

## Identificação computacional de genes miRNA e seus alvos no genoma da soja

---

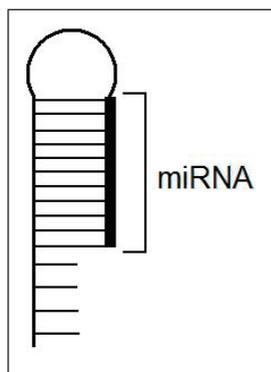
SILLA, P.R.<sup>1</sup>; CAMARGO-BRUNETTO, M.A. de O.<sup>1</sup>; BINNECK, E.<sup>2</sup>

<sup>1</sup>Universidade Estadual de Londrina – UEL, paulo.silla@gmail.com; <sup>2</sup>Embrapa Soja

No último século, muitas informações a respeito da estrutura molecular dos organismos vivos foram decifradas e, nas últimas décadas, a velocidade e facilidade de identificação dos constituintes moleculares foi melhorada pelo sequenciamento do genoma de vários organismos, de modo que muitos processos que ocorrem nas células foram caracterizados (Bruggeman et al., 2007). Um dos importantes avanços recentes da pesquisa biológica foi a descoberta de moléculas de RNA versáteis que regulam a expressão dos genes (Ghosh et al., 2007). Tais moléculas, chamadas miRNA (microRNA), possuem em torno de 22 nucleotídeos e são importantes reguladoras da expressão gênica ao nível pós-transcricional (Sunkar & Jagadeeswaran, 2008). Geralmente, o miRNA regula de modo negativo o produto de seu gene alvo (clivagem e degradação ou atenuação da tradução), mas existem exceções em que a influência é positiva (Meyers et al., 2006).

Entender a origem e a estrutura dos genes miRNA é um importante fator e serve como base para sua posterior predição. Mendes et al. (2009), Ghosh et al. (2007) e Chen (2005) apresentam de modo detalhado a biogênese das moléculas de miRNA: as proteínas que estão envolvidas na transformação do miRNA primário (pri-miRNA) em miRNA precursor (pre-miRNA) no núcleo da célula, a passagem deste para o citoplasma,

o processo de maturação do miRNA e sua ligação com o complexo de silenciamento induzido por RNA (*RNA-Induced Silencing Complex* – *RISC*) que, enfim, atua sobre o RNA mensageiro realizando papel regulatório. A Fig. 1 mostra a estrutura de um pre-miRNA, destacando o miRNA que será formado ao fim do processo.



**Fig. 1.** Precursor de um miRNA (pre-miRNA), gerado no núcleo da célula e enviado ao citoplasma pela ação de proteínas transportadoras.

Existem muitos problemas computacionais associados ao contexto dos miRNAs. Os mais importantes e, também, aqueles que têm chamado mais atenção são a busca por genes miRNA e a predição de seus alvos (Mendes et al., 2009). Genes miRNA candidatos podem ser identificados a partir do genoma e/ou de sequências EST através de procura pela presença de formação de potenciais “hairpin” (pareamento de bases) englobando a sequência pronta do miRNA existente na biblioteca (Subramanian et al., 2008). As abordagens mais utilizadas para a busca por genes miRNA em animais e plantas são as seguintes:

- Baseada em filtro: consiste na identificação de pequenos conjuntos com alta qualidade de miRNA candidatos conservados, os quais possuem boa chance de serem confirmados biologicamente como miRNAs verdadeiros (Mendes et al., 2009). A aplicação deste método é limitada devido ao seu alto custo, consumo de tempo e aleatoriedade

dos resultados (Zhang et al., 2005). A maioria dos miRNAs conhecidos, porém, foram identificados com essa abordagem (Meyers et al., 2006);

- Alvo centralizado: nova estratégia que beneficiou a criação de alguns pressupostos sobre a estrutura dos precursores de miRNA, a fim de identificar miRNAs em larga escala (Mendes et al., 2009).
- Baseada em homologia: miRNAs conservados entre espécies podem ser encontrados com este método. Em plantas, porém, a maioria dos genes miRNA não são conservados, o que limita a utilização dessa abordagem (Zhang et al., 2005). Outro fator negativo da busca baseada em homologia é que ela não identifica novos genes miRNA.
- Aprendizado de máquina: aplicados recentemente no descobrimento de miRNA, métodos derivados de aprendizado de máquina têm apresentado bons resultados (Yousef et al., 2009). Basicamente, esta técnica necessita de um conjunto de dados de miRNAs conhecidos para ser utilizado como parâmetro de treinamento. Após o treinamento, o conjunto de teste formado por dados onde se deseja descobrir novos miRNA é apresentado ao algoritmo e este estabelece os prováveis miRNAs existentes. Em seguida, uma validação biológica é necessária para que os miRNAs identificados sejam confirmados como miRNAs verdadeiros ou descartados.

Sabe-se que os genes miRNA afetam numerosos aspectos da biologia da planta, especialmente seus padrões de crescimento e desenvolvimento, ou mesmo em mecanismos de defesa contra estresses bióticos e abióticos (Jones-Rhoades et al., 2006). Uma vez identificado o gene miRNA, o passo seguinte é saber qual a sua influência na expressão gênica do organismo. Em plantas, existe uma complementaridade quase perfeita entre os miRNAs e seus alvos, o que facilita esta tarefa (Ghosh et al., 2007).

Em contrapartida, Bentwich (2005) diz que a predição computacional dos alvos de miRNAs apresenta-se como uma tarefa desafiadora, devido à não existência de um grupo suficientemente grande de alvos de miRNA conhecidos que possa ser utilizado como conjunto de treinamento e também por conta da dificuldade de validação, que é mais complexa que a validação do miRNA.

Os objetivos gerais deste trabalho são:

- Identificar genes miRNA na soja, utilizando as abordagens baseadas em homologia e aprendizado de máquina.
- Estabelecer o alvo dos miRNAs identificados (qual o seu papel na regulação da expressão gênica).

Os objetivos específicos são:

- Desenvolver um conjunto de algoritmos que possam ser utilizados na detecção de genes miRNA em plantas e predição de seus alvos.
- Testar, adaptar e otimizar no genoma da soja métodos computacionais e seus algoritmos já desenvolvidos para predição de miRNAs/precursosores de plantas e seus alvos.

Durante a realização deste trabalho, a busca por homologia e o aprendizado de máquina serão utilizados a fim de identificar miRNA na soja. Para o caso da busca por homologia, o genoma da soja será confrontado, utilizando algoritmos de pareamento, com dados de miRNA existentes de outras espécies de plantas, obtidos em bancos de dados públicos disponíveis na Internet. Nesse sentido, existe uma série de algoritmos prontos que deverão ser utilizados.

O próximo passo será utilizar técnicas de aprendizagem de máquina com o objetivo de encontrar genes miRNAs ainda não descobertos e/ou exclusivos na Soja. Durante essa etapa, pode-se utilizar algum

conjunto de *software* já existente ou então, como forma de tornar a pesquisa mais flexível, desenvolver novos algoritmos, utilizando modelos estocásticos (como Cadeias Ocultas de Markov), Redes Neurais Artificiais, Lógica *Fuzzy*, técnicas de Computação Evolutiva ou um modelo híbrido que combine as principais vantagens de cada um. Novamente, dados de expressão serão obtidos em bancos de dados públicos disponíveis para *download* na Internet.

Não existem regras que possam ser aplicadas para distinguir absolutamente precursores miRNAs de outras sequências (Johnson & Sundaesan, 2007). Logo, o conjunto de candidatos obtidos nas etapas anteriores será analisado biologicamente, a fim de estabelecer quais realmente são miRNAs.

Uma vez definido o conjunto final de miRNAs (validados biologicamente), estes serão estudados a fim de estabelecer seu papel no organismo, ou seja, seus alvos regulatórios. Nesta etapa, pretende-se desenvolver um método computacional novo, empregando alguma das técnicas citadas no passo de busca de genes miRNA.

Finalmente, uma nova análise biológica deverá ser realizada sobre os alvos hipoteticamente controlados pelos miRNAs encontrados, com o objetivo de validar ou não essas informações.

Todos os algoritmos desenvolvidos serão analisados e aqueles que necessitarem de alto poder computacional e possuírem características de natureza paralela, serão escritos desse modo e executados em um aglomerado (*cluster*) de computadores que, por sua vez, será construído utilizando as estações de trabalho disponíveis no laboratório de bioinformática da Embrapa Soja.

## Referências

BENTWICH, B. Prediction and validation of microRNAs and their targets. **FEBS Letters**, v. 579, p. 5904–5910, 2005.

BRUGGEMAN, F. J.; HORNBERG, J. J.; BOOGERD, F. C.; WESTERHOFF, H. V. Introduction to systems biology. In: BAGINSKY, S.; FERNIE, A. R. (Ed.). **Plant systems biology**. Basel: Birkhäuser Verlag, 2007. EXS 97, p. 1–20.

CHEN, X. MicroRNA biogenesis and function in plants. **FEBS Letters**, v. 579, p. 5923–5931, 2005.

GHOSH, Z.; CHAKRABARTI, J.; MALLICK, B. Mirnomics - The bioinformatics of microRNA genes. **Biochemical and Biophysical Research Communications**, v. 363, p. 6–11, 2007.

JOHNSON, C.; SUNDARESAN, V. Regulatory small RNA in plants. In: BAGINSKY, S.; FERNIE, A. R. (Ed.). **Plant systems biology**. Basel: Birkhäuser Verlag, 2007. EXS 97, p. 99–113.

JONES-RHOADES, M. W.; BARTEL, D. P.; BARTEL, B. MicroRNAs and their regulatory roles in plants. **Annual Review of Plant Biology**, v. 57, p. 19–53, 2006.

MENDES, N. D.; FREITAS, A. T.; SAGOT, M.-F. Current tools for the identification of miRNA genes and their targets. **Nucleic Acids Research**, v. 37, p. 2419–2433, 2009.

MEYERS, B. C.; SOURET, F. F.; LU, C.; GREEN, P. J. Sweating the small stuff: microRNA discovery in plants. **Current Opinion in Biotechnology**, v. 17, p. 139–146, 2006.

SUBRAMANIAN, S.; FU, Y.; SUNKAR, R.; BARBAZUK, W. B.; ZHU, J. K.; YU, O. Novel and nodulation-regulated microRNA in soybean roots. **BMC Genomics**, v. 9, 2008.

SUNKAR, R.; JAGADEESWARAN, G. In silico identification of conserved microRNA in large number of diverse plant species. **BMC Plant Biology**, v. 8:37, 2008.

YOUSEF, M.; SHOWE, L.; SHOWE, M. A study of microRNAs in silico and in vivo: bioinformatics approaches to microRNA discovery and target identification. **FEBS Journal**, v. 276, p. 2150–2156, 2009.

ZHANG, B.; PAN, X.; COBB, G. P.; ANDERSON, T. A. Plant microRNA: a small regulatory molecule with big impact. **Developmental Biology**, v. 289, p. 3–16, 2005