

文章编号: 1674-7054(2012)03-0222-06

小果野蕉 microRNAs 及其靶基因的生物信息学预测

叶可勇¹ 陈瑶² 李瑞梅¹ 符少萍¹ 郭建春¹

(1. 中国热带农业科学院 热带生物技术研究所在海南海口 571101;

2. 中国热带农业科学院 科技信息研究所 海南海口 571101)

摘要: 通过基于表达序列标签(EST)、基因组勘测序列(GSS)和高通量测序基因组序列(HTGS)的同源序列比对搜索,以及一系列的标准筛选,最终预测到属于8个家族的10条小果野蕉 miRNAs。在线软件 psRNA-Target 预测到24对 miRNAs 与靶基因的互作,这些靶基因主要参与小果野蕉的新陈代谢、生长发育以及胁迫响应等过程。

关键词: 小果野蕉; microRNAs (miRNAs); 靶基因; 生物信息学

中图分类号: S 668.1

文献标志码: A

小果野蕉(*Musa acuminata*)为姜目(Zingiberales)芭蕉科(Musaceae)芭蕉属的植物,分布于印度北部、缅甸、泰国、越南、马来西亚、菲律宾以及我国云南、广西等地,多生长于海拔1 000~1 200 m的阴湿的沟谷、沼泽、半沼泽及坡地上,具有适应性强、分布广等特点。小果野蕉是目前世界上栽培香蕉的亲本种之一,其假茎可作猪饲料^[1]。

microRNAs(miRNA)普遍存在于真核生物中,是一类内源、非编码的单链小分子RNA,其前体具有茎-环结构。在植物体内具有高度保守性、时序性和组织特异性^[2-3]。在转录后水平,成熟的miRNA通过特异性的碱基互补配对与靶基因结合,从而降解靶基因或抑制其翻译来调控基因的表达^[4]。相比于双链siRNA,转录后的miRNA前体以单链的形式存在,并自身折叠形成发夹结构^[5]。miRNA前体通过Dicer 1酶(DCL1)切割,产生长约21 bp的miRNA:miRNA*,然后被运输到细胞质中。成熟的miRNA与RNA沉默复合体结合,其中1条miRNA保留在复合体中,通过与靶基因互补区配对来调控靶基因表达^[6]。

相比动物的miRNA,植物miRNA的研究起步晚,随着miRNA研究方法以及新一代测序技术的快速发展,植物中大量的miRNA被发现。2011年10月4日,miRNA数据库miRBase数据库升级至18.0版本,记录了168个物种中的18 266条miRNAs序列^[7]。到目前为止,对植物miRNA的研究主要集中在拟南芥、水稻、玉米等基因组已被解析的模式植物上,而在非模式植物中研究较少,miRBase数据库中尚未储存小果野蕉miRNAs的相关研究信息。本研究以miRNA的保守性为基础,利用生物信息学方法对小果野蕉的miRNA及其靶基因进行预测,为进一步研究小果野蕉的miRNA奠定基础。

1 材料与方法

1.1 miRBase中的miRNAs序列 在miRBase数据库(<http://www.mirbase.org/>)中下载拟南芥、水稻、玉米、毛果杨、大豆等53种植物的4 014条miRNAs序列。为避免重复的miRNAs序列产生冗余性结果,将碱基序列相同的miRNAs去除,以去重后剩余的miRNAs为参考序列,来预测小果野蕉中的miRNAs。

1.2 小果野蕉的EST、GSS和HTGS序列 小果野蕉的表达序列标签(EST)、基因组勘测序列(GSS)和

收稿日期: 2012-08-10

基金项目: 国家973重大基础性研究项目(2010CB126600); 国家高技术研究发展计划(863计划)(2009AA10Z102); 国家自然科学基金(30860151; 31170234; 31160061); 中央级公益性科研院所基本科研课题(ITBB110218); 国家木薯产业技术体系项目

作者简介: 叶可勇(1986-),男,浙江温州人,中国热带农业科学院热带生物技术研究所在海南海口研究实习生。

通信作者: 郭建春,女,博士,研究员, E-mail: jianchunguo@163.com

高通量测序基因组序列(HTGS) 在美国国家生物技术信息中心(National Center for Biotechnology Information , NCBI) 的 GenBank 核酸序列数据库(<http://www.ncbi.nlm.nih.gov>) 中下载。总共获得 15 560 条 EST 序列 31 458 条 GSS 序列 33 条 HTGS 序列。

1.3 小果野蕉 miRNAs 的生物信息学预测 小果野蕉 miRNAs 的预测流程如图 1 所示。首先, 将 miR-Base 中非重复的植物 miRNAs 序列与小果野蕉的 EST, GSS 和 HTGS 序列进行 blastn 比对, 允许小于等于 2 个的碱基错配, 初步筛选获得候选 miRNAs 序列。将候选 miRNAs 序列相互间进行比对, 去除重复的 miRNAs 序列。然后, 在 NCBI 蛋白质数据库中将筛选后获得的候选 miRNAs 序列进行 blastx (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) 比对, 去除编码蛋白质的序列。最后, 使用 Mfold(<http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi>) 软件对剩余的候选 miRNAs 前体序列的二级结构进行折叠, 分析筛选二级结构获得小果野蕉 miRNAs^[8]。

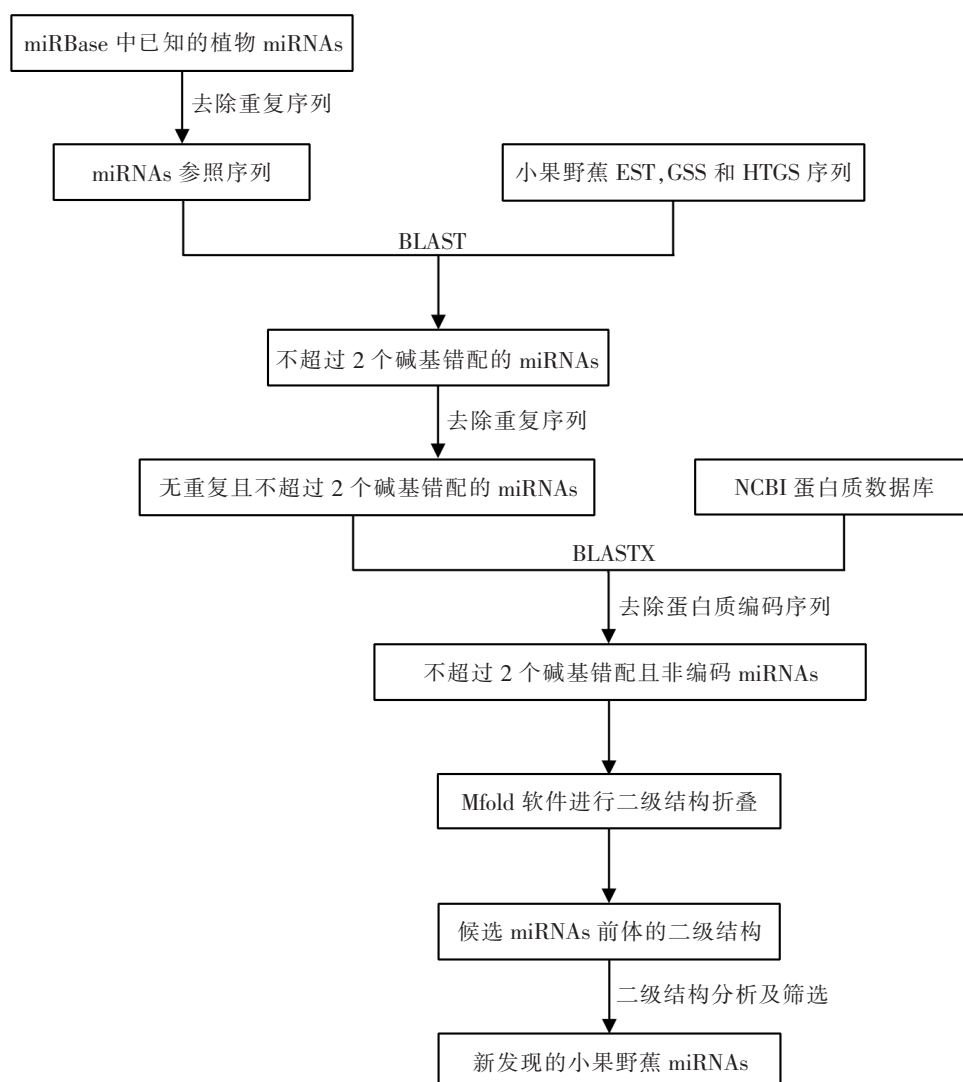


图1 小果野蕉 miRNAs 的生物信息学预测流程

小果野蕉 miRNAs 的筛选标准: 1) 小果野蕉新预测的 miRNAs 与参考 miRNAs 序列碱基错配数小于等于 2; 2) miRNAs 前体具有二级发夹结构, 并且成熟 miRNAs 序列与 miRBase 中同一家族的成熟 miRNAs 位于相同的臂上; 3) miRNAs 前体序列的最小折叠自由能(Minimal folding free energy , MFE) 绝对值要较高, 并且最小折叠自由能系数(Minimal folding free energy index , MFEI) 要大于 0.85; 4) 成熟的 miRNAs 与卫星 miRNA* 互补序列的错配碱基数不超过 5 个, 并且不允许有较大的环或缺口; 5) blastx 分析, E 值要大于 $1e-5$ ^[9]。

1.4 小果野蕉 miRNA 靶基因的预测 预测植物 miRNAs 靶基因的软件有多种,如 psRNATarget ,miRNAassist ,PatScan 和 miRU 等。本研究利用 psRNATarget 在线软件(<http://plantgrn.noble.org/psRNATarget/>),根据植物 miRNAs 与其靶基因具有高度的序列互补性,在小果野蕉 EST 数据库中预测其 miRNAs 的靶基因^[10]。最大期望值设置为 3,其他参数设置为默认。psRNATarget 对靶基因的预测是根据罚分机制来筛选,每 20 nt 长的 miRNAs 序列得分要小于等于 3,G:U 摆动配对的罚分值为 0.5,插入或缺失罚分值为 2,其他非沃森-克里克碱基配对罚分值为 1,总得分小于等于 3,且靶基因易亲近指数小于 25.0 kcal/mol 的序列为 miRNAs 的潜在靶基因^[10-11]。将筛选得到的 miRNAs 靶基因与 NCBI 蛋白质数据库进行 blastx 比对,将这些靶基因进行功能注释。

2 结果与分析

2.1 预测的小果野蕉 miRNAs 根据 miRNAs 在物种间具有高度的保守性及其前体的二级发夹结构特征,通过同源预测的方法,以一系列的标准进行筛选,最终在小果野蕉 EST(3 条)、GSS(6 条)和 HTGS(1 条)数据库中预测到 10 条保守的 miRNAs(表 1),它们属于 8 个 miRNAs 家族。其中 miR156 和 miR160 家族均预测到 2 个 miRNAs 成员,其他 6 个 miRNAs 家族只有 1 个成员。

表 1 小果野蕉中发现的 miRNAs 及其特征

miRNA	miRNA 成熟序列	ML	基因 ID	起始位置	终止位置	PL	Arm	miRBase best hit ID	E-value
mac-miR156a	TGACAGAAGAGACTGAGCAC	20	ES434836.1(EST)	114	211	98	6'	zma-MIR156d	2.00E-13
mac-miR156b	TTGACAGAAGATAGAGAGCAC	21	HN242063.1(GSS)	131	216	86	5'	pte-MIR156j	4.00E-22
mac-miR160a	TGCCTGGCTCCCTGTACGCCA	21	AC186747.2(HTGS)	77930	77841	90	5'	sbi-MIR160a	2.00E-17
mac-miR160b	TGCCTGGCTCCCTGTATGCCA	21	HN248011.1(GSS)	536	620	86	5'	osa-MIR160b	7.00E-23
mac-miR162a	TCGATAAACCTTGCATCCGG	21	HN247504.1(GSS)	814	682	133	3'	zma-MIR162	3.00E-14
mac-miR166a	TCGGACCAGGCTTCATTCCCC	21	FF558732.1(EST)	149	231	83	3'	bnm-MIR166d	1.00E-09
mac-miR168a	TCGCTTGGTGCAGGTCCGGAA	21	HN253213.1(GSS)	117	24	94	5'	sof-MIR168a	7.00E-13
mac-miR172a	GAATCTTGATGATGCTGCAC	20	HN239201.1(GSS)	119	2	118	3'	gma-MIR172a	5.00E-08
mac-miR396a	TTCCACAGCTTTCTTGAAGCTG	21	DX457510.1(GSS)	482	585	104	5'	lja-MIR396	7.00E-12
mac-miR2118a	TTGCCGATCCCTCCCATGCCGA	22	FF557264.1(EST)	63	221	159	3'	osa-MIR2118b	4.00E-08

注: ML: 成熟 miRNA 的长度; PL: miRNA 前体序列的长度; Arm: 成熟 miRNA 在前体中的位置

如表 1 所示,小果野蕉成熟 miRNAs 的长度为 20~22 nt,其中 7 条 miRNAs 成熟序列的长度为 21 nt,2 条 miRNAs 成熟序列的长度为 20 nt,1 条 miRNA 成熟序列的长度为 22 nt。同时,从表 1 中也能看出成熟 miRNA 序列在前体结构上的位置偏向于靠近 5'端,其中有 6 个成熟 miRNAs 位于茎环结构的 5'端,4 个位于 3'端。miRNAs 的前体序列平均长度约为 105 nt,变化范围为 83~159 nt。将 10 条 miRNAs 经 Mfold 进行二级结构折叠,详细的发夹结构见图 2,其中成熟的 miRNAs 序列用阴影显示。

2.2 小果野蕉 miRNAs 靶基因的预测 利用发现的 10 条小果野蕉 miRNAs,通过在线软件 psRNATarget 在小果野蕉 EST 数据库中进行靶基因预测,在 10 个 miRNAs 和 19 个 mRNA 之间预测了 24 对 miRNA:mRNA 互作对,其中大约 74%的靶基因有功能注释。miR160a 和 miR166a 的靶基因没有被预测到,完整的靶基因信息见表 2。小果野蕉中的这些靶基因涉及不同的新陈代谢和生理途径,如转移酶、转录调控因子 SPL、AP2 类调控因子、抗病性蛋白、突触融合蛋白和钙调结合蛋白等。

3 讨论

miRNAs 作为一种新型的非编码单链小分子 RNA,对真核生物的基因表达起快速有效的调控作用,它的发现使人们对基因表达调控有进一步的认识。目前,对 miRNAs 的研究集中在基因组已被解析的植物中,如拟南芥、水稻、毛果杨、玉米等^[12-15]。然而,在热带植物中,对 miRNAs 的研究比较滞后。本研究预测了小果野蕉 miRNAs 序列及其靶基因,为进一步研究小果野蕉 miRNAs 的功能和演变奠定了基础。

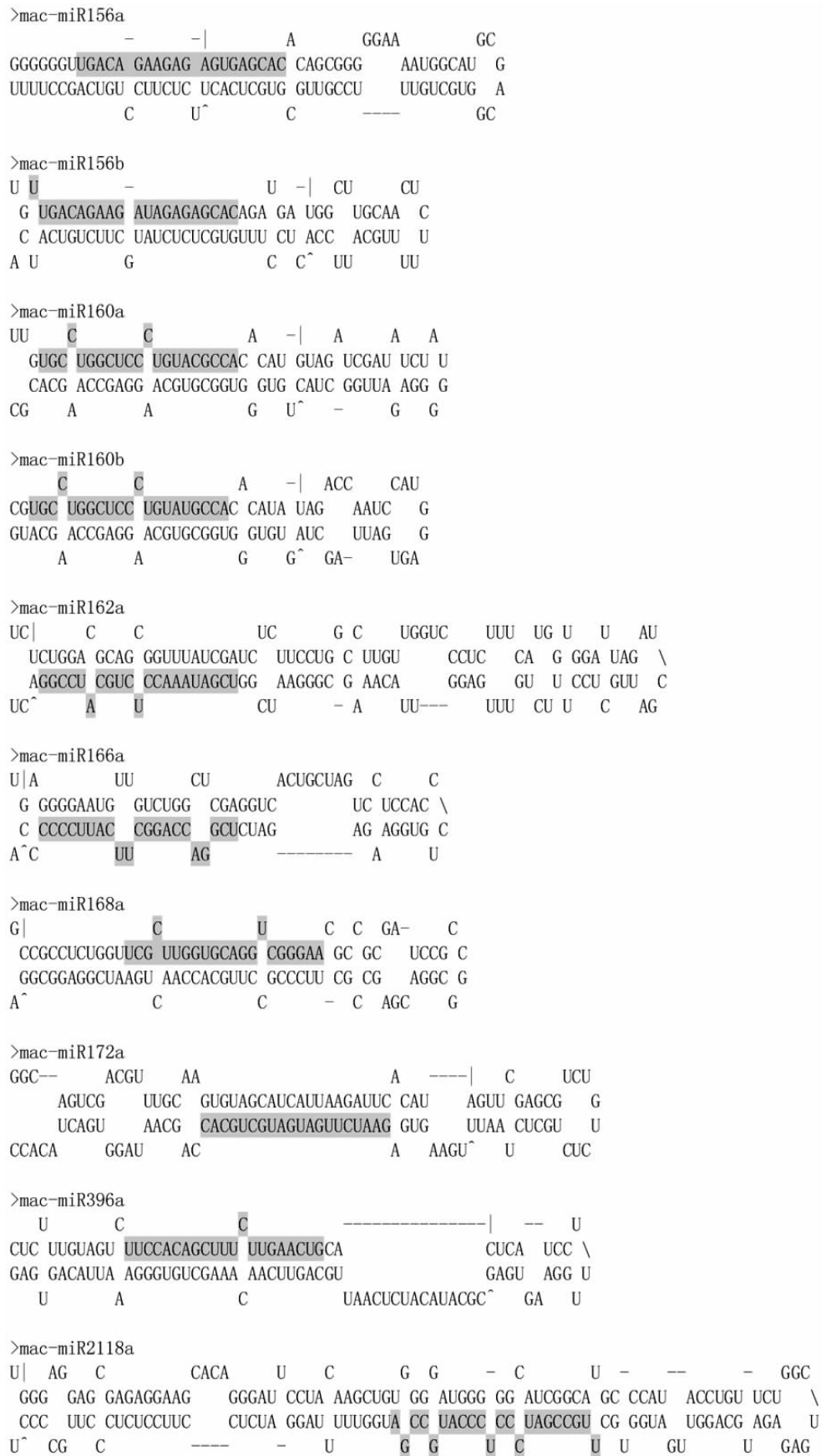


图 2 10 个 miRNAs 前体二级结构图

阴影部分表示成熟序列, 结构图由 Mfold 软件产生

表 2 小果野蕉中预测到的 miRNA 靶基因及其功能

miRNA_ID	靶基因总数	靶基因_ID	最大期望值	易接近值	靶基因功能
mac-miR156a	6	FF558536	1	12.616	SPL 类转录因子
		ES434709	1	12.682	SPL 类转录因子
		ES431356	1	16.633	SPL 类转录因子
		FF560486	1	22.572	SPL 类转录因子
		ES432808	2	11.517	SPL 类转录因子
		ES435047	2.5	15.624	功能未知
mac-miR156b	5	FF560486	1.5	23.555	SPL 类转录因子
		ES432808	2	11.231	SPL 类转录因子
		FF558536	2	12.65	SPL 类转录因子
		ES434709	2	12.845	SPL 类转录因子
		ES431356	2.5	16.332	SPL 类转录因子
mac-miR160a	0				
mac-miR160b	1	ES437336	3	20.303	功能未知
mac-miR162a	1	ES431955	3	15.047	二肽基肽酶
mac-miR166a	0				
mac-miR168a	1	FF560697	3	16.931	功能未知
mac-miR172a	4	ES436968	1.5	15.863	APETALA2 - 类蛋白
		FF559887	1.5	16.782	APETALA2 - 类蛋白
		FF562345	3	20.517	1 β - 半乳糖基转移酶
		FF560010	3	23.201	1 β - 半乳糖基转移酶
mac-miR396a	4	FF559506	3	18.501	功能未知
		FF560944	3	21.018	突触融合蛋白
		FF561310	3	23.73	转录调控蛋白 SNF2
		ES434228	3	24.765	功能未知
mac-miR2118a	2	ES436760	2	21.136	NBS-LRR 抗病性蛋白
		ES435616	2.5	14.812	钙调结合蛋白

利用生物信息学对 miRNAs 进行同源性预测是一种快速、有效和经济的方法。目前 miRBase 数据库储存了 291 条拟南芥 miRNAs 序列、581 条水稻 miRNAs 序列、234 条毛果杨 miRNAs 序列、362 条大豆 miRNAs 序列、163 条葡萄 miRNAs 序列、172 条玉米 miRNAs 序列和 63 条蓖麻 miRNAs 序列等,然而还没有储存小果野蕉 miRNAs 的相关信息。本研究在小果野蕉中发现了 8 个 miRNA 家族的 10 条序列,并且利用在线软件 psRNATarget 对 10 条 miRNAs 序列进行靶基因预测,获得 24 对 miRNA : mRNA 互作对,小果野蕉中的这些靶基因涉及不同的新陈代谢和生理途径。由于目前小果野蕉的基因组还未被解析,本研究不能够全面地预测到小果野蕉的 miRNAs 及其靶基因。随着小果野蕉基因组测序的进行,将有更多的小果野蕉 miRNAs 及其靶基因被发现,通过研究其对靶基因的调控功能,将进一步揭示小果野蕉的生长发育等机理。

参考文献:

- [1] MENG L, GAO X, CHEN J, et al. Spatial and temporal effects on seed dispersal and seed predation of *Musa acuminata* in southern Yunnan, China[J]. Integrative Zoology, 2012, 7(1): 30 - 40.
- [2] JONES-RHOADES M W, BARTEL D P, BARTEL B. MicroRNAs and their regulatory roles in plants[J]. Annual review of Plant Biology, 2006, 57: 19 - 53.
- [3] MALLORY A C, VAUCHERET H. MicroRNAs: something important between the genes[J]. Current Opinion in Plant biology, 2004, 7(2): 120 - 125.
- [4] BARTEL D P. MicroRNAs: genomics, biogenesis, mechanism, and function[J]. Cell, 2007, 131: 11 - 29.
- [5] BARTEL D P. MicroRNAs: genomics, biogenesis, mechanism, and function[J]. Cell, 2004, 116(2): 281 - 297.

- [6] LLAVE C , XIE Z , KASSCHAU K D , et al. Cleavage of *Scarecrow-like* mRNA targets directed by a class of *Arabidopsis* miRNA[J]. *Science* , 2002 , 297(5589) : 2053 – 2056.
- [7] GRIFFITHS-JONES S , GROCOCK R J , VAN DONGEN S , et al. miRBase: microRNA sequences , targets and gene nomenclature[J]. *Nucleic Acids Research* , 2006 , 34: 140 – 144.
- [8] ZUKER M. Mfold web server for nucleic acid folding and hybridization prediction [J]. *Nucleic acids research* 2003 31(13) : 3406 – 3415.
- [9] ZENG C , WANG W , ZHENG Y , et al. Conservation and divergence of microRNAs and their functions in Euphorbiaceous plants[J]. *Nucleic Acids Research* 2010 , 38(3) : 981 – 995.
- [10] DAI X , ZHAO P X. psRNATarget: a plant small RNA target analysis server[J]. *Nucleic Acids Research* , 2011 , 39: 155 – 159.
- [11] ZHAO C Z , XIA H , FRAZIER T P , et al. Deep sequencing identifies novel and conserved microRNAs in peanuts (*Arachis hypogaea* L.) [J]. *BMC Plant Biology* , 2010 , 10: 3.
- [12] LI B S , YIN W L , XIA X L. Identification of microRNAs and their targets from *Populus euphratica* [J]. *Biochemical and Biophysical Research Communications* , 2009 , 388(2) : 272 – 277.
- [13] ARCHAK S , NAGARAJU J. Computational prediction of rice (*Oryza sativa*) miRNA targets[J]. *Genomics , Proteomics & Bioinformatics* 2007 5(3/4) : 196 – 206.
- [14] LU S , SUN YH , SHI R , et al. Novel and mechanical stress-responsive MicroRNAs in *Populus trichocarpa* that are absent from *Arabidopsis* [J]. *The Plant Cell* , 2005 , 17(8) : 2186 – 2203.
- [15] ZHANG L , CHIA JM , KUMARI S , et al. A genome-wide characterization of microRNA genes in maize[J]. *PLoS Genetics* , 2009 , 5(11) : e1000716.

Bioinformatic Prediction of Conserved microRNAs and Their Target Genes in *Musa acuminata*

YE Ke-yong¹ , CHEN Yao² , LI Rui-mei¹ , FU Shao-ping¹ , GUO Jian-chun¹

(1. Institute of Tropical Bioscience and Biotechnology , Chinese Academy of Tropical Agricultural Sciences , Haikou 571101 , China;

2. Scientific and Technological Information Research Institute , Chinese Academy of Tropical Agricultural Sciences , Haikou 571101 , China)

Abstract: MicroRNAs (miRNAs) are small RNAs (sRNA) that regulate post-transcriptional gene expression by cleaving or inhibiting the translation of target gene transcripts. Based on the conservation of miRNAs in many plant species and the currently available genome sequence , 10 individual conserved miRNAs have been predicted in *Musa acuminata* , and these miRNAs were found to belong to 8 families. Online psRNATagret software was run and 24 pairs of miRNAs were predicted to interact with their targets in *Musa acuminata* , and these target genes are involved in metabolism , development , response to biotic and abiotic stresses , and other cellular processes. This study has provided information on *Musa acuminata* miRNAs and their targets for future studies.

Key words: *Musa acuminata*; microRNAs (miRNAs); target gene; bioinformatics