

拟南芥 miRNA 及其靶序列配对特征分析

刘冲^{1,2}, 孙杰¹, 范云六², 王磊²

(1. 石河子大学农学院, 绿洲生态实验室, 新疆 832003;
2. 中国农业科学院生物技术研究所, 国家农作物基因资源与基因改良重大科学工程, 北京 100081)

摘要: microRNAs (miRNAs) 是真核生物中一类长度约为 22 nt 的非编码小分子 RNA, miRNA 与 AGO 等蛋白形成 RISC 沉默复合物, 通过剪切或翻译抑制对靶基因起负调控作用。对拟南芥 miRNA 序列及其配对的靶序列间的特征进行了统计分析, 结果表明 miRNA 序列 5' 端富含 A、U, 第 1、第 19 碱基位对 U、C 具有较强的倾向性; miRNA 与靶序列间常有 1~4 个碱基错配, 错配碱基常出现在第 1、第 2 和第 21 位, 而第 3~第 6, 第 9~第 10, 第 16~第 17 碱基配对较为保守, 为人工合成 miRNA 的设计及 miRNA 靶基因的预测以提供了依据。

关键词: 拟南芥; 基因沉默; RNAi; 微小 RNA; 人工小 RNA

中图分类号: Q522+.6 文献标识码: A 文章编号: 1008-0864(2009)04-0053-05

Characterization of *Arabidopsis* microRNAs and Matching to its Targets

LIU Chong^{1,2}, SUN Jie¹, FAN Yun-liu², WANG Lei²

(1. College of Agriculture, Key Laboratory of Oasis Ecology Agriculture, Shihezi University, Xinjiang Shihezi 832003;

2. Biotechnology Research Institute, National Key Facility of Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing 100081, China)

Abstract: MicroRNAs (miRNAs) are a class of endogenous non-coding single-stranded RNA with about 22 nucleotides length. MicroRNAs function as sequence-specific negative regulators in post-transcriptional gene silencing by base pairing with target mRNAs, which lead to mRNA cleavage or translational repression. We examined the sequence characteristics of all known *Arabidopsis* miRNAs and the base-pairing feature between miRNAs and their targets. The miRNA sequences displayed higher A and U content at the 5' end, and higher U content at position 1, and higher C content around position 19. There often exist one to four mismatching between miRNAs and their targets, which showed that mismatches occurred frequently at position 1, 2 or 21, but hardly occurred at position 3~6, 9~10, 16~17. The data is potentially applicable to the design of artificial microRNA and prediction of microRNAs targets.

Key words: *Arabidopsis*; gene silencing; RNAi; microRNA; artificial microRNA (amiRNA)

microRNAs (miRNAs) 是近年来发现的一类约 22 nt 的小分子 RNA, 广泛存在于真核生物体内。miRNAs 作为一类非编码 RNA (non-coding RNA), 在基因组上通常具有独立的基因座位 (locus), 在 RNA 聚合酶 II 作用下合成 pri-miRNA (primary miRNA, 初始 miRNA), pri-miRNA 自身折叠成不完全配对的发卡结构 (hairpin), 在核酸酶的作用下产生约 21 nt 大小的成熟 miRNA^[1]。

成熟的 miRNAs 首先与 AGO 等蛋白形成

RISC (RNA induced silencing complex) 复合物, RISC 与靶基因的 mRNA 以碱基配对方式结合, 干扰基因的表达^[2]。植物 miRNA 与其靶 mRNA 碱基互补配对后主要通过靶基因 mRNAs 的切割或翻译抑制来调节基因的表达。miRNA 参与了多种生物学功能的调控, 如细胞增殖、组织发育、分化、凋亡、癌变、环境应答以及病毒感染等。

RNA 干扰 (RNAi) 是近年来发现的普遍存在于生物界中调节 mRNA 的生物学现象, miRNA

收稿日期: 2009-03-31; 修回日期: 2009-06-22

基金项目: 国家 863 计划项目 (2007AA10Z147); 绿洲生态农业重点实验室开放课题发展基金 (200704) 资助。

作者简介: 刘冲, 硕士研究生, 研究方向为植物代谢工程。E-mail: liu_chong02@163.com。通讯作者: 王磊, 研究员, 博士, 从事基因沉默的研究。E-mail: wanglei@caas.net.cn

是也 RNAi 现象的一种,是内源性非编码的单链 RNA,大多数 miRNA 在进化上具有高度保守性。实验证明,miRNA 已成为继 siRNA 之后 RNA 干扰领域一种重要的研究手段和实验工具^[3~6]。本文运用生物信息学的方法统计了拟南芥内源 miRNA 的序列特征,以及 miRNA 与其靶基因配对序列的规律,从而为 miRNA 靶基因的预测以及设计合成人工 miRNA 时序列的选择提供了依据。

1 材料与方法

1.1 选用的 miRNA 序列

拟南芥 (*Arabidopsis thaliana*) 小分子 RNA 序列共 254 条,来源于美国国立生物技术信息中心 (National Center for Biotechnology Information, NCBI) GeneBank 数据库 (<http://www.ncbi.nlm.nih.gov/Genbank>) 和 The miRNA Registry (<http://www.sanger.ac.uk/>) 数据库及其他文献报道序列。

1.2 对候选 miRNA 与靶基因碱基配对的统计方法

候选源 miRNA 序列通过比对数据库的靶基因,统计 miRNA 序列碱基出现的频率和靶基因碱基配对的规律,以及错配碱基出现的频率,方法参考 WebLogo 统计方法^[7]。

2 结果

2.1 拟南芥 miRNAs 及其前体的序列特征

经统计分析发现 miRNA 的长度一般为约

22 nt,所有的 miRNA 都在 15 ~ 30 nt 之间,其中 21 ~ 23 nt 长度的 miRNA 占大多数,约为 84%。成熟的 miRNA 的 5' 端有一个磷酸基团,3' 端为羟基且通常被甲基化,这一特点使它与大多数寡核苷酸和功能 RNA 降解片段区分开,这也是 miRNA 一些的基本特征。

在对已发现的拟南芥 miRNA 序列分析后发现:①miRNA 序列第 1 和第 19 碱基位较为保守,第 1 个碱基是 U 的序列占总数的 58%,第 19 个碱基是 C 的占总数的 35%。②第 7、第 8、第 10、第 20 碱基位出现 A、G、A、U 的概率分别是 18%、16.7%、12% 和 13.2%。③其他碱基位的碱基几乎是随机出现(图 1)。④miRNA 碱基组成的特点为:miRNA 5' 端富含 A、U,3' 端富含 G、C。这可能有利于 miRNA 双链在 5' 端解链和 RISC 复合体的形成。

与动物 miRNA 相比,植物 miRNA 的前体结构(stem-loop)更大、更复杂,其前体常形成分子内茎环结构并且含有大量的 U:G 碱基对。大多数 miRNA 基因以单拷贝、多拷贝或基因簇(cluster)的形式存在于基因组中。动物 miRNA 基因簇内,miRNA 基因排列紧密,常常是基因头尾相接,仅被短序列分开。另外,一些 miRNA 的基因结构和功能在进化中呈现保守性。约 12% 的 miRNA 在线虫、果蝇、哺乳动物和植物中呈现保守性。在秀丽隐杆线虫 (*C. elegans*) 中所发现的 miRNA 85% 都可以在桅杆线虫 (*C. briggsae*) 的基因组中找到同源 miRNA^[8]。

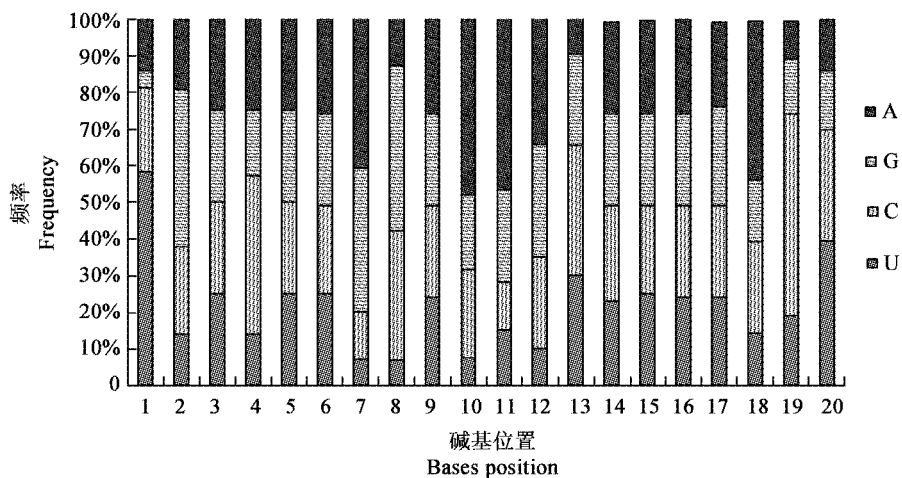


图 1 拟南芥中 miRNA 序列碱基偏好性分布特征

Fig. 1 Base biases of endogenous miRNAs from *Arabidopsis thaliana*.

2.2 miRNA 和靶基因碱基配对分析

miRNA 通过与靶 mRNA 碱基互补配对,引起靶 mRNA 的降解或翻译抑制。为了了解这种序列配对特征,我们对拟南芥的 miRNA 和靶基因序列进行了分析,结果表明:①从 miRNA 碱基位来看,miRNA 与靶基因的错配碱基常出现在 3' 端和 5' 端的末端,在第 1 和第 21 碱基位碱基错配率分别为 58.3% 和 43%。其次第 2、第 7、第 8、第 13、第 15 和第 20 碱基位的碱基错配比率都超过了 20% (图 2)。②miRNA 的第 3~第 4、第 9~第 12、第 16~第 17 等碱基位和靶基因碱基配对较为保守(图 2)。③在 254 条 miRNA 中,和靶基因完全配对的只有 5 条,约占 1.8%;只有 1 个碱基错配的序列有 42 条,占总数的 15%;2 个碱基错

配序列有 68 条,占总数的 24%;3 个碱基错配序列有 92 条,占总数的 37%;4 个碱基错配序列有 41 条,占总数的 14% (图 3)。④拟南芥中的错配碱基多数是 G:U,但实验证实 G:U 的碱基错配并不影响 RNA 干扰和抑制翻译的作用,所以 G:U 可以被看成配对碱基^[9]。如果把 G:U 作为配对碱基来看,则拟南芥 miRNAs 与靶基因完全配对的碱基有 7 条;1 个碱基错配序列有 54 条,占总数的 21%;2 个碱基错配序列有 71 条,占总数的 27%;3 个碱基错配序列有 83 条,占总数的 33%;4 个碱基错配序列有 41 条,占总数的 13% (图 3)。说明 miRNA 与其调控的靶基因间的错配碱基数大多数为 1~3 个。

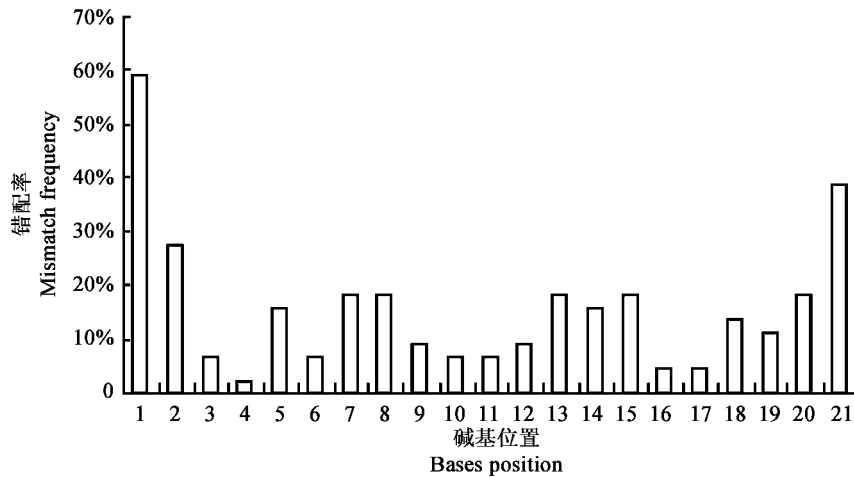


图 2 拟南芥中 miRNAs 与靶基因在不同碱基位置的错配机率

Fig. 2 Mismatch frequency of *Arabidopsis* miRNAs and their targets at different position.

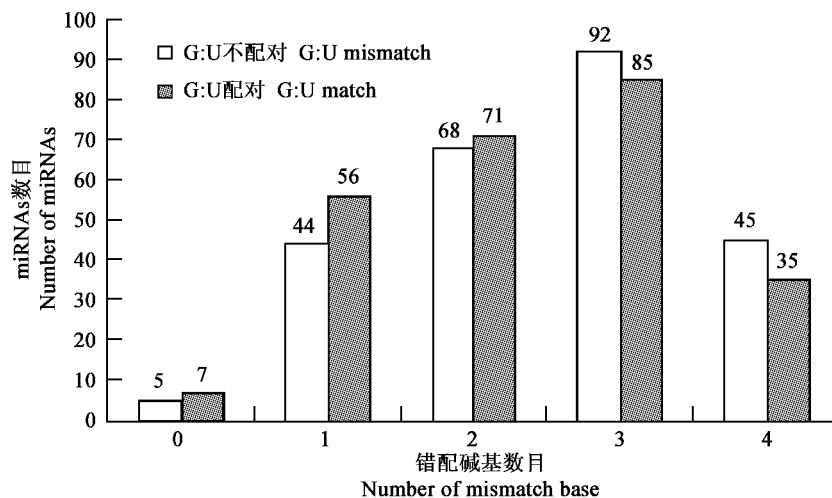


图 3 拟南芥 miRNAs 及其靶基因间错配碱基统计

Fig. 3 Statistic of mismatch between *Arabidopsis* miRNAs and their targets.

3 讨论

人工合成 miRNA 同样能够在生物体进行生物代谢,干扰其目的基因的表达,引起目的基因沉默的效果,但沉默的效率不尽相同。对 miRNA 序列特征以及和靶基因碱基配对特征的分析表明,miRNA 序列 5'端富含 A、C,并且序列第 1 个碱基对 U 有强烈的倾向性,而对 G 却有抗性,第 2~第 4 个碱基缺乏 U,第 8 个碱基偏向 G,第 19 个碱基偏向 C,这是拟南芥内源 miRNA 碱基组成的特点。miRNA 序列第 1 和第 19 碱基是 U 和 C,同 Stephan Ossowski 报道的一致^[10];而且 miRNA 5'端的 2~8 个核苷酸几乎无一例外地与靶 mRNA 3'端 UTR 区完全互补^[11]。植物中 miRNA 与靶基因的碱基配对较为保守,仅在 miRNA 3'端和 5'端

末端有少量碱基不配对。miRNA 特点及其与靶基因碱基配对的规律为人工 miRNA (amiRNA) 沉默载体的设计提供了有利的依据。

在 miRNA 与靶基因的相互作用中,人们利用 miRNA 和靶基因碱基配对的特点,开发出了一些 amiRNA 序列分析软件,如 WMD 等(表 1)。由于可供选择的 amiRNA 序列很多,如何从中选择一个更好的 amiRNA 依然是人们面临的一个问题。因此本文通过对已有的 miRNA 序列及其与靶基因的互作规律分析,统计出拟南芥 miRNA 的序列特征及其与靶序列碱基配对的特点;在此基础上,结合 miRNA 与靶基因的热力学稳定性和二级结构分析软件,如 MFold、RNAfold 和 RNAhybrid 等(表 1),对高效 amiRNA 的设计、筛选具有重要的参考价值。

表 1 目前在线 miRNA 靶基因预测工具

Table 1 Online resources for miRNA target predictions.

数据库名称 Method	使用方法 Method availability	网址 Website	参考文献 Reference
WMD	在线 online search	http://wmd2.weigelworld.org	[12]
miRanda	下载 download	http://www.microrna.org/miranda	[13]
mirBase	在线 online search	http://pictar.bio.nyu.edu/	
picTar	在线 online search	http://microrna.sanger.ac.uk/targets/v2/	
TargetscanS	在线 online search	http://www.targetscan.org/	[14]
RNAhybrid	下载 download	http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/	[10]
DIANA-MicroT	下载 download	http://diana.pcbi.upenn.edu	[10]
Tarbase	在线 online search	http://diana.pcbi.upenn.edu/tarbase	

参 考 文 献

- [1] Lee Y, Ahn C, Han J, *et al.*. The nuclear RNase III Drosha initiates microRNA processing [J]. *Nature*, 2003, 425: 415-419.
- [2] Hutvagner G. Small RNA asymmetry in RNAi: Function in RISC assembly and gene regulation [J]. *FEBS Lett.*, 2005, 579: 5850-5857.
- [3] Niu Q W, Lin S S, Reyes J L, *et al.*. Expression of artificial microRNAs in transgenic *Arabidopsis thaliana* confers virus resistance [J]. *Nat. Biotechnol.*, 2006, 11: 1420-1428.
- [4] Warthmann N, Chen H, Ossowski S, *et al.*. Highly specific gene silencing by artificial miRNAs in rice [J]. *PLoS ONE*, 2008, 3(3): e1829.
- [5] Schwab R, Ossowski S, Riester M, *et al.*. Highly specific gene silencing by artificial microRNAs in *Arabidopsis* [J]. *Plant Cell*, 2006, 18: 1121-1133.
- [6] Khraiweh B, Ossowski S, Weigel D, *et al.*. Specific gene silencing by artificial microRNAs in *Physcomitrella patens*: An alternative to targeted gene knockouts [J]. *Plant Physiol.*, 2008, 148: 684-693.
- [7] Crooks G E, Hon G, Chandonia J M, *et al.*. WebLogo: a sequence logo generator [J]. *Genome Res.*, 2004, 14: 1188-1190.
- [8] Lau N C, Lim L P, Weinstein E G. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans* [J]. *Science*, 2001, 294(5543): 858-862.
- [9] Schwab R, Palatnik J f, Riester M, *et al.*. Specific effects of

- microRNAs on the plant transcriptome[J]. *Dev. Cell*, 2005, 8:517-527.
- [10] Kluiver J, Van Den Berg A, De Jong D, *et al.* Regulation of pri-microRNA BIC transcription and processing in *Burkitt lymphoma*[J]. *Oncogene*, 2007, 26(26):3769-3776.
- [11] Lewis B P, Burge C B, Bartel D P. Conserved seed pairing, often lanked by adenosines, indicates that thousands of human genes are microRNA targets[J]. *Cell*, 2005, 120:15-20.
- [12] Ossowski S, Schwab R A D W. Gene silencing in plants using artificial microRNAs and other small RNAs[J]. *Plant Cell*, 2007, 53:674-690.
- [13] Mourelatos Z, Dostie J, Paushkin S, *et al.* miRNPs; a novel class of ribonucleoproteins containing numerous microRNAs[J]. *Genes Dev.*, 2006, 16(6):720-728.
- [14] Josee D, Zissimos M, Michael Y, *et al.* Numerous microRNPs in neuronal cells containing novel microRNAs[J]. *RNA*, 2003(9):180-186.

【863 课题介绍】

课题名称:油菜 miR169 家族及新的候选 miRNA 家族的功能鉴定

课题编号:2007AA10Z147

课题内容、目标:

①分离油菜 miR169 家族的前体,构建过表达载体并转化油菜,鉴定 miR169 家族的功能。

②Northern 杂交分析 miR169 在油菜中的表达特性。

③分离 miR169 的靶基因 *BnHAP*,通过在油菜中的过表达分析油菜 *BnHAP* 的功能。

④研究 *bn-miR169* 及其与靶基因之间的互作关系,阐明 miR169 在植物中的调控作用。

课题进展:

已分离了 miR169 前体,构建植物表达载体,获得了转基因拟南芥;分离了 miR169 的靶基因 *HAP*,并对其进行了突变,构建了相应的植物表达载体,获得了转基因植株,正在对其进行功能分析。