Cand03	<i>Arabidopsis thaliana</i> (MIR825)	0.33	Possível Ortólogo
gma-MIR-Cand04	<i>Glycine max</i> (MIR1530)	0.032	Possível Parálogo
gma-MIR-Cand05	<i>Glycine max</i> (MIR1530)	6E-05	Possível Parálogo
gma-MIR-Cand06	<i>Oryza sativa</i> (MIR5543)	0.25	Possível Ortólogo
gma-MIR-Cand07	<i>Oryza sativa</i> (MIR1874)	0.22	Possível Ortólogo
gma-MIR-Cand08	<i>Medicago truncatula</i> (MIR2591)	7E-04	Possível Ortólogo
gma-MIR-Cand09	<i>Oryza sativa</i> (MIR2921)	0.15	Possível Ortólogo
gma-MIR-Cand10	<i>Glycine max</i> (MIR 1520)	0.036	Possível Parálogo
gma-MIR-Cand11	<i>Medicago truncatula</i> (MIR2608)	0.002	Possível Ortólogo
gma-MIR-Cand12	<i>Glycine max</i> (MIR1530)	0.012	Possível Parálogo
gma-MIR-Cand13	<i>Glycine max</i> (MIR4393)	0.040	Possível Parálogo
gma-MIR-Cand14	<i>Medicago truncatula</i> (MIR2609)	0.001	Possível Ortólogo
gma-MIR-Cand15	<i>Glycine max</i> (MIR1522)	0.28	Possível Parálogo
gma-MIR-Cand16	<i>Oryza sativa</i> (MIR396)	0.030	Possível Ortólogo
gma-MIR-Cand17	<i>Arabidopsis thaliana</i> (MIR414)	7E-08	Possível Ortólogo
gma-MIR-Cand18	<i>Arabidopsis thaliana</i> (MIR403)	0.64	Possível Ortólogo
gma-MIR-Cand19	<i>Sorghum bicolor</i> (MIR437)	0.010	Possível Ortólogo

Os demais candidatos apresentaram similaridade com precursores de outras espécies de plantas, como as famílias MIR414 (gma-MIR-Cand01; gma-MIR-Cand17), MIR825 (gma-MIR-Cand03) e MIR403 (gma-MIR-Cand18) de *Arabidopsis thaliana*; MIR 812 (gma-MIR-Cand02), MIR5543 (gma-MIR-Cand06), MIR1874 (gma-MIR-Cand07); MIR2921 (gma-MIR-Cand09) e MIR396 (gma-MIR-Cand16) de *Oryza sativa*; MIR2591 (gma-MIR-Cand08), MIR2608 (gma-MIR-Cand11) e MIR2609 (gma-MIR-Cand14) de *Medicago truncatula*, e MIR437 (gma-MIR-Cand19) de *Sorghum bicolor*. Para a maioria dos precursores das famílias de miRNAs em que os candidatos apresentaram similaridade ainda não havia relato de envolvimento no processo de nodulação.

No caso da família MIR396 de *Oryza sativa*, esta se mostra bastante conservada entre as diversas espécies de plantas, sendo encontrada, além da soja, em *Arabidopsis thaliana*, *Sorghum bicolor*, *Saccharum officinarum*, *Zea mays*, *Populus trichocarpa*, *Medicago truncatula*, entre outras. Em soja, foi identificada nos estágios iniciais de nodulação (SUBRAMANIAN *et al.*, 2008) e em resposta a estresses (KULCHESKI *et al.*, 2011). As famílias MIR2591, MIR2608 e MIR2609 de *Medicago truncatula*, que apresentaram similaridade com os candidatos precursores,

também foram identificadas como diferencialmente reguladas durante a nodulação (LELANDAIS-BRIÈRE *et al.*, 2009).

Através desses alinhamentos, foi possível observar que as sequências obtidas neste experimento apresentaram certo grau de similaridade com as sequências de precursores previamente identificados, no entanto, não tiveram similaridade com seus respectivos miRNAs maduros. Essa é uma constatação muito importante, evidenciando que tais pré-miRNAs candidatos possam originar miRNAs, ainda não identificados, decorrentes da nodulação da soja. O alinhamento de cada sequência pode ser verificado com maiores detalhes no Anexo E.

A formação da estrutura secundária *stem-loop* foi obtida utilizando o programa Mfold (Anexo F), um algoritmo bastante utilizado na predição computacional de miRNAs em plantas (ZHANG *et al.*, 2005; QIU *et al.*, 2007; SUBRAMANIAN *et al.*, 2008; WANG *et al.*, 2009; KULCHESKI *et al.*, 2011), baseado principalmente na ΔG mínima obtida para a formação da estrutura secundária (ZUKER, 2003). Essa energia livre dos precursores de miRNAs é, em geral, menor do que em outros ncRNAs (BONNET *et al.*, 2004), sendo, portanto, um parâmetro importante na predição de novos miRNAs, pois representa a quantificação da estabilidade da estrutura secundária (pareamento) formada (MATHEWS *et al.*, 1999; MATHEWS; TURNER, 2006).

O estudo realizado por Thakur e colaboradores (2011) deixou claro que existe grande variação na energia livre das estruturas de pré-miRNAs em plantas. Na análise da frequência de ΔG em sequências de precursores de *S. bicolor*, *O. sativa*, *A. thaliana* e *M. truncatula*, a distribuição ficou entre aproximadamente -20 a -100 kcal/mol, estando a maioria concentrada entre cerca de -30 a -70 kcal/mol. Os autores também mostraram que o tamanho do pré-miRNA pode influenciar no valor de ΔG . Os valores de ΔG dos candidatos deste estudo variaram de -11,10 a -60,60 kcal/mol (Tabela 9), onde as melhores estruturas foram aquelas com $\Delta G \leq -20$ kcal/mol, representando, possivelmente, autênticos precursores. Os tamanhos dos precursores foram de 53 a 162 nucleotídeos (Tabela 7). Sabe-se que em plantas há uma variação maior de comprimento dessa sequência do que em animais, podendo variar de 50 até 350 nucleotídeos (JUNG *et al.*, 2009).

O processamento do pré-miRNA origina a sequência do miRNA maduro que irá atuar na regulação negativa, por degradação ou repressão da

tradução, dos genes alvos expressos nas células das raízes durante a nodulação. A fim de obter essa sequência reguladora, foi realizada uma predição *in silico* através do programa MaturePred, capaz de localizar o miRNA maduro a partir de novos precursores candidatos, baseado principalmente nos pré-miRNAs de plantas verificados experimentalmente e depositados no miRBase (XUAN *et al.*, 2011).

A maioria das sequências dos miRNAs maduros tiveram um tamanho de 21 nucleotídeos, e apenas um candidato apresentou 18 nucleotídeos. Dentre estes, a predição de dez sequências de miRNAs foi identificada no “braço” posição 3’ do precursor e nove no “braço” 5’ (Tabela 9). A localização destas sequências nas suas respectivas estruturas secundárias está apresentada com detalhes no Anexo F.

Quanto a sua localização, utilizando o genoma da soja (Phytozome) como referência, os candidatos preditos foram transcritos de regiões intrônicas, intergênicas, de região 5’ e 3’ UTR e, ainda, houve sequências transcritas de região CDS, como mostra a Tabela 9. Apesar da maioria dos pré-miRNAs de plantas serem transcritos de regiões intergênicas e intrônicas, estes também podem advir de regiões UTR e CDS, conforme observado por Fahlgren *et al.* (2010); Kulcheski *et al.* (2011) e Nozawa *et al.* (2012), o que vem a fortalecer os resultados.

Tabela 9 – Resultado da predição *in silico* do miRNA maduro.

Candidato miRNA	Energia livre da Estrutura	Sequência miRNA maduro	^a Tam.	^b Score	^c Braço	^d Crom.	Coordenada miRNA predito	Região
*gma-MIR-Cand01	$\Delta G = -37,30$ kcal/mol	UGAUCGUGAUCGUGAUCGUGC	21 nt	0.649064	3'	Gm15	7382273 - 7382293	^e CDS
*gma-MIR-Cand02	$\Delta G = -29,10$ kcal/mol	UAGUAUAAAAAGUAUUAAAU	21 nt	0.653008	5'	Gm17	2217418 - 2217438	intrônica
*gma-MIR-Cand03	$\Delta G = -25,50$ kcal/mol	UUCUAUAGAAGUCAAGUGCAC	21 nt	0.582132	3'	Gm20	819015 - 819035	intrônica
*gma-MIR-Cand04	$\Delta G = -32,20$ kcal/mol	UCUAUCUAUCUAUCUAUCUAU	21 nt	0.712194	5'	Gm17	2225498 - 2225518	intrônica
gma-MIR-Cand05	$\Delta G = -22,10$ kcal/mol	UUUAUUUCUAAAUAUUUAUUG	21 nt	0.579147	5'	Gm06	41501044 - 41501064	3' UTR
gma-MIR-Cand06	$\Delta G = -16,90$ kcal/mol	ACAAAGUGGUUAAUUAACAAC	21 nt	0.546925	5'	Gm16	20464796 - 20464816	CDS/intergênica
*gma-MIR-Cand07	$\Delta G = -60,60$ kcal/mol	AUCCAGAAGGAGUCAACUCUC	21 nt	0.698796	3'	Gm10	47394262 - 47394282	CDS
gma-MIR-Cand08	$\Delta G = -20,80$ kcal/mol	AUGAUAAUCUAUAAGGCAA	21 nt	0.44891	3'	Gm04	49212058 - 49212078	5'UTR
gma-MIR-Cand09	$\Delta G = -15,80$ kcal/mol	AUAAGUUUAAGCUUGAACAC	21 nt	0.516254	3'	Gm20	34162445 - 34162465	intergênica
gma-MIR-Cand10	$\Delta G = -11,90$ kcal/mol	AAUUCGUUAUUUAUCUAUC	18 nt	0.528889	3'	Gm04	4131636 - 4131653	3'UTR
*gma-MIR-Cand11	$\Delta G = -23,20$ kcal/mol	UCUCUCUCUCUCUUUGUUUUU	21 nt	0.522843	5'	Gm12	20305882 - 20305902	5'UTR
gma-MIR-Cand12	$\Delta G = -11,40$ kcal/mol	GUUUUCUGUAUAAUUUCAUA	21 nt	0.683337	5'	Gm03	36784285 - 36784305	intrônica
*gma-MIR-Cand13	$\Delta G = -20,50$ kcal/mol	UGUUUUUAUAUGAAAAGGUGA	21 nt	0.520922	5'	Gm04	6683683 - 6683703	intrônica
gma-MIR-Cand14	$\Delta G = -11,70$ kcal/mol	UUUUAGUGACGAAUAUUUCA	21 nt	0.745488	5'	Gm10	9075172 - 9075192	intergênica
gma-MIR-Cand15	$\Delta G = -11,10$ kcal/mol	CUUCUCAUUUAUGAAUUAUGA	21 nt	0.46095	3'	Gm20	37227792 - 37227812	intrônica
*gma-MIR-Cand16	$\Delta G = -56,20$ kcal/mol	AGGUGAAGGACUACGCCGAUU	21 nt	0.741033	3'	Gm03	36536344 - 36536364	CDS
*gma-MIR-Cand17	$\Delta G = -32,80$ kcal/mol	UGAUGGUGAUGAUGAUGAUGA	21 nt	0.622788	5'	Gm08	6375449 - 6375469	CDS
gma-MIR-Cand18	$\Delta G = -17,00$ kcal/mol	ACACCAAUGAGGAACAAUGUA	21 nt	0.569925	3'	Gm09	40325686 - 40325706	CDS
gma-MIR-Cand19	$\Delta G = -13,30$ kcal/mol	AAUUUAAAAUUUAGCCAUUUA	21 nt	0.687659	3'	Gm04	4458909 - 4458929	3' UTR

a. Tamanho do miRNA maduro em nucleotídeos(nt)

b. Score: pontuação obtida pelo preditor de miRNA maduro (MaturePred)

c. Localização do miRNA maduro dentro da estrutura *stem-loop*

d. Cromossomo em que o miRNA predito foi identificado

e. CDS- *coding sequence* (sequência codificadora)

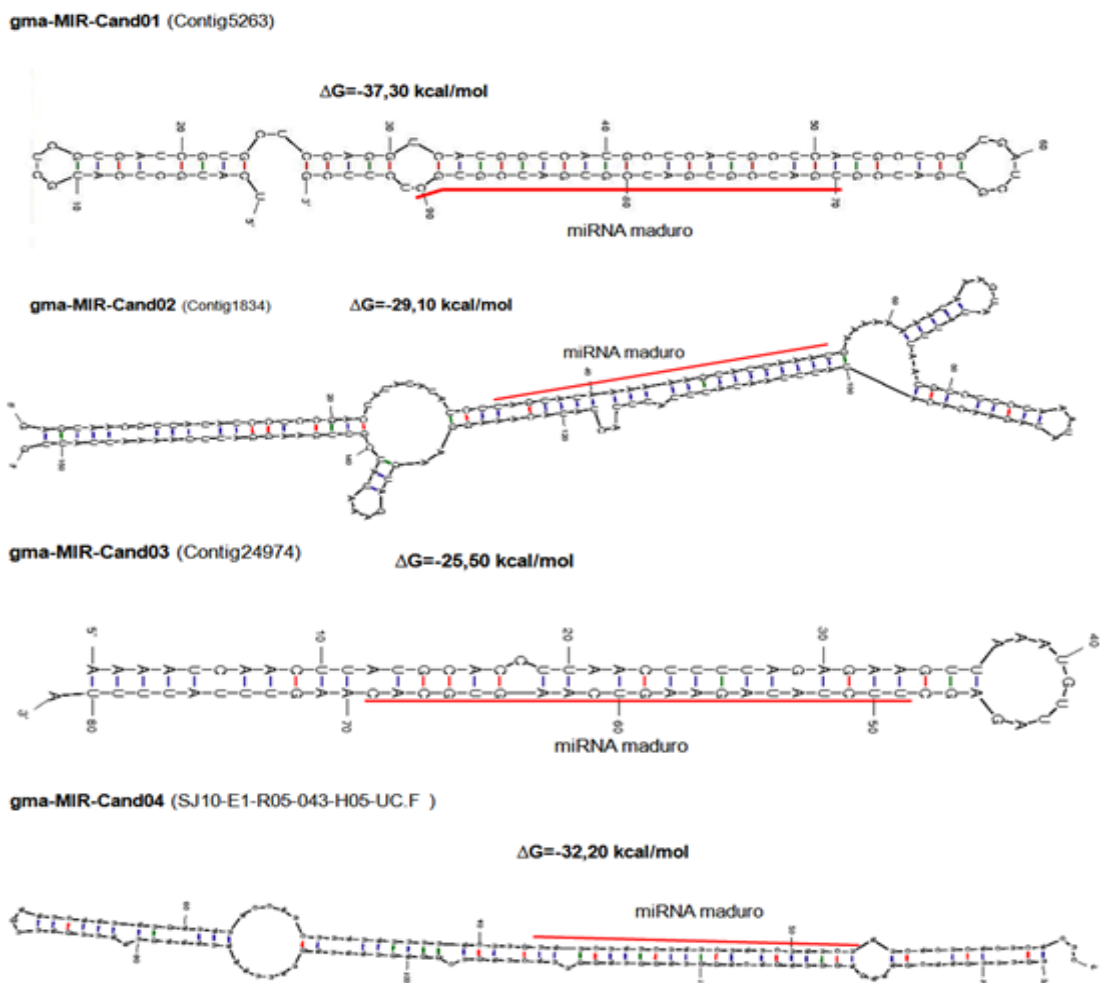
f.UTR- *untranslated region* (região não traduzida)

*. Potenciais novos miRNAs de soja preditos *in silico*.

Dessa forma, através dos critérios citados anteriormente, este estudo aponta nove potenciais miRNAs (Tabela 9), cuja estrutura é apresentada na Figura 7.

Na visualização das estruturas secundárias dos precursores, a sequência de cada miRNA maduro predita está destacada com uma linha, onde cada um apresenta seus respectivos valores de ΔG . Conseqüentemente, com base nas fortes evidências e nos critérios estabelecidos, estes são possivelmente novos miRNAs de soja diferencialmente expressos na nodulação. Note-se, através da Figura 7, que a sequência do miRNA maduro encontra-se dentro do “braço” da estrutura precursora (*stem-loop*), e ainda a presença de *mismatches* no pareamento miRNA:miRNA* foi ≤ 4 . É importante ressaltar que o tamanho real dos precursores pode ser ligeiramente maior ou menor à apresentada aqui.

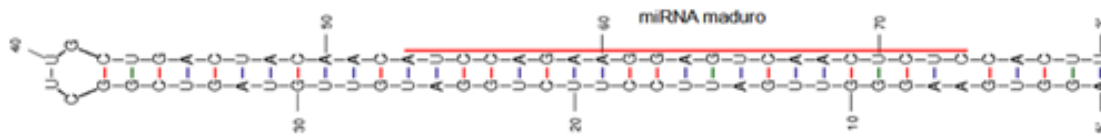
Figura 7 – Estruturas secundárias dos pré-miRNAs e dos miRNAs maduros de *Glycine max* preditos *in silico* (segundo os algoritmos Mfold e MaturePred respectivamente).



Continuação

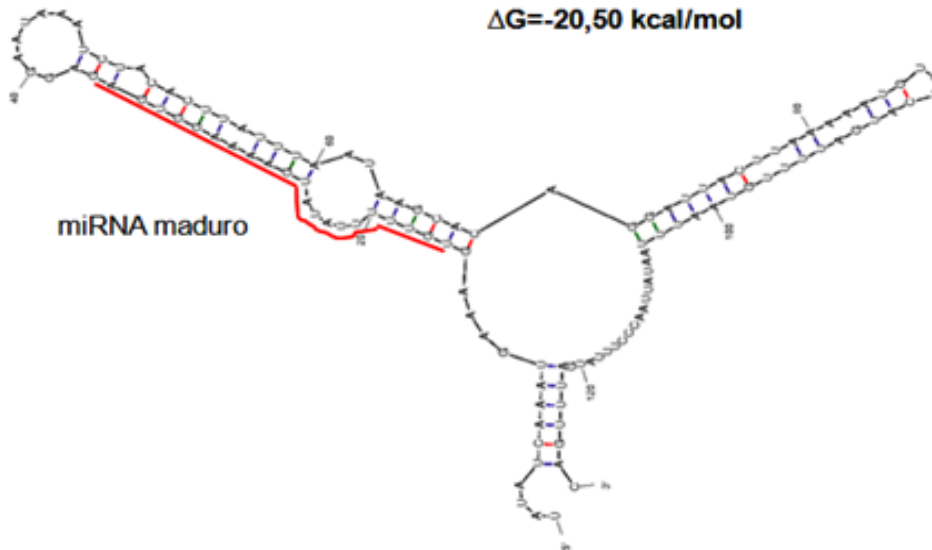
gma-MIR-Cand07 (Contig26154)

$\Delta G = -60,60$ kcal/mol



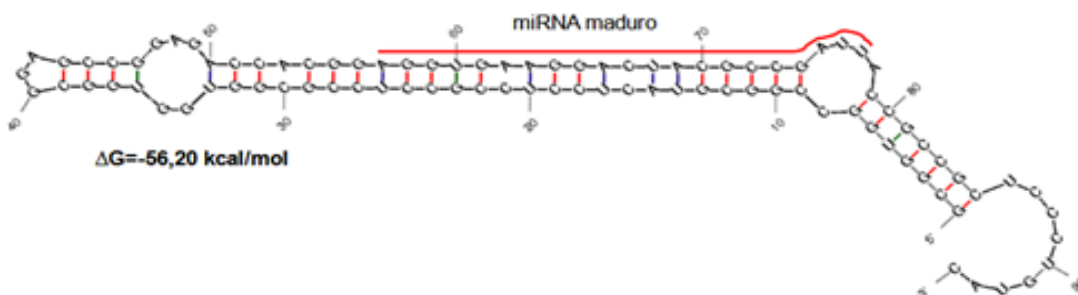
gma-MIR-Cand13 (Contig7552)

$\Delta G = -20,50$ kcal/mol



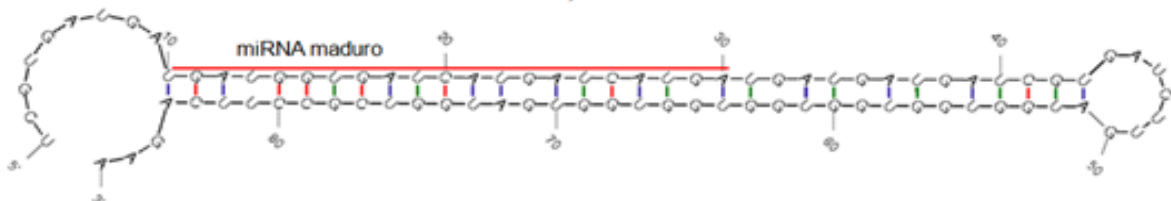
gma-MIR-Cand16 (Contig21304)

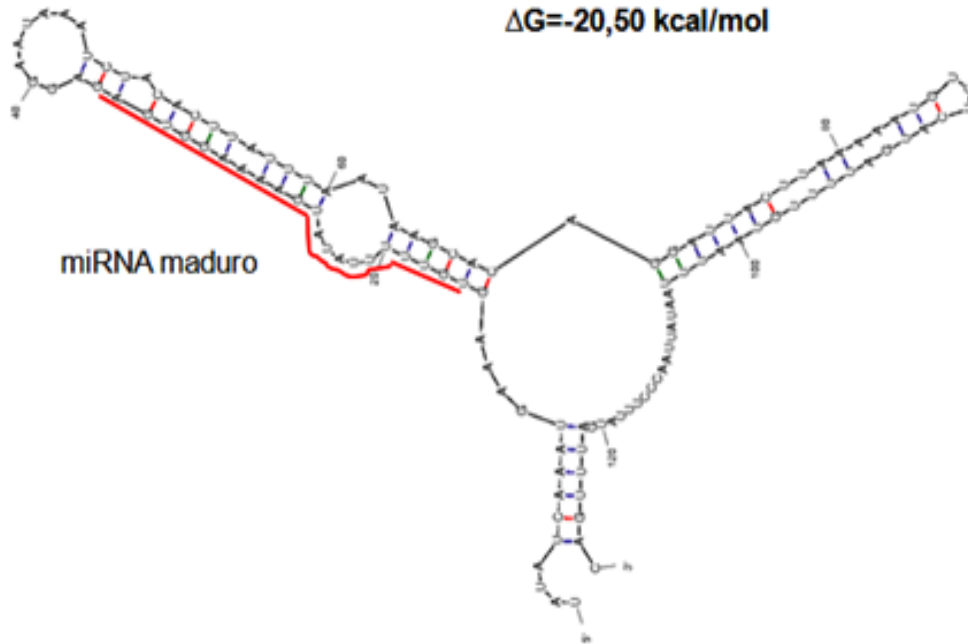
$\Delta G = -56,20$ kcal/mol



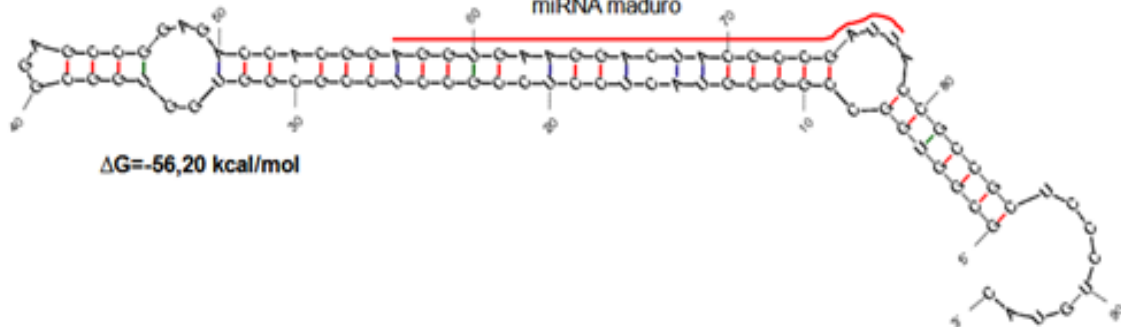
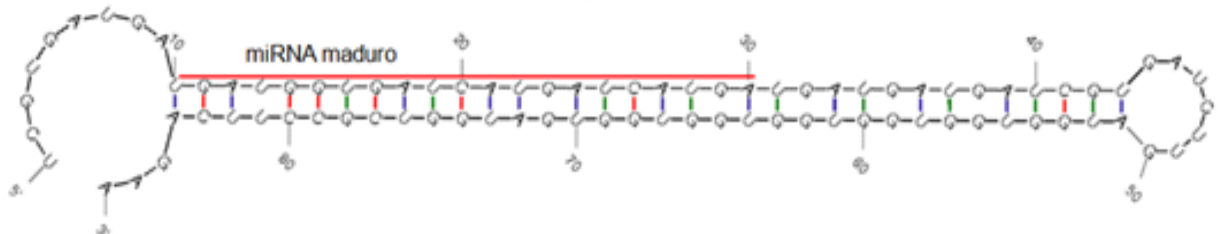
gma-MIR-Cand17 (Glyma08g08970.1)

$\Delta G = -32,80$ kcal/mol



gma-MIR-Cand13 (Contig7552) $\Delta G = -20,50 \text{ kcal/mol}$ **gma-MIR-Cand16 (Contig21304)**

miRNA maduro

 $\Delta G = -56,20 \text{ kcal/mol}$ **gma-MIR-Cand17 (Glyma08g08970.1)** $\Delta G = -32,80 \text{ kcal/mol}$ 

Após a identificação dos miRNAs maduros, para àqueles miRNAs potenciais, foram buscados possíveis alvos de regulação. A predição dos alvos foi realizada pelo programa psRNA Target (<http://plantgrn.noble.org/psRNATarget/>), e as sequências desses alvos foram analisadas individualmente no Phytozome, o que possibilitou identificar os genes, juntamente com sua provável função (Tabela 10).

Tabela 10 – Transcritos alvos de *Glycine max* para os miRNAs preditos.

miRNA	Gene alvo	Descrição	Função
gma-MIR-Cand01	Glyma13g11090	Proteína PLTs	Constituinte da parede celular
	Glyma04g40740	Fator de <i>splicing</i> alternativo	Processamento de mRNA
gma-MIR-Cand02	Glyma04g02900	Predita alfa/beta hidrolase	Catabolismo
gma-MIR-Cand03	Glyma06g05020	Serina carboxipeptidase	Proteólise
gma-MIR-Cand04	Glyma07g35050	Fator de <i>splicing</i>	Processamento de mRNA
	Glyma06g05880	Proteína integral de membrana	Transporte
gma-MIR-Cand07	Glyma20g27950	Ubiquitina	Modificação de proteínas
gma-MIR-Cand11	Glyma08g17120	CTP sintase	Síntese de nucleotídeos
gma-MIR-Cand13	Glyma13g19640	Epimerase dependente de NAD	Metabolismo
gma-MIR-Cand16	Glyma18g23590	Oligonucleotídeo de ligação	Replicação/transcrição ou tradução
gma-MIR-Cand17	Glyma03g40730	Fator de transcrição bZIP	Regulação de transcrição

Os nove miRNAs identificados que provavelmente estão participando da nodulação apresentaram pelo menos um gene alvo, sendo que para gma-MIR-Cand01 e gma-MIR-Cand04 foi possível identificar dois possíveis alvos. Os alvos identificados estão basicamente envolvidos com o metabolismo e com a regulação da transcrição/tradução, processos alterados na planta pela presença do microssimbionte, que desencadeia a formação dos nódulos (LIBAULT *et al.*, 2010; OLDROYD *et al.*, 2011). E, conforme já comentado, os miRNAs participam ativamente na regulação da interação simbiótica, do balanço nutricional da planta, do desenvolvimento nodular e da própria fixação de nitrogênio (SUBRAMANIAN *et al.*, 2008; WANG *et al.*, 2009; SIMON *et al.*, 2009).

Os miRNAs gma-MIR-Cand01 e gma-MIR-Cand04 parecem estar regulando fatores de *splicing* que atuam no processamento de RNA mensageiro (LOPATO *et al.*, 1996). Os gma-MIR-Cand016 e gma-MIR-Cand017 também aparecem regulando genes alvos que participam da expressão gênica sincronizada durante os dez dias pós-inoculação. Aqui, o fator transcricional bZIP parece ser alvo do gma-MIR-Cand017. Este fator de transcrição (FT) regula negativamente a nodulação (NISHIMURA *et al.*, 2002), e está presente na reposta de defesa da planta (DRÖGE-LASER *et al.*, 1997). Outros FTs têm sido apresentados como

sendo regulados por miRNAs durante a nodulação, como o MtHAP2, um regulador importantíssimo durante o desenvolvimento nodular, controlado pelo miR169; o HD-ZIPIII, regulado pelo miR166, e o NAC1 pelo miR164, todos identificados durante a nodulação em *Medicago truncatula* (COMBIER *et al.*, 2006; BOUALEM *et al.*, 2008; D'HAESELEER *et al.*, 2011). Em soja, o trabalho de Subramanian *et al.* (2008) também mostrou que durante a nodulação, os FTs estão sendo alvos dos miRNAs, dado confirmado mais recentemente por Joshi e colaboradores (2010).

Os outros alvos que foram identificados, regulados pelos miRNAs preditos neste estudo, estão basicamente atuando na modulação das alterações metabólicas da planta, o que também foi observado aos 3 DAI com *B. japonicum* em soja, onde os miRNAs estavam regulando uma variedade de alvos, dentre eles, proteases, canais de água, enzimas metabólicas, entre outros (SUBRAMANIAN *et al.*, 2008).

Em meio aos genes alvos apresentados (Tabela 10), apenas o Glyma20g27950 foi encontrado na biblioteca subtrativa, uma forte evidência de que os miRNAs preditos estão regulando negativamente os genes alvos apontados, visto que, estes estão ausentes entre os genes diferencialmente expressos após a inoculação com *B. japonicum* identificados neste estudo.

Estratégias computacionais vêm sendo extremamente eficientes e importantes na identificação de novos miRNAs, como nos trabalhos de Qiu *et al.* (2007), Xie *et al.* (2007) e Lu & Yang (2010) que identificaram novos miRNAs e genes alvos através de análises *in silico* baseado em ESTs de *Gossypium hirsutum*, *Brassica napus* e *Vigna unguiculata*, respectivamente, demonstrando mais uma vez o valor destas análises. Assim, é possível salientar a importância deste estudo na identificação computacional de miRNAs a partir de cDNAs diferencialmente expressos, propiciando a ampliação no conhecimento sobre os miRNAs que regulam a nodulação em soja.

Os resultados acumulam evidências de que a regulação da expressão gênica pelos miRNAs durante a nodulação facilita e coordena os vários processos da planta, desde a interação simbiótica até a organogênese e funcionamento dos nódulos. Como exemplo, podem ser citados os miR482, miR1512 e miR1515, que provocam um aumento significativo no número de nódulos em raiz de soja inoculada com *B. japonicum* (LI *et al.*, 2010).

Portanto, os estudos de predição e identificação de miRNAs e seus alvos ajudam a esclarecer a modulação da fisiologia vegetal que leva à FBN.

CONCLUSÃO

Através da técnica de hibridização subtrativa supressiva foi possível detectar os principais processos biológicos desencadeados na cultivar de soja Conquista, após a inoculação com a estirpe CPAC 15 de *B. japonicum*. Dentre esses processos, destacaram-se as vias metabólicas do metabolismo primário, de modificação da parede celular e a via do sistema de defesa antioxidante. Funções putativas na simbiose de alguns desses genes foram atribuídas pela primeira vez na simbiose soja-*Bradyrhizobium*.

Diante do interesse em compreender melhor os mecanismos de regulação da nodulação, foi também realizada a predição computacional de novos miRNAs. Foram identificados *in silico* nove miRNAs potenciais, que preencheram os requisitos estabelecidos, bem como seus possíveis genes alvos.

A análise de perfis transcricionais da soja é um requisito essencial para compreender o funcionamento das células da raiz da soja durante a nodulação. Nesse contexto, os resultados obtidos neste estudo contribuem para elucidar os mecanismos regulatórios da organogênese nodular. Cabe salientar, ainda, a importância de avaliar a transcrição de genes em uma simbiose envolvendo uma cultivar e uma estirpe de *B. japonicum* amplamente utilizadas no Brasil.

REFERÊNCIAS

- ALVES, B.J.R.; BODDEY, R.M.; URQUIAGA, S. The success of BNF in soybean in Brazil. **Plant and Soil**, v. 252, p. 1–9, 2003.
- ALWINE, J.C.; KEMP, D.J.; STARK, G.R. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. **Proceedings of the National Academy of Sciences of the United States of America**, v. 74, p. 5350-5354, 1977.
- AMBROS, V.; BARTEL, B.; BARTEL, D.P.; BURGE, C.B.; CARRINGTON, J.C.; CHEN, X.; DREYFUSS, G.; EDDY, S.R.; GRIFFITHS-JONES, S.; MARSHALL, M.; MATZKE, M.; RUVKUN, G.; TUSCHL, T. A uniform system for microRNA annotation. **RNA**, v. 9, p. 277-279, 2003.
- ASAMIZU, E.; NAKAMURA, Y.; SATO, S.; TABATA, S. Comparison of the transcript profiles from the root and the nodulating root of the model legume *Lotus japonicus* by serial analysis of gene expression. **Molecular Plant-Microbe Interactions**, v.18, p. 487-498, 2005.
- BARON, C.; ZAMBRYSKI, P.C. The plant response in pathogenesis, symbiosis, and wounding: variations of a common theme? **Annual Review of Genetics**, v. 29, p. 107–129, 1995.
- BARTEL, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. **Cell**, v. 116, p. 281–297, 2004.
- BATEMAN, A.; BIRNEY, E.; CERRUTI, L.; DURBIN, R.; ETWILLER, L.; GRIFFITHS-JONES, S.; HOWE, K.L.; MARSHALL, M.; SONNHAMMER, E.L.L. The Pfam protein families database. **Nucleic Acids Research**, v. 30, p. 276-280, 2002.
- BECANA, M.; MATAMOROS, M.A.; UDVARDI, M.; DALTON, D.A. Recent insights into antioxidant defenses of legume root nodules. **New Phytologist**, v. 188, p. 960–976, 2010.
- BENEDITO, V.A.; DAI, X.; HE, J.; ZHAO, P.X.; UDVARDI, M.K.. Functional genomics of plant transporters in legume nodules. **Functional Plant Biology**, v. 33, p. 731–736, 2006.
- BENEDITO, V.A.; LI, H.; DAI, X.; WANDREY, M.; HE, J.; KAUNDAL, R.; TORRES-JEREZ, I.; GOMEZ, S.K.; HARRISON, M.J.; TANG, Y.; ZHAO, P.X.; UDVARDI, M.K. Genomic inventory and transcriptional analysis of *Medicago truncatula* transporters. **Plant Physiology**, v. 152, p. 1716–1730, 2010.
- BLANCAFLOR, E.B.; WANG, Y.; MOTES, C.M. Organization and function of the actin cytoskeleton in developing root cells. **International Review of Cytology**, v. 252, p. 219-264, 2006.
- BONNET, E.; WUYTS, J.; ROUZÉ, P.; VAN DE PEER, Y. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. **Bioinformatics**, v. 20, p. 2911-2917, 2004.

- BOUALEM, A.; LAPORTE, P.; JOVANOVIC, M.; LAFFONT, C.; PLET, J.; COMBIER, J.P.; NIEBEL, A.; CRESPI, M.; FRUGIER, F. MicroRNA166 controls root and nodule development in *Medicago truncatula*. **The Plant Journal**, v. 54, 876–887, 2008.
- BRECHENMACHER, L.; KIM, M.Y.; BENITEZ, M.; LI, M.; JOSHI, T.; CALLA, B.; LEE, M.P.; LIBAULT, M.; VODKIN, L.O.; XU, D.; LEE, S.H.; CLOUGH, S.J.; STACEY, G. Transcription profiling of soybean nodulation by *Bradyrhizobium japonicum*. **Molecular Plant-Microbe Interactions**, v. 21, p. 631-645, 2008.
- BROUGHTON, W. J.; DILWORTH, M. J. Methods in legume-*rhizobium* technology: plant nutrient solutions. In: SOMASEGARAN, P.; HOBEN, H.J. (Ed.). **Handbook for rhizobia**. Hawaii: NifTAL Project and University of Hawaii, 1970, p. 245–249.
- CAPOEN, W.; GOORMACHTIG, S.; DE RYCKE, R.; SCHROEYERS, K.; HOLSTERS, M. SrSymRK, a plant receptor essential for symbiosome formation. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, p. 10369–10374, 2005.
- CÁRDENAS, L.; MARTÍNEZ, A.; SÁNCHEZ, F.; QUINTO, C. Fast, transient and specific intracellular ROS changes in living root hair cells responding to Nod factors (NFs). **The Plant Journal**, v. 56, p. 802-813, 2008.
- CÁRDENAS, L.; VIDALI, L.; DOMINGUEZ, J.; PÉREZ, H.; SÁNCHEZ, F.; HEPLER, P.K.; QUINTO, C. Rearrangement of actin microfilaments in plant root hairs responding to *Rhizobium etli* nodulation signals. **Plant Physiology**, v. 116, p. 871–877, 1998.
- CARRINGTON, J.C.; AMBROS, V. Role of microRNAs in plant and animal development. **Science**, v. 301, p. 336–338, 2003.
- CHANG, C.; DAMIANI, I.; PUPPO, A.; FRENDI, P. Redox changes during the legume-*Rhizobium* symbiosis. **Molecular Plant**, v. 2, p. 370–377, 2009.
- CHEN, R.; HU, Z.; ZHANG, H. Identification of microRNAs in wild soybean (*Glycine soja*). **Journal of Integrative Plant Biology**, v. 51, p. 1071–1079, 2009.
- COMBIER, J.P.; FRUGIER, F.; DE BILLY, F.; BOUALEM, A.; EL-YAHYAOU, F.; MOREAU, S.; VERNIÉ, T.; OTT, T.; GAMAS, P.; CRESPI, M.; NIEBEL, A. MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. **Genes & Development**, v. 20, p. 3084-3088, 2006.
- COLEBATCH, G.; DESBROSSES, G.; OTT, T.; KRUSELL, L.; MONTANARI, O.; KLOSKA, S.; KOPKA, J.; UDVARDI, M.K. Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. **The Plant Journal**, v. 39, p. 487–512, 2004.
- COLEBATCH, G.; KLOSKA, S.; TREVASKIS, B.; FREUND, S.; ALTMANN, T.; UDVARDI, M. K. Novel aspects of symbiotic nitrogen fixation uncovered by transcript profiling with cDNA arrays. **Molecular Plant-Microbe Interactions**, v. 15, p. 411-420, 2002.

- CONESA, A.; GÖTZ, S. Blast2GO: a comprehensive suite for functional analysis in plant genomics. **International Journal of Plant Genomics**, 2008:619832, 2008.
- COOPER, J.E. Early interactions between legumes and rhizobia: disclosing complexity in a molecular dialogue. **Journal of Applied Microbiology**, v. 103, p. 1355–1365, 2007.
- CHRISTENSEN, A.; SVENSSON, K.; PERSSON, S.; JUNG, J.; MICHALAK, M.; WIDELL, S.; SOMMARIN, M. Functional characterization of *Arabidopsis* calreticulin1a: a key alleviator of endoplasmic reticulum stress. **Plant Cell Physiology**, v. 49, p. 912–924, 2008.
- CROUZET, J.; TROMBIK, T.; FRAYSSE, A.S.; BOUTRY, M. Organization and function of the plant pleiotropic drug resistance ABC transporter family. **FEBS Letters**, v. 580, p. 1123–1130, 2006.
- CURIONI, P.M.G.; HARTWIG, U.A.; NÖSBERGER, J.; SCHULLER, K.A. Glycolytic flux is adjusted to nitrogenase activity in nodules of detopped and argon-treated alfalfa plants. **Plant Physiology**, v. 119, p. 445–453, 1999.
- DAI, X.; ZHAO, P.X. psRNATarget: a plant small RNA target analysis server. **Nucleic Acids Research**, v. 39, (suppl 2): W155-W159, 2011.
- DIATCHENKO, L.; LAU, Y.F.; CAMPBELL, A.P.; CHENCHIK, A.; MOQADAM, F.; HUANG, B.; LUKYANOV, S.; LUKYANOV, K.; GURSKAYA, N.; SVERDLOV, E.D.; SIEBERT, P.D. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. **Proceedings of the National Academy of Sciences of the United States of America**, v. 93, p. 6025-6030, 1996.
- DING, Y.; OLDROYD, G.E. Positioning the nodule, the hormone dictum. **Plant Signaling & Behavior**, v. 4, p. 89-93, 2009.
- DESBROSSES, G.J.; STOUGAARD, J. Root nodulation: a paradigm for how plant-microbe symbiosis influences host developmental pathways. **Cell Host & Microbe**, v. 10, p. 348-358, 2011.
- DOLATABADIAN, A.; SANAVY, S.A.M.M.; GHANATI, F.; GRESSHOFF, P.M. Morphological and physiological response of soybean treated with the microsymbiont *Bradyrhizobium japonicum* pre-incubated with genistein. **South African Journal of Botany**, v. 79, p. 9–18, 2012.
- DRÖGE-LASER, W.; KAISER, A.; LINDSAY, W.P.; HALKIER, B.A.; LOAKE, G.J.; DOERNER, P.; DIXON, R.A.; LAMB, C. Rapid stimulation of a soybean protein-serine kinase that phosphorylates a novel bZIP DNA-binding protein, G/HBF-1, during the induction of early transcription-dependent defenses. **The EMBO Journal**, v. 16, p. 726–738, 1997.
- D'HAESELEER, K.; HERDER, G.D.; LAFFONT, C.; PLET, J.; MORTIER, V.; LELANDAIS-BRIÈRE, C.; DE BODT, S.; DE KEYSER, A.; CRESPI, M.; HOLSTERS, M.; FRUGIER, F.; GOORMACHTIG, S. Transcriptional and post transcriptional

- regulation of a NAC1 transcription factor in *Medicago truncatula* roots. **New Phytologist**, v. 191, p. 647–661, 2011.
- EDDY, S.R. A memory-efficient dynamic programming algorithm for optimal alignment of a sequence to an RNA secondary structure. **BMC Bioinformatics**, v. 3: 8, 2002.
- EL YAHYAOU, F.; KÜSTER, H.; BEN AMOR, B.; HOHNJEC, N.; PÜHLER, A.; BECKER, A.; GOUZY, J.; VERNIÉ, T.; GOUGH, C.; NIEBEL, A.; GODIARD, L.; GAMAS, P. Expression profiling in *Medicago truncatula* identifies more than 750 genes differentially expressed during nodulation, including many potential regulators of the symbiotic program. **Plant Physiology**, v. 136, p. 3159-3176, 2004.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **SOJA em números (safra 2010/2011)**. Londrina: Embrapa Soja, 2012. Disponível em: <http://www.cnpso.embrapa.br/index.php?op_page=294&cod_pai=17>. Acesso em: fev 2012.
- FAHLGREN, N.; JOGDEO, S.; KASSCHAU, K.D.; SULLIVAN, C.M.; CHAPMAN, E.J.; LAUBINGER, S.; SMITH, L.M.; DASENKO, M.; GIVAN, S.A.; WEIGEL, D.; CARRINGTON, J.C. MicroRNA gene evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. **The Plant Cell**, v. 22, p. 1074-1089, 2010.
- FEDOROVA, E.E.; FELIPE, M.R.; PUEYO, J.J.; LUCAS, M.M. Conformation of cytoskeletal elements during the division of infected *Lupinus albus* L. nodule cells. **Journal of Experimental Botany**, v. 58, p. 2225–2236, 2007.
- FEDURCO, M.; ROMIEU, A.; WILLIAMS, S.; LAWRENCE, I.; TURCATTI, G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. **Nucleic Acids Research**, v. 34, p. 1-13, 2006.
- FERGUSON, B.J.; INDRASUMUNAR, A.; HAYASHI, S.; LIN, M.H.; LIN, Y.H.; REID, D.E.; GRESSHOFF, P.M. Molecular analysis of legume nodule development and autoregulation. **Journal of Integrative Plant Biology**, v. 52, p. 61–76, 2010.
- FERGUSON, B.J.; MATHESIUS, U. Signaling interactions during nodule development. **Journal of Plant Growth Regulation**, v. 22, p. 47–72, 2003.
- FOUCHER, F.; KONDOROSI, E. Cell cycle regulation in the course of nodule organogenesis in *Medicago*. **Plant Molecular Biology**, v. 43, p. 773-786, 2000.
- FRENDO, P.; HARRISON, J.; NORMAN, C.; JIMÉNEZ, M.J.H.; VAN DE SYPE, G.; GILABERT, A.; PUPPO, A. Glutathione and homoglutathione play a critical role in the nodulation process of *Medicago truncatula*. **Molecular Plant-Microbe Interactions**, v. 18, p. 254–259, 2005.
- GRAHAM, P.H.; VANCE, C.P. Legumes: importance and constraints to greater use. **Plant Physiology**, v. 131, p. 872-877, 2003.
- GRIFFITHS-JONES, S.; BATEMAN, A.; MARSHALL, M.; KHANNA, A.; EDDY, S.R. Rfam: an RNA family database. **Nucleic Acids Research**, v. 31, p. 439-441, 2003.

- GRIFFITHS-JONES, S.; GROCOCK, R.J.; VAN DONGEN, S.; BATEMAN, A.; ENRIGHT, A.J. miRBase: microRNA sequences, targets and gene nomenclature. **Nucleic Acids Research**, v. 34, D140–D144, 2006.
- GUPTA, D.; TUTEJA, N. Chaperones and foldases in endoplasmic reticulum stress signaling in plants. **Plant Signaling & Behavior**, v. 6, p. 232-236, 2011.
- HEID, C.A.; STEVENS, J.; LIVAK, K.J.; WILLIAMS, P.M. Real time quantitative PCR. **Genome Research**, v. 6, p. 986-994, 1996.
- HERNANDEZ, D.; FRANÇOIS, P.; FARINELLI, L.; ØSTERÅS, M.; SCHRENZEL, J. De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. **Genome Research**, v. 18, p. 802-809, 2008.
- HERTEL, J.; HOFACKER, I.L.; STADLER, P.F. SnoReport: computational identification of snoRNAs with unknown targets. **Bioinformatics**, v. 24, p. 158-164, 2008.
- HINDERHOFER, K; ZENTGRAF, U. Identification of a transcription factor specifically expressed at the onset of leaf senescence. **Planta**, v. 213, n. 3, p. 469-473, 2001.
- HUNGRIA, M.; BODDEY, L.H.; SANTOS, M.A.; VARGAS, M.A.T. Nitrogen fixation capacity and nodule occupancy by *Bradyrhizobium japonicum* and *B. elkanii* strains. **Biology and Fertility of Soils**, v. 27, p. 393-399, 1998.
- HUNGRIA, M.; CAMPO, R.J.; MENDES, I.C. **A importância do processo de fixação biológica do nitrogênio para a cultura da soja**: componente essencial para a competitividade do produto brasileiro. Londrina: Embrapa Soja, 2007. 80p. (Embrapa Soja. Documentos, 283).
- HUNGRIA, M.; CAMPO, R.J.; MENDES, I.C.; GRAHAM, P.H. Contribution of biological nitrogen fixation to the N nutrition of grain crops in the tropics: the success of soybean (*Glycine max* L. Merr.) in South America. In: SINGH, R.P.; SHANKAR, N.; JAIWAL, P.K., eds. **Nitrogen nutrition and sustainable plant productivity**. Houston, Texas: Studium Press, LLC, 2006. p. 43-93.
- JAMES, E.K. Nitrogen fixation in endophytic and associative symbiosis. **Field Crops Research**, v. 65, 197-209, 2000.
- JENSEN, E.S; HAUGGAARD-NIELSEN, H. How can increased use of biological N₂ fixation in agriculture benefit the environment? **Plant and Soil**, v. 252, p. 177–186, 2003.
- JIA, X.Y.; HE, L.H.; JING, R.L.; LIA, R.Z. Calreticulin: conserved protein and diverse functions in plants. **Physiologia Plantarum**, v. 136, p. 127–138, 2009.
- JIANG, H.; WONG, W.H. Statistical inferences for isoform expression in RNA-Seq. **Bioinformatics**, v. 25, p. 1026-1032, 2009.
- JOHANSSON, I.; KARLSSON, M.; JOHANSON, U.; LARSSON, C.; KJELLBOM, P. The role of aquaporins in cellular and whole plant water balance. **Biochimica et Biophysica Acta**, v. 1465, p. 324-334, 2000.

- JONES-RHOADES, M.W.; BARTEL, D.P. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. **Molecular Cell**, v. 14, p. 787–799, 2004.
- JOSHI, T.; YAN, Z.; LIBAULT, M.; JEONG, D.H.; PARK, S.; GREEN, P.J.; SHERRIER, D.J.; FARMER, A.; MAY, G.; MEYERS, B.C.; XU, D.; STACEY, G. Prediction of novel miRNAs and associated target genes in *Glycine max*. **BMC Bioinformatics**, v. 11 (Suppl 1): S14, 2010.
- JUNG, J.H.; SEO, P.J.; PARK, C.M. MicroRNA biogenesis and function in higher plants. **Plant Biotechnology Reports**, v. 3, p. 111–126, 2009.
- KADRI, S.; HINMAN, V.; BENOS, P.V. HHMMiR: efficient *de novo* prediction of microRNAs using hierarchical hidden Markov models. **BMC Bioinformatics**, v. 10 (Suppl 1): S35, 2009.
- KAWSURALIKHIT, S.; YOKOYAMA, T.; KOUCHI, H.; ARIMA, Y. Comprehensive analysis of plant gene expression in soybean root nodules at different growth stages. **Soil Science & Plant Nutrition**, v. 61, p. 535-547, 2005.
- KANEHISA, M.; GOTO, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. **Nucleic Acids Research**, v. 28, p. 27-30, 2000.
- KASPRZEWSKA, A. Plant chitinases: regulation and function. **Cellular & Molecular Biology Letters**, v. 8, p. 809 – 824, 2003.
- KJELLBOM, P.; LARSSON, C.; JOHANSSON, I.; KARLSSON, M.; JOHANSON, U. Aquaporins and water homeostasis in plants. **Trends in Plant Science**, v. 4, p. 308-314, 1999.
- KOCH, K. Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. **Current Opinion in Plant Biology**, v. 7, p. 235–246, 2004.
- KONDOROSI, E.; REDONDO-NIETO, M.; KONDOROSI, A. Ubiquitin-mediated proteolysis. To be in the right place at the right moment during nodule development. **Plant Physiology**, v. 137, p. 1197–1204, 2005.
- KOSKI, L.B.; GRAY, M.W.; LANG, B.F.; BURGER, G. AutoFACT: an automatic functional annotation and classification tool. **BMC Bioinformatics**, v. 6: 151, 2005.
- KOUCHI, H.; SHIMOMURA, K.; HATA, S.; HIROTA, A.; WU, G.J.; KUMAGAI, H.; TAJIMA, S.; SUGANUMA, N.; SUZUKI, A.; AOKI, T.; HAYASHI, M.; YOKOYAMA, T.; OHYAMA, T.; ASAMIZU, E.; KUWATA, C.; SHIBATA, D.; TABATA, S. Large-scale analysis of gene expression profiles during early stages of root nodule formation in a model legume, *Lotus japonicus*. **DNA Research**, v. 11, p. 263-274, 2004.
- KOZOMARA, A.; GRIFFITHS-JONES, S. miRBase: integrating microRNA annotation and deep-sequencing data. **Nucleic Acids Research**, v. 39, p. D152-D157, 2011.
- KULCHESKI, F.R.; 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: 307, 2011.
- KUSANO, M.; FUKUSHIMA, A.; REDESTIG, H.; SAITO, K. Metabolomic approaches toward understanding nitrogen metabolism in plants. **Journal of Experimental Botany**, v. 62, p.1439-1453, 2011.
- LASLETT, D.; CANBACK, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. **Nucleic Acids Research**, v. 32, p. 11-16, 2004.
- LEE, H.; HUR, C.G.; OH, C.J.; KIM, H.B.; PAKR, S.Y.; AN, C.S. Analysis of the root nodule-enhanced transcriptome in soybean. **Molecules and Cells**, v. 18, p. 53-62, 2004.
- LELANDAIS-BRIÈRE, 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. **The Plant Cell**, v. 21, p. 2780-2796, 2009.
- LI, H.; DENG, Y.; WU, T.; SUBRAMANIAN, S.; YU, O. Misexpression of miR482, miR1512, and miR1515 increases soybean nodulation. **Plant Physiology**, v. 153, p. 1759–1770, 2010.
- LI, R.; LI, Y.; KRISTIANSEN, K.; WANG, J. SOAP: short oligonucleotide alignment program. **Bioinformatics**, v. 24, p. 713-714, 2008.
- LIBAULT, M.; FARMER, A.; BRECHENMACHER, L.; DRNEVICH, J.; LANGLEY, R.J.; BILGIN, D.D.; RADWAN, O.; NEECE, D.J.; CLOUGH, S.J.; MAY, G.D.; STACEY, G. Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. **Plant Physiology**, v. 152, p. 541-552, 2010.
- LIEVENS, S.; GOORMACHTIG, S.; HERMAN, S.; HOLSTERS, M. Patterns of pectin methylesterase transcripts in developing stem nodules of *Sesbania rostrata*. **Molecular Plant-Microbe Interactions**, v. 16, p. 164-168, 2002.
- LIMPENS, E.; BISSELING, T. Signaling in symbiosis. **Current Opinion in Plant Biology**, v. 6, p. 343–350 2003.
- LOHAR, D.P.; SHAROPOVA, N.; ENDRE, G.; PEÑUELA, S.; SAMAC, D.; TOWN, C.; SILVERSTEIN, K.A.; VANDENBOSCH, K.A. Transcript analysis of early nodulation events in *Medicago truncatula*. **Plant Physiology**, v. 140, p. 221-234, 2006.
- LOPATO, S.; MAYEDA, A.; KRAINER, A.R.; BARTA, A. Pre-mRNA splicing in plants: characterization of Ser/Arg splicing factors. **Proceedings of the National Academy of Sciences of the United States of America**, v. 93, p. 3074-3079, 1996.
- LOWE, T.M.; EDDY, S.R. A computational screen for methylation guide snoRNAs in yeast. **Science**, v. 283, p. 1168-1171, 1999.

- LOWE, T.M.; EDDY, S.R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. **Nucleic Acids Research**, v. 25, p. 955-964, 1997.
- LU, Y.; YANG, X. Computational identification of novel microRNAs and their targets in *Vigna unguiculata*. **Comparative and Functional Genomics**, p. 1-17, 2010.
- MARDIS, E.R. Next-generation DNA sequencing methods. **The Annual Review of Genomics and Human Genetics**, v. 9, p. 387-402, 2008.
- MARINO, D.; PUCCIARIELLO, C.; PUPPO, A.; FRENDI, P. The redox state, a referee of the legume–rhizobia symbiotic game. **Advances in Botanical Research**, v. 52, p. 116-139, 2009.
- MATHEWS, D.H.; SABINA, J.; ZUKER, M.; TURNER, D.H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. **Journal of Molecular Biology**, v. 288, p. 911-940, 1999.
- MATHEWS, D.H.; TURNER, D.H. Prediction of RNA secondary structure by free energy minimization. **Current Opinion in Structural Biology**, v. 16, p. 270-278, 2006.
- MATTICK, J.S.; MAKUNIN, I.V. Non-coding RNA. **Human Molecular Genetics**, v. 15, p. 17–29, 2006.
- McKERNAN, K.; BLANCHARD, A.; KOTLER, L.; COSTA, G. **Reagents, methods, and libraries for bead-based sequencing**. Inf. C1.12Q 435/6. U.S. 2008/0003571. 1 feb 2006, 3 jan 2008.
- MILLAR, A.A.; WATERHOUSE, P.M. Plant and animal microRNAs: similarities and differences. **Functional & Integrative Genomics**, v. 5, p. 129–135, 2005.
- MINIC, Z. Physiological roles of plant glycoside hydrolases. **Planta**, v. 227, p. 723-740, 2008.
- MITHÖFER, A. Suppression of plant defence in rhizobia–legume symbiosis. **Trends in Plant Science**, v. 7, p. 440-444, 2002.
- MORTAZAVI, A.; WILLIAMS, B.A.; McCUE, K.; SCHAEFFER, L.; WOLD, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. **Nature Methods**, v. 5, p. 621-628, 2008.
- NAQVI, A.R.; SARWAT, M.; HASAN, S.; CHOUDHURY, N.R. **Biogenesis, functions and fate of plant microRNAs**. *Journal of Cellular Physiology*, 2012. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/jcp.24052/pdf>>. Acesso em: jan 2012.
- NAWROCKI, E.P.; KOLBE, D. L.; EDDY, S.R. Infernal 1.0: inference of RNA alignments. **Bioinformatics**, v. 25, p. 1335-1337, 2009.
- NISHIMURA, R.; OHMORI, M.; KAWAGUCHI, M. The novel symbiotic phenotype of enhanced-nodulating mutant of *Lotus japonicus*: *astray* mutant is an early nodulating mutant with wider nodulation zone. **Plant and Cell Physiology**, v. 43, p. 853-859, 2002.

- NOZAWA, M.; MIURA, S.; NEI, M. **Origins and evolution of microRNA genes in plant species**. *Genome Biology and Evolution*, 2012. Disponível em: <<http://gbe.oxfordjournals.org/content/early/2012/01/04/gbe.evs002.short?rss=1>>. Acesso em: jan 2012.
- OEHRLE, N.W.; SARMA, A.D.; WATERS, J.K.; EMERICH, D.W. Proteomic analysis of soybean nodule cytosol. **Phytochemistry**, v. 69, p. 2426–2438, 2008.
- OLDROYD, G.E.; DOWNIE, J.A. Coordinating nodule morphogenesis with rhizobial infection in legumes. **Annual Review of Plant Biology**, v. 59, p. 519-546, 2008.
- OLDROYD, G.E.; DOWNIE, J.A. Calcium, kinases and nodulation signalling in legumes. **Nature Reviews Molecular Cell Biology**, v. 5, p. 566–576, 2004.
- OLDROYD, G.E.; MURRAY, J.D.; POOLE, P.S.; DOWNIE, J.A. The rules of engagement in the legume-rhizobial symbiosis. **Annual Review of Genetics**, v. 45, p. 119-144, 2011.
- PASCHOAL, A.R.; NUNES, F.M.F.; CRISTINO A.S.; SIMÕES, Z.L.P.; DURHAM, A.M. **In silico annotation of non-coding RNA candidates**. In: X-MEETING 2009, 2009, Angra dos Reis.
- PASCHOAL, A.R.; MARACAJA-COUTINHO, V.; SETUBAL, J.C.; SIMÕES, Z.L.P.; VERJOVSKI-ALMEIDA, S.; DURHAM, A.M. **NRDR: non-coding RNA databases resource**. Disponível em: <<http://www.ncrnadatabases.org/>>. Acesso em: jan 2012.
- PEOPLES, M.B.; CRASWELL, E.T. Biological nitrogen fixation: Investments, expectations and actual contributions to agriculture. **Plant and Soil**, v. 141, p. 13-39, 1992.
- PRELL, J.; WHITE, J.P.; BOURDES, A.; BUNNEWELL, S.; BONGAERTS, R.J.; POOLE, P.S. Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, p. 12477-12482, 2009.
- PUCCIARIELLO, C.; INNOCENTI, G.; VAN DE VELDE, W.; LAMBERT, A.; HOPKINS, J.; CLÉMENT, M.; PONCHET, M.; PAULY, N.; GOORMACHTIG, S.; HOLSTERS, M.; PUPPO, A.; FRENO, P. (Homo)glutathione depletion modulates host gene expression during the symbiotic interaction between *Medicago truncatula* and *Sinorhizobium meliloti*. **Plant Physiology**, v. 151, p. 1186–1196, 2009.
- QIU, C.X.; XIE, F.L.; ZHU, Y.Y.; GUO, K.; HUANG, S.Q.; NIE, L.; YANG, Z.M. Computational identification of microRNAs and their targets in *Gossypium hirsutum* expressed sequence tags. **Gene**, v. 395, p. 49–61, 2007.
- RONAGHI, M.; UHLÉN, M.; NYRÉN, P. A sequencing method based on real-time pyrophosphate. **Science**, v. 281, p. 363-365, 1998.
- ROTHBERG, J.M.; HINZ, W.; REARICK, T.M.; SCHULTZ, J.; MILESKI, W.; DAVEY, M.; LEAMON, J.H.; JOHNSON, K.; MILGREW, M.J.; EDWARDS, M.; HOON, J.; SIMONS, J.F.; MARRAN, D.; MYERS, J.W.; DAVIDSON, J.F.; BRANTING, A.; NOBILE, J.R.; PUC, B.P.; LIGHT, D.; CLARK, T.A.; HUBER, M.; BRANCIFORTE,

- J.T.; STONER, I.B.; CAWLEY, S.E.; LYONS, M.; FU, Y.; HOMER, N.; SEDOVA, M.; MIAO, X.; REED, B.; SABINA, J.; FEIERSTEIN, E.; SCHORN, M.; ALANJARY, M.; DIMALANTA, E.; DRESSMAN, D.; KASINSKAS, R.; SOKOLSKY, T.; FIDANZA, J.A.; NAMSARAEV, E.; MCKERNAN, K.J.; WILLIAMS, A.; ROTH, G.T.; BUSTILLO, J. An integrated semiconductor device enabling non-optical genome sequencing. **Nature**, v. 475, p. 348-352, 2011.
- SALAH, I.B.; ALBACETE, A.; ANDÚJAR, C.M.; HAOUALA, R.; LABIDI, N.; ZRIBI, F.; MARTINEZ, V.; PÉREZ-ALFOCEA, F.; ABDELLY, C. Response of nitrogen fixation in relation to nodule carbohydrate metabolism in *Medicago ciliaris* lines subjected to salt stress. **Journal of Plant Physiology**, v. 166, p. 477-488, 2009.
- SANCHEZ-FERNANDEZ, R., FRICKER, M., CORBEN, L. B., WHITE, N. S., SHEARD, N., LEAVER, C. J., VAN MONTAGU, M., INZÉ, D., AND MAY, M. J. Cell proliferation and hair tip growth in the *Arabidopsis* root are under mechanistically different forms of redox control. **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, p. 2745-2750, 1997.
- SANTOS, P.; FORTUNATO, A.; RIBEIRO, A.; PAWLOWSKI, K. **Plant Biotechnology**, v. 25, p. 299–307, 2008.
- SANTOS, R.; HEROUART, D.; SIGAUD, S.; TOUATI, D.; PUPPO, A. Oxidative burst in alfalfa-*Sinorhizobium meliloti* symbiotic interaction. **Molecular Plant -Microbe Interactions**, v. 14, p. 86–89, 2001.
- SCHENA, M.; SHALON, D.; DAVIS, R.W.; BROWN, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. **Science**, v. 270, p. 467–470, 1995.
- SCHMIDT, S.M.; PANSTRUGA, R. Cytoskeleton functions in plant–microbe interactions. **Physiological and Molecular Plant Pathology**, v. 71, p. 135–148, 2007.
- SCHMUTZ, J.; CANNON, S.B.; SCHLUETER, J.; MA, J.; MITROS, T.; NELSON, W.; HYTEN, D.L.; SONG, Q.; THELEN, J.J.; CHENG, J.; XU, D.; HELLSTEN, U.; MAY, G.D.; YU, Y.; SAKURAI, T.; UMEZAWA, T.; BHATTACHARYYA, M.K.; SANDHU, D.; VALLIYODAN, B.; LINDQUIST, E.; PETO, M.; GRANT, D.; SHU, S.; GOODSTEIN, D.; BARRY, K.; FUTRELL-GRIGGS, M.; ABERNATHY, B.; DU, J.; TIAN, Z.; ZHU, L.; GILL, N.; JOSHI, T.; LIBAULT, M.; SETHURAMAN, A.; ZHANG, X.C.; SHINOZAKI, K.; NGUYEN, H.T.; WING, R.A.; CREGAN, P.; SPECHT, J.; GRIMWOOD, J.; ROKHSAR, D.; STACEY, G.; SHOEMAKER, R.C.; JACKSON, S.A. Genome sequence of the palaeopolyploid soybean. **Nature**, v. 463, p. 178–183, 2010.
- SEARLE, I.R.; MEN, A.E.; LANIYA, T.S.; BUZAS, D.M.; ITURBE-ORMAETXE, I.; CARROLL, B.J.; GRESSHOFF, P.M. Long-distance signaling in nodulation directed by a CLAVATA1-like receptor kinase. **Science**, v. 299, p. 109-112, 2003.
- SIMON, S.A.; MEYERS, B.C.; SHERRIER, D.J. MicroRNAs in the rhizobia legume symbiosis. **Plant Physiology**, v. 151, p. 1002–1008, 2009.
- STACEY, G.; LIBAULT, M.; BRECHENMACHER, L.; WAN, J.; MAY, G.D. Genetics and functional genomics of legume nodulation. **Current Opinion in Plant Biology**, v. 9, p. 110-121, 2006.

- STAIGER, C.J. Signaling to the actin cytoskeleton in plants. **Annual Review of Plant Physiology and Plant Molecular Biology**, v. 51, p. 257–288, 2000.
- STARKER, C.G.; PARRA-COLMENARES, A.L.; SMITH, L.; MITRA, R.M.; LONG, S.R. Nitrogen fixation mutants of *Medicago truncatula* fail to support plant and bacterial symbiotic gene expression. **Plant Physiology**, v. 140, p. 671-680, 2006.
- STOUGAARD, J. Genetics and genomics of root symbiosis. **Current Opinion in Plant Biology**, v. 4, p. 328–335, 2001.
- STOUGAARD, J. Regulators and regulation of legumes root nodule development. **Plant Physiology**, v. 124, p. 531-540, 2000.
- STRACKE, S.; KISTNER, C.; YOSHIDA, S.; MULDER, L.; SATO, S.; KANEKO, T.; TABATA, S.; SANDAL, N.; STOUGAARD, J.; SZCZYGLOWSKI, K.; PARNISKE, M. A plant receptor-like kinase required for both bacterial and fungal symbiosis. **Nature**, v. 417, p. 959-962, 2002.
- SUBRAMANIAN, S.; STACEY, G.; YU, O. Distinct, crucial roles of flavonoids during legume nodulation. **Trends in Plant Science**, v. 12, p. 282-285, 2007.
- SUBRAMANIAN, S.; STACEY, G.; YU, O. Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*. **The Plant Journal**, v. 48, p. 261–273, 2006.
- SUBRAMANIAN, S.; FU, Y.; SUNKAR, R.; BARBAZUK, W.B.; ZHU, J.K.; YU, O. Novel and nodulation-regulated microRNAs in soybean roots. **BMC Genomics**, v. 9: 160, 2008.
- SUGIYAMA, A.; SHITAN, N.; YAZAKI, K. Signaling from soybean roots to rhizobium. **Plant Signaling & Behavior**, v. 3, p. 38-40, 2008.
- THAKUR, V.; WANCHANA, S.; XU, M.; BRUSKIEWICH, R.; QUICK, W.P.; MOSIG, A.; ZHU, XIN-GUANG. Characterization of statistical features for plant microRNA prediction. **BMC Genomics**, v. 12:108 2011.
- TURCATTI, G.; ROMIEU, A.; FEDURCO, M.; TAIRI, A.P. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. **Nucleic Acids Research**, v.36, p. 1-13, 2008.
- UEHLEIN, N.; FILESCHI, K.; ECKERT, M.; BIENERT, G.P.; BERTEL, A.; KALDENHOFF, R. Arbuscular mycorrhizal symbiosis and plant aquaporin expression. **Phytochemistry**, v. 68, p.122-129, 2007.
- van RHIJN, P.; VANDERLEYDEN, J. The *Rhizobium*-plant symbiosis. **Microbiological Reviews**, v. 59, 124-142, 1995.
- VANCE, C.P.; GANTT, J.S. Control in nitrogen and carbon metabolism in root-nodules. **Physiologia Plantarum**, v. 85, p. 266-274, 1992.
- VELCULESCU, V.E.; ZHANG, L.; VOGELSTEIN, B.; KINZLER, K.W. Serial analysis of gene expression. **Science**, v. 270, p. 484–487, 1995.

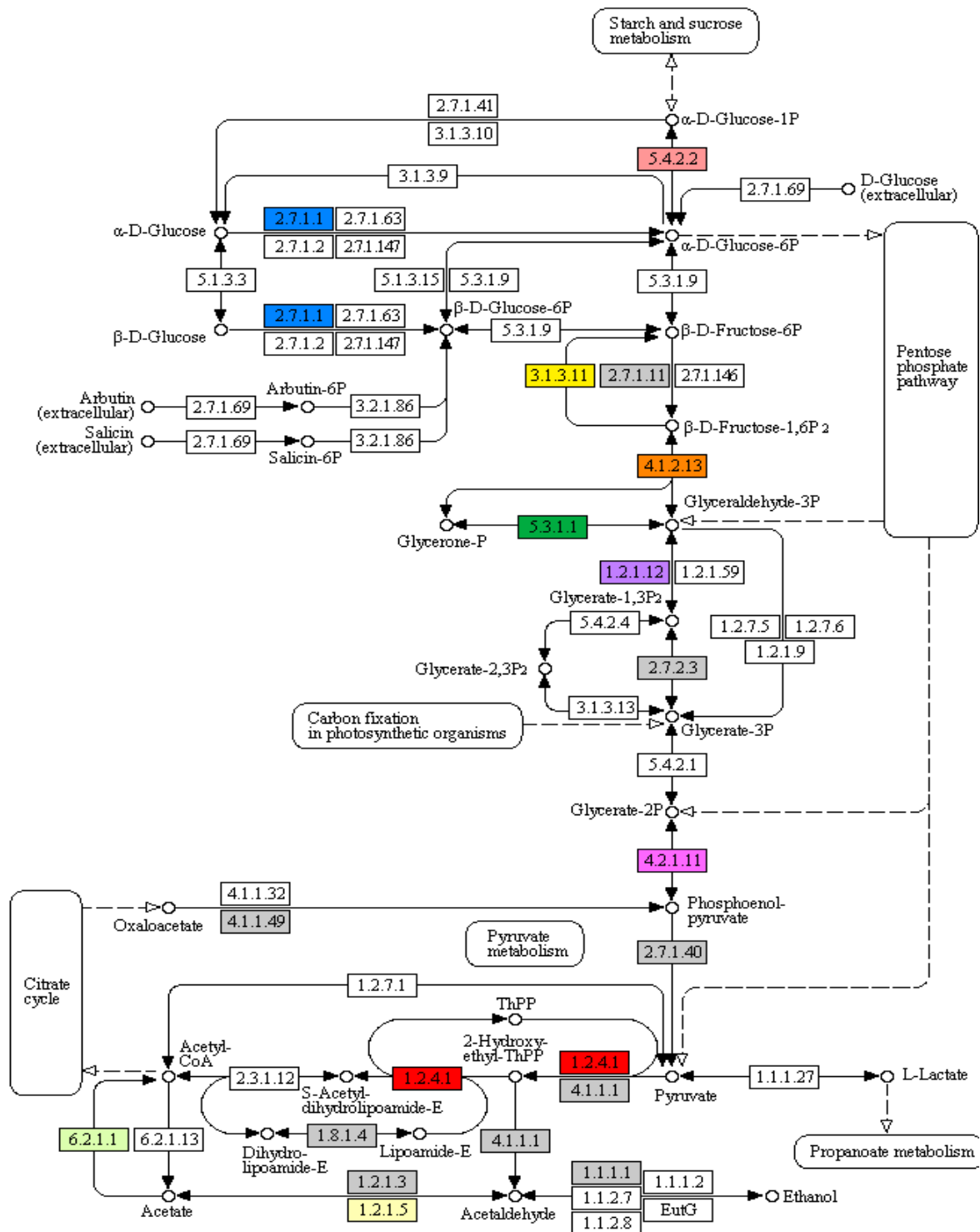
- VINCENT, J.M. **A manual for the practical study of root nodule bacteria**. Science Blackwell Publishing, Oxford , 1970. 164p.
- VOINNET, O. Origin, biogenesis, and activity of plant microRNAs. **Cell**, v. 136, p. 669–687, 2009.
- XIE, F.L.; HUANG, S.Q.; GUO, K.; XIANG, A.L.; ZHU, Y.Y.; NIE, L.; YANG, Z.M. Computational identification of novel microRNAs and targets in *Brassica napus*. **FEBS Letters**, v. 581, p. 1464-1474, 2007.
- XIE, Z.P.; STAEHELIN, C.; WIEMKEN, A.; BROUGHTON, W.J.; MÜLLER, J.; BOLLER, T. Symbiosis-stimulated chitinase isoenzymes of soybean (*Glycine max* (L.) Merr.). **Journal of Experimental Botany**, v. 50, p. 327-333, 1999.
- XUAN, P.; GUO, M.; HUANG, Y.; LI, W.; HUANG, Y. MaturePred: efficient identification of microRNAs within novel plant pre-miRNAs. **PLoS ONE**, v. 6, e27422, 2011.
- WAIS, R.J.; GALERA, C.; OLDROYD, G.; CATOIRA, R.; PENMETSA, R.V.; COOK, D.; GOUGH, C.; DENARIÉ, J.; LONG, S.R. Genetic analysis of calcium spiking responses in nodulation mutants of *Medicago truncatula*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 97, p. 13407-13412, 2000.
- WANG, K.L.; LI, H.; ECKER, J.R. Ethylene biosynthesis and signaling networks. **The Plant Cell**, v. 14, p. S131-S151, 2002.
- WANG, L.; LI, P.; BRUTNELL, T.P. Exploring plant transcriptomes using ultra high-throughput sequencing. **Briefings in Functional Genomics**, v. 9, p. 118-128, 2010.
- WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics**, v. 10, p. 57-63, 2009.
- WANG, Y.; LI, P.; CAO, X.; WANG, X.; ZHANG, A.; LI, X. Identification and expression analysis of miRNAs from nitrogen-fixing soybean nodules. **Biochemical and Biophysical Research Communications**, v. 378, p. 799-803, 2009.
- WEISSMAN, A.M. Themes and variations on ubiquitylation. **Nature Reviews Molecular Cell Biology**, v. 2, p. 169–178, 2001.
- WEIZHONG, L.; JAROSZEWSKI, L.; GODZIK, A. Clustering of highly homologous sequences to reduce the size of large protein databases. **Bioinformatics**, v.17, p. 282-283, 2001.
- ZHANG, B.H.; PAN, X.P.; WANG, Q.L.; COBB, G.P.; ANDERSON, T.A. Identification and characterization of new plant microRNAs using EST analysis. **Cell Research**, v. 15, p. 336-360, 2005.
- ZHU, H.; RIELY, B.K.; BURNS, N.J.; ANÉ, J.M. Tracing nonlegume orthologs of legume genes required for nodulation and arbuscular mycorrhizal symbioses. **Genetics**, v. 172, p. 2491-2499, 2006.

ZUKER, M. Mfold web server for nucleic acid folding and hybridization prediction.
Nucleic Acids Research, v. 31, p. 3406-3415, 2003.

ANEXOS

ANEXO A

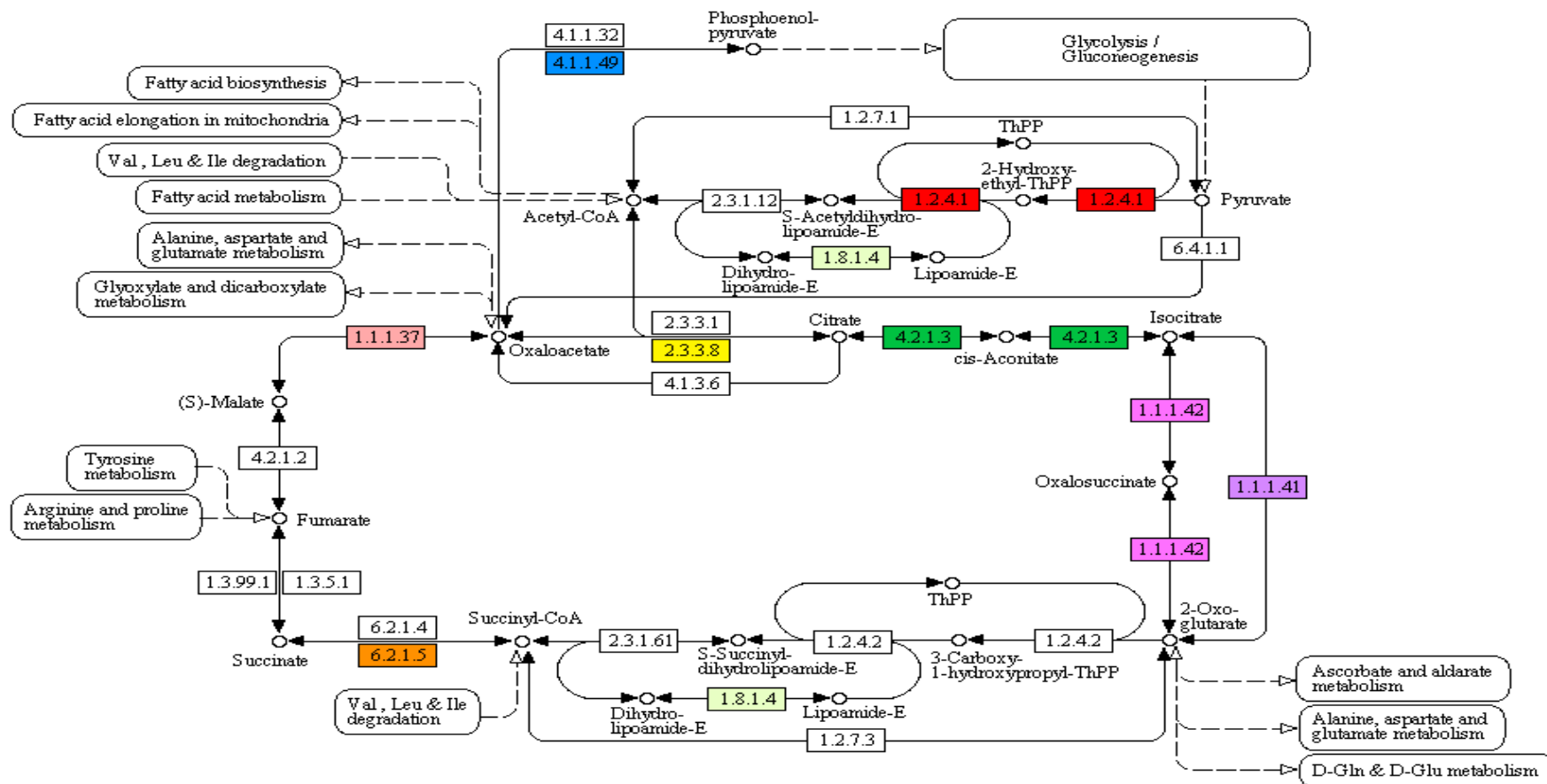
Via glicolítica ativa durante a nodulação (ID Kegg-map00010).



Observação: As enzimas codificadas pelos transcritos presentes na biblioteca subtrativa estão destacadas com cores diferentes, nas quais, cores iguais representam o mesmo produto.

ANEXO B

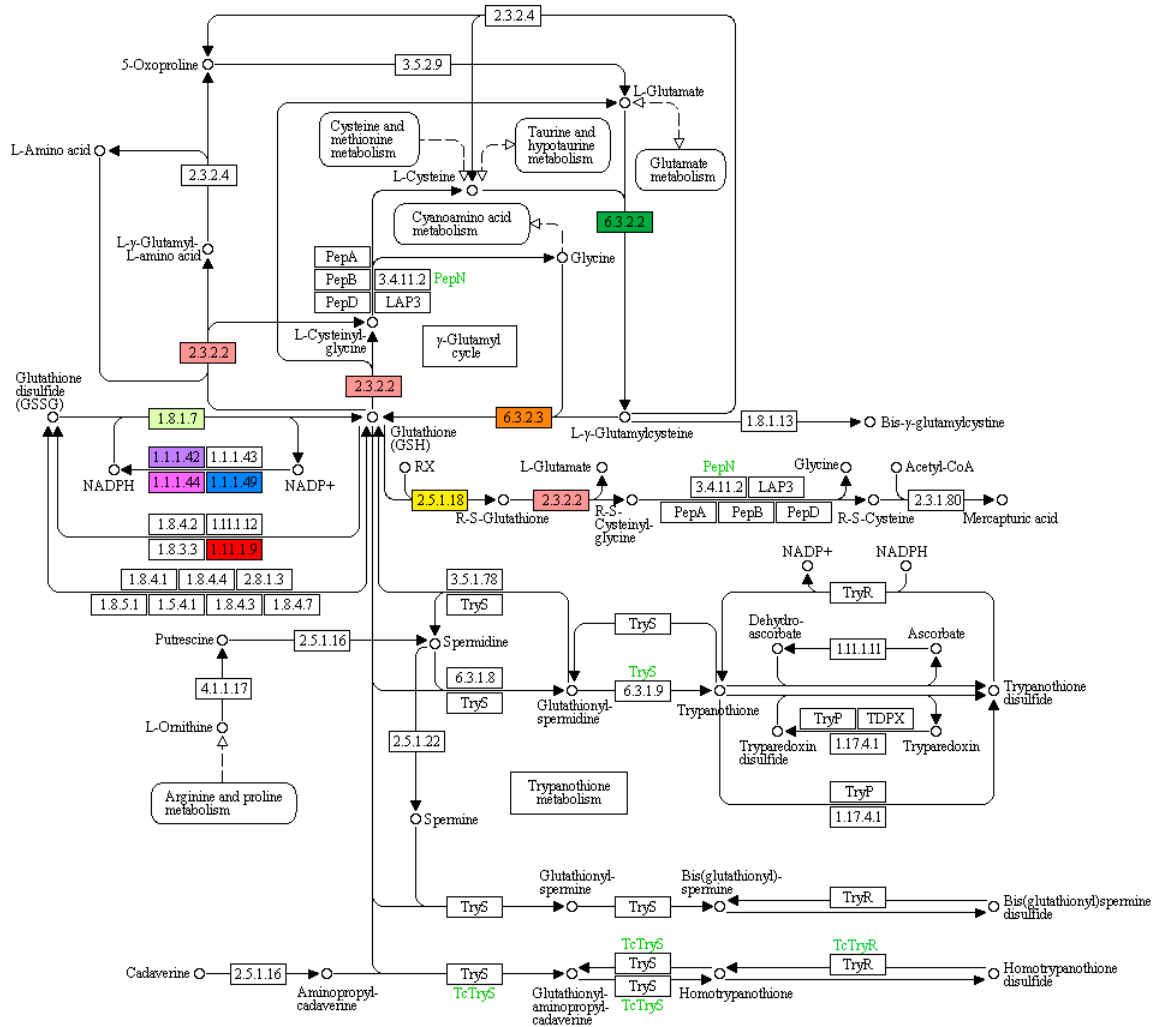
Ciclo de Krebs ativo durante a nodulação (ID Kegg-map00020).



Observação: As enzimas codificadas pelos transcritos presentes na biblioteca subtrativa estão destacadas com cores diferentes, nas quais, cores iguais representam o mesmo produto.

ANEXO C

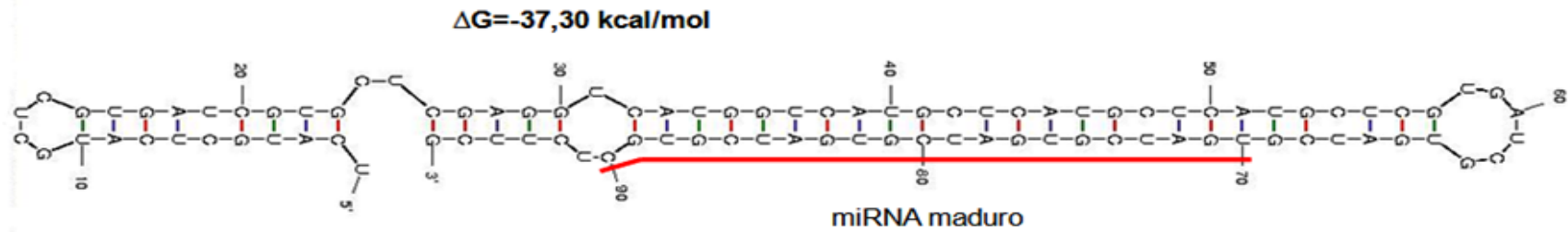
Via metabólica da glutatona – sistema de defesa antioxidante presente na nodulação (ID Kegg- map00480).



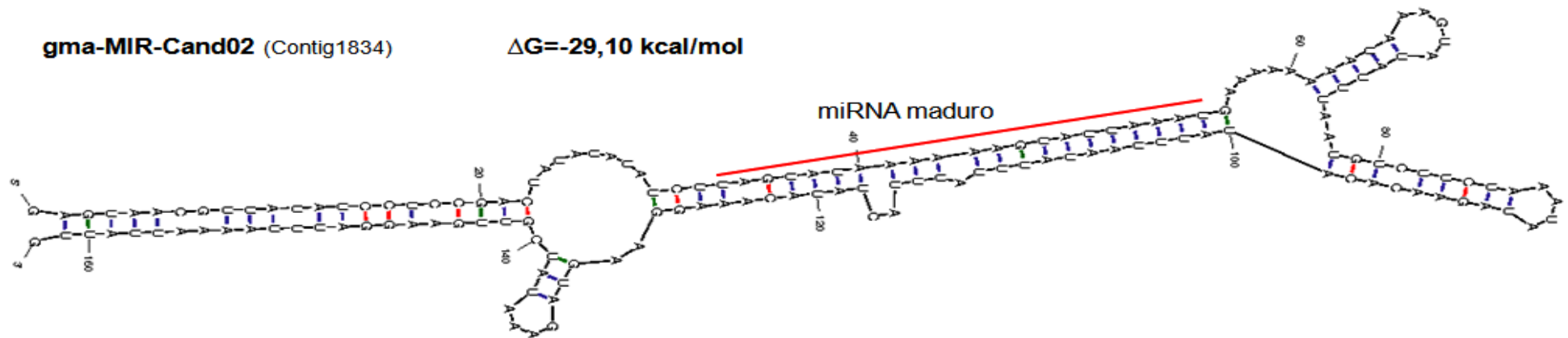
Observação: As enzimas codificadas pelos transcritos presentes na biblioteca subtrativa estão destacadas com cores diferentes, nas quais, cores iguais representam o mesmo produto.

ANEXO F

Resultado da predição das estruturas secundárias dos candidatos precursores de miRNAs e a localização dos miRNAs maduros.
(Segundo os programas Mfold e MaturePred respectivamente).

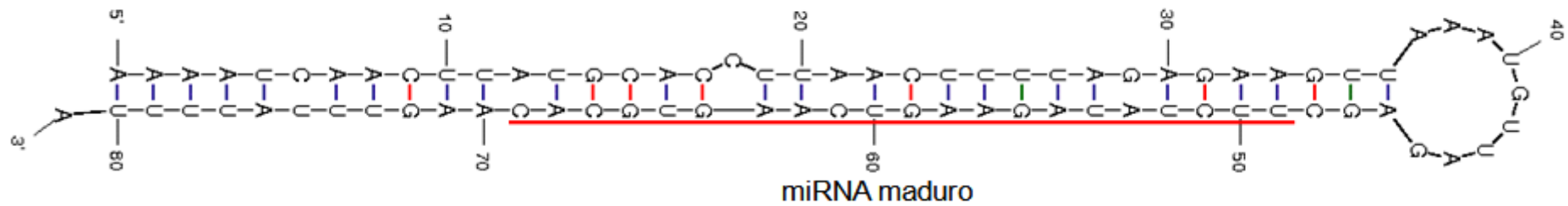
gma-MIR-Cand01 (Contig5263)**gma-MIR-Cand02 (Contig1834)**

$\Delta G = -29,10$ kcal/mol



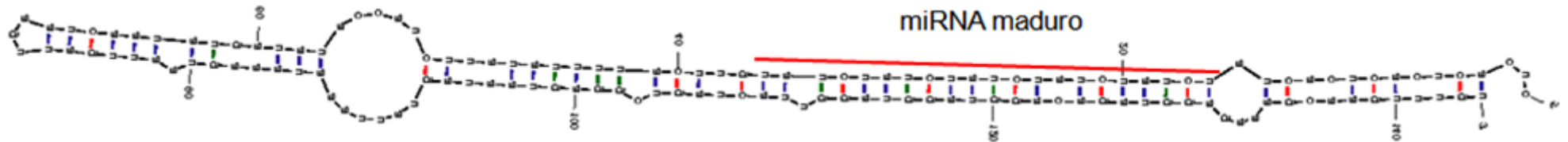
gma-MIR-Cand03 (Contig24974)

$\Delta G = -25,50$ kcal/mol



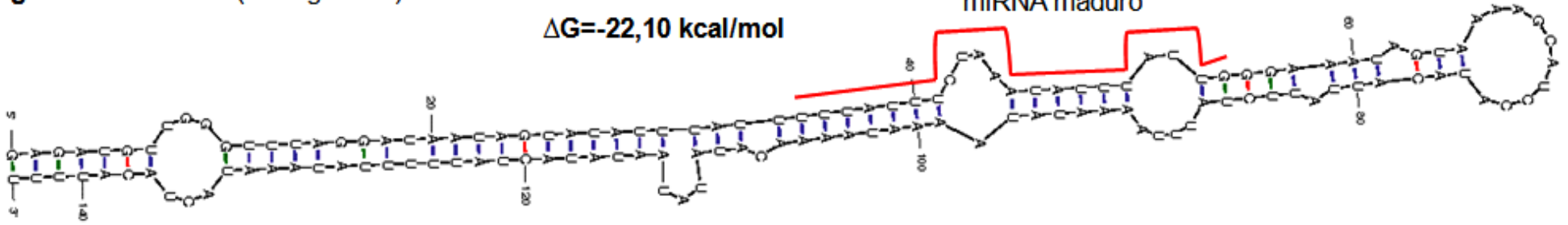
gma-MIR-Cand04 (SJ10-E1-R05-043-H05-UC.F)

$\Delta G = -32,20$ kcal/mol



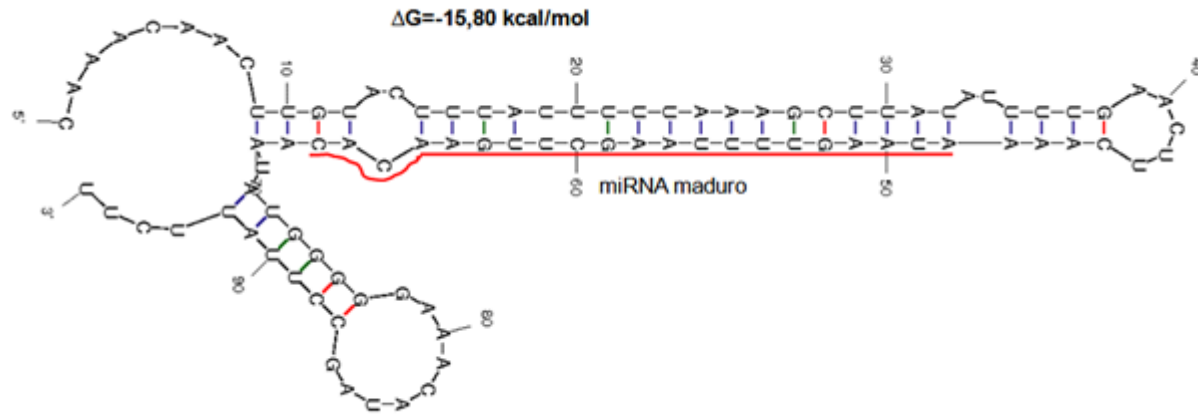
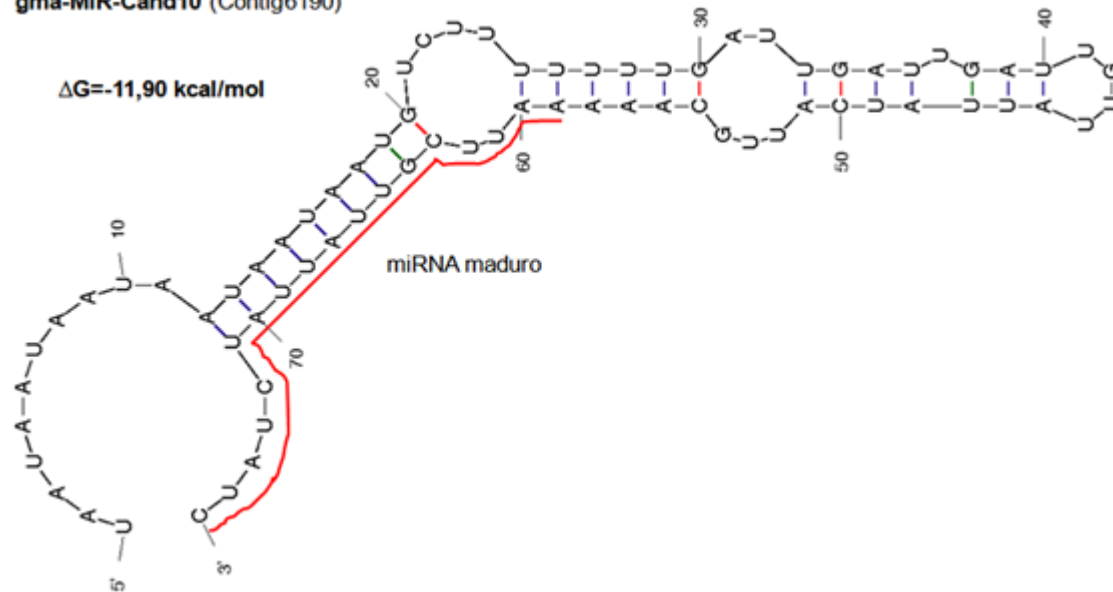
gma-MIR-Cand05 (Contig16659) $\Delta G = -22,10 \text{ kcal/mol}$

miRNA maduro

**gma-MIR-Cand06 (Contig13869)** $\Delta G = -16,90 \text{ kcal/mol}$

miRNA maduro



gma-MIR-Cand09 (Contig24959)**gma-MIR-Cand10 (Contig6190)**

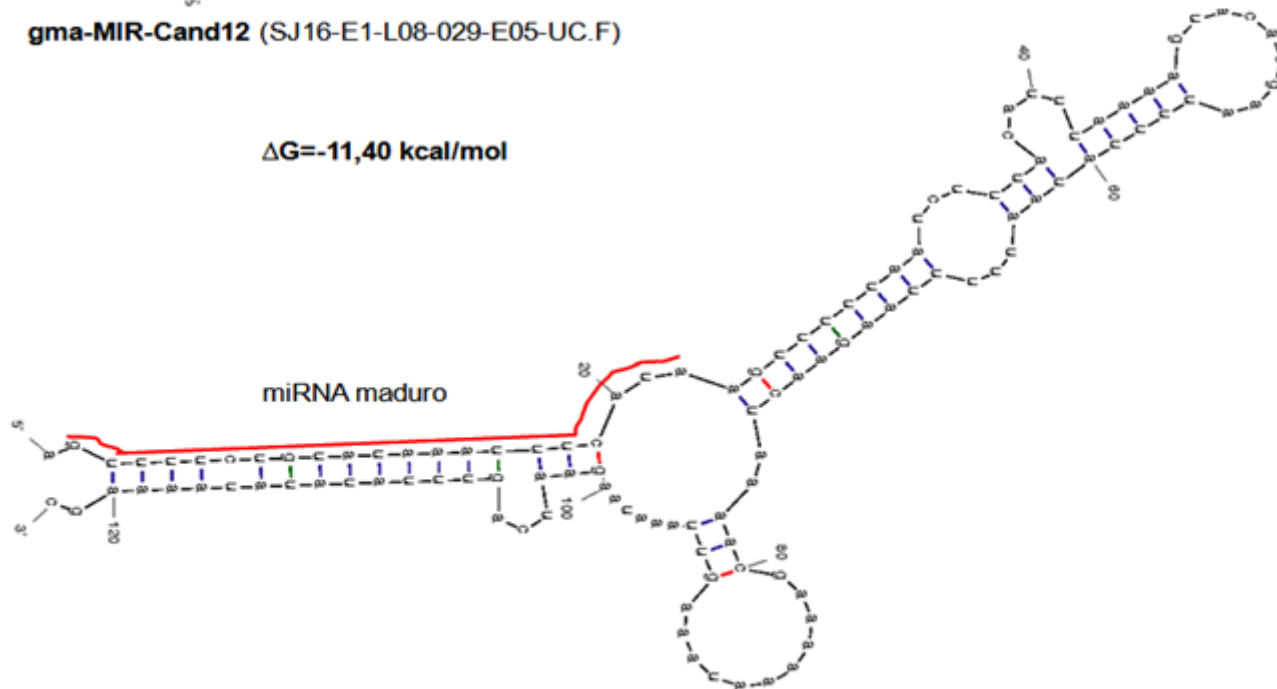
gma-MIR-Cand11 (Contig29200)

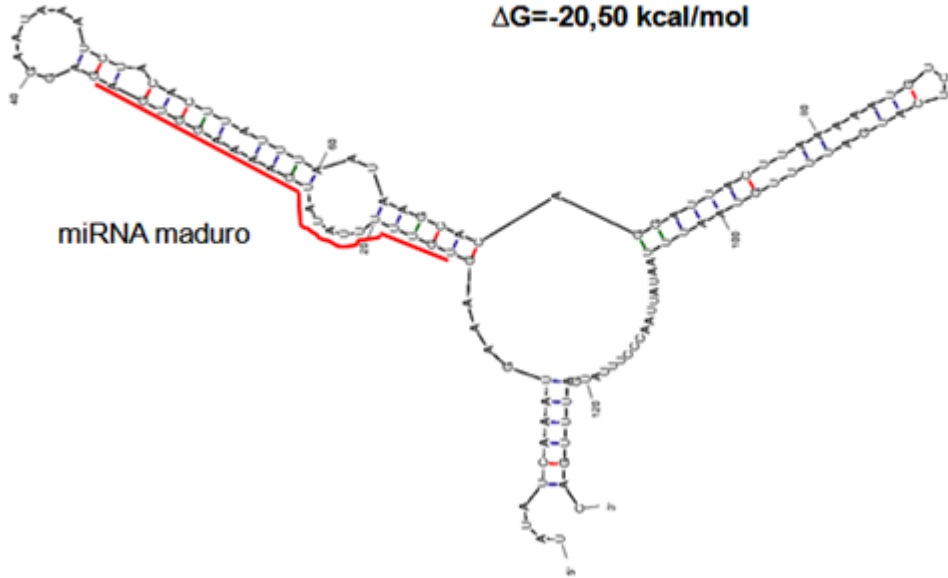
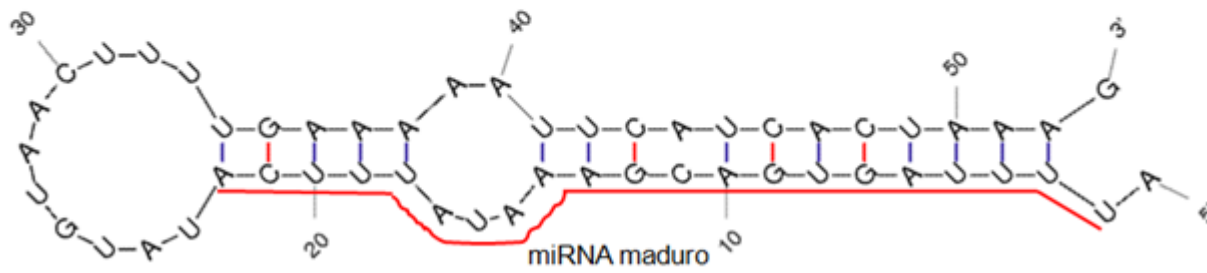
$\Delta G = -23,20$ kcal/mol



gma-MIR-Cand12 (SJ16-E1-L08-029-E05-UC.F)

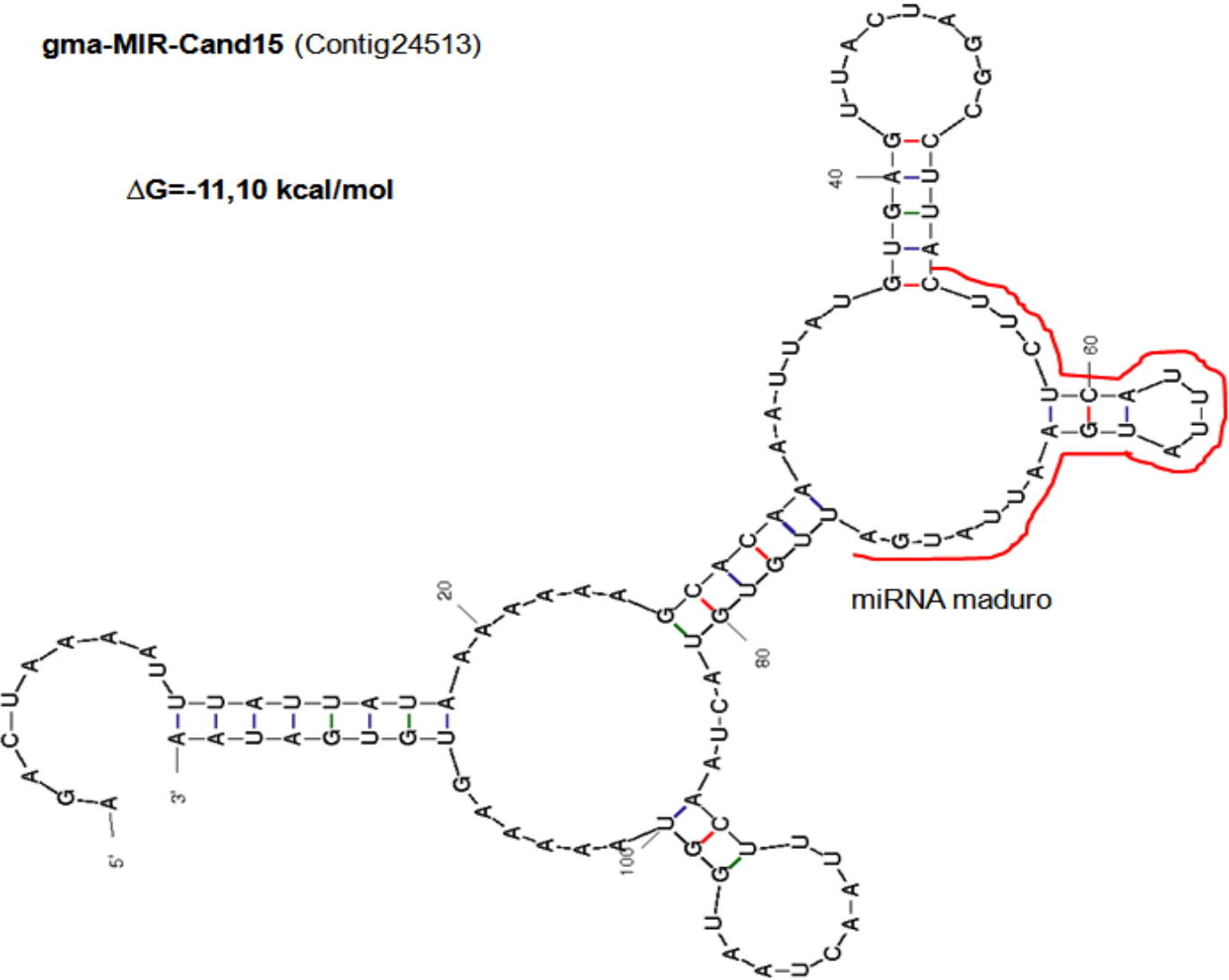
$\Delta G = -11,40$ kcal/mol

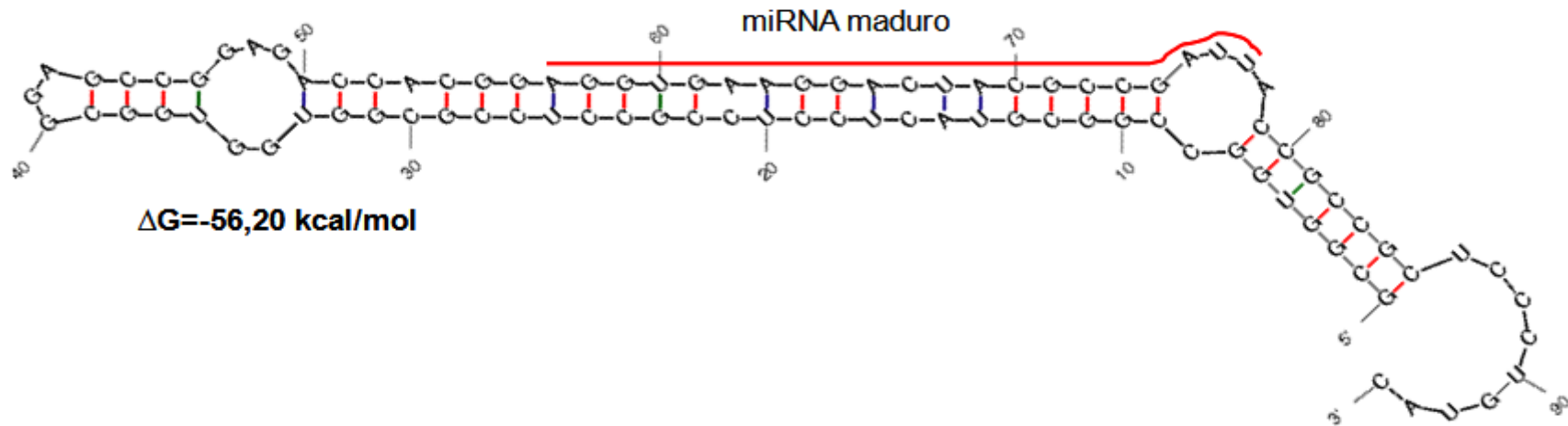
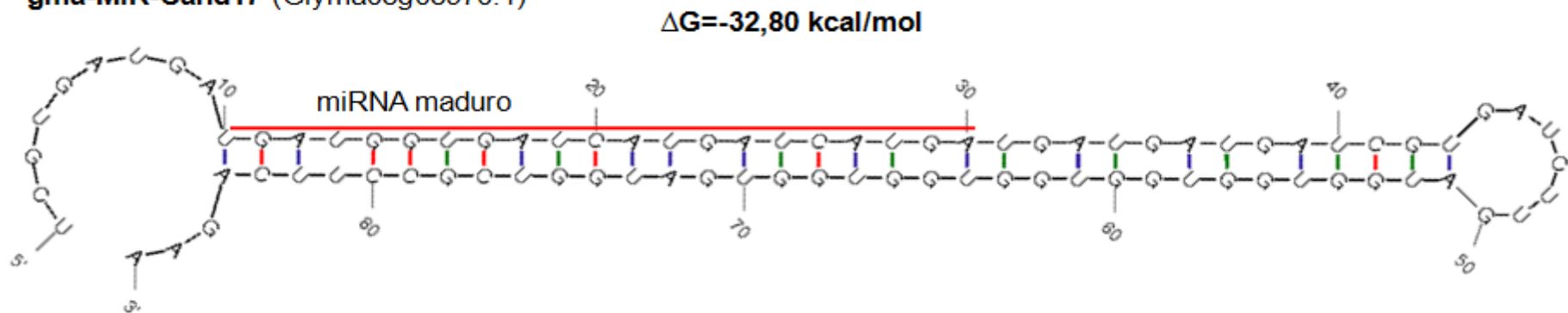


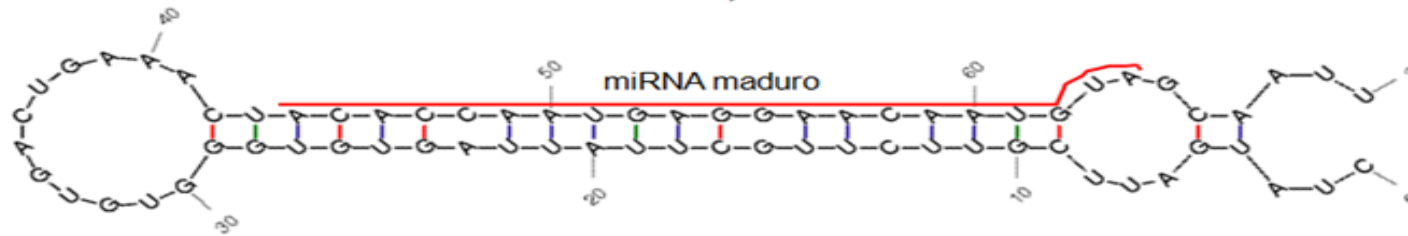
gma-MIR-Cand13 (Contig7552) $\Delta G = -20,50$ kcal/mol**gma-MIR-Cand14 (Contig27883)** $\Delta G = -11,70$ kcal/mol

gma-MIR-Cand15 (Contig24513)

$\Delta G = -11,10$ kcal/mol



gma-MIR-Cand16 (Contig21304)**gma-MIR-Cand17 (Glyma08g08970.1)**

gma-MIR-Cand18 (Contig19125) $\Delta G = -17,00 \text{ kcal/mol}$ **gma-MIR-Cand19 (Contig8009)** $\Delta G = -13,30 \text{ kcal/mol}$ 