

文章编号: 1674-7054(2016)02-0237-09

## 香蕉 miRNA 及其靶基因的生物信息学预测

彭 赟, 龙海飞, 贺 瑞, 徐 敏, 曾灿彬, 骆 娟, 齐 钊,  
闫 臻, 阮云泽, 何朝族, 汤 华

(海南大学 农学院/海南省热带生物资源可持续利用重点实验室, 海口, 570228)

**摘 要:** 利用生物信息学方法预测香蕉中的 miRNA, 即将 miRBase 数据库中的已知植物的 miRNA 与香蕉 EST 和 GSS 数据库进行 BLAST 比对搜索, 筛选出潜在的 miRNA, 最后得到 16 条香蕉 miRNA, 分属于 9 个 miRNA 家族, 其中 3 条来源于 EST 数据库, 13 条来源于 GSS 数据库, 并利用在线软件 psRNATarget 预测其靶基因, 共得到 168 个靶基因, 分别编码多种功能蛋白, 参与各生物学过程。

**关键词:** 香蕉; microRNA; 靶基因; 生物信息学分析

中图分类号: Q 789

文献标志码: A

DOI: 10.15886/j.cnki.rdsxb.2016.02.017

microRNA(miRNA)是一类在真核生物中发现的非编码、内源性的单链小分子 RNA, 长度大约为 21 个核苷酸, 它可以通过对靶基因的切割或者抑制翻译来发挥转录后水平的负调控作用<sup>[1]</sup>。在植物中, 成熟 miRNA 不是基因直接转录的产物, 而是由 1 kb 左右的初始转录产物 pri-miRNA 经 Dicer-like1 酶剪切产生具有茎环结构的 pre-miRNA, 最终形成成熟 miRNA<sup>[2]</sup>。miRNA 与靶基因的转录产物互补结合从而进行调控, 如果完全互补, miRNA 就切割 mRNA, 如果不完全互补, 则引起翻译抑制。作为重要的调控分子, miRNA 参与植物生命过程中一系列重要进程, 主要包括个体生长发育的调控、细胞、组织分化、激素分泌、信号转导以及抗逆反应等<sup>[3]</sup>。目前, miRNA 的鉴定方法主要有 2 种: 分子生物学和生物信息学<sup>[4]</sup>。其中分子生物学主要包括直接克隆<sup>[5]</sup>和高通量测序<sup>[6]</sup>等方法。直接克隆法是通过从总 RNA 中富集片段大小在 21nt 左右的小 RNA 分子来构建不同组织、不同发育时期或不同生长条件下的 miRNA 基因库, 但这种方法很难鉴定表达丰度低或者特异性时空表达的 miRNA<sup>[7]</sup>。随着高通量测序技术的发展, 为 miRNA 鉴定和定量分析提供了有力的技术支持, 具有灵活性强、灵敏度高等特点, 然而, 还是存在周期长、成本较高等问题。与上述 2 种方法相比, 利用生物信息学预测 miRNA 成为当前研究 miRNA 的重要策略之一。通过比较基因组学研究发现, 从低等的苔藓、裸子植物到高等的单子叶、双子叶植物, 大量已知的 miRNA 在植物中具有高度的保守性<sup>[8]</sup>。根据 miRNA 的这一特点, 可以用同源搜索的方法在其他物种中预测潜在的 miRNA。这是一种相对方便、经济和快速的方法。近年来, 利用 EST 序列以及 GSS 序列, 通过同源性搜索方法在植物物种中发现了大量的保守 miRNA, 包括番茄<sup>[9]</sup>、棉花<sup>[10]</sup>、玉米<sup>[11]</sup>等。香蕉是我国最大宗的热带水果, 产区主要在广东、广西、海南等省份, 其营养价值高, 市场需求量大, 是我国重要的经济作物。目前对于香蕉 miRNA 的研究还处于起始阶段。寻找香蕉 miRNA 可以为香蕉的生长发育以及抗逆反应提供更多生物学信息。笔者利用植物中 miRNA 序列和结构的保守性, 依据 miRBase 上已知的多种植物的 miRNA 序列信息, 与香蕉的 EST、GSS 数据库比较, 预测香蕉的 miRNA 及其靶基因, 对 miRNA 的序列特征

收稿日期: 2015-11-25

基金项目: 国家自然科学基金资助项目(31360364); 教育部热带作物新品种选育工程中心与作物学重点学科联合资助项目(lhxm-2012-2); 海南省重大科技项目(ZDX2013023)及“中央财政支持中西部高校提升综合实力专项”子课题

作者简介: 彭赟(1991-), 女, 海南大学农学院 2013 级硕士研究生, E-mail: pyrt107@sina.com

通信作者: 汤华(1974-), 男, 教授, 博士, 研究方向: 作物遗传育种与分子生物学, E-mail: thtiger@163.com

和其靶基因的功能等方面进行深入分析,对研究香蕉 miRNA 的调控机理具有重要意义。

## 1 材料与方法

**1.1 相关序列信息数据库** 所有的成熟 miRNA 和 miRNA 前体序列均来源于小 RNA 数据库 miRBase database(<http://microrna.sanger.ac.uk/sequences/index.shtml>)。香蕉的 EST(Expressed Sequence Tag)和 GSS(Genomic Survey Sequence)序列来源于美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI)的 GenBank 核酸序列数据库网站(<http://banana-genome.cirad.fr/home>)。蛋白质序列来源于 NCBI 蛋白质数据库(<http://www.ncbi.nlm.nih.gov/>)。靶基因预测的数据库为香蕉基因组数据库([http://banana-genome.cirad.fr/download/musa\\_cds.fna.gz](http://banana-genome.cirad.fr/download/musa_cds.fna.gz))。

**1.2 生物信息学分析相关软件** 核酸和蛋白质序列比对软件分别是 BLASTN 和 BLASTX(<http://www.ncbi.nlm.nih.gov/BLAST/>)。多序列比对分析所用软件是 ClustalW(<http://www.ebi.ac.uk/Tools/msa/clustalw2/>)。运用 RNA 在线折叠软件 Mfold 3.2(<http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form>)。分析比对得到的序列,判断是否具有 miRNA 前体的基本结构特征。靶基因预测的在线软件为 psRNATarget(<http://plantgrn.noble.org/psRNATarget/>)。

**1.3 香蕉 miRNAs 的生物信息学预测** miRNAs 计算机预测基本流程:生物信息学预测 miRNA 的一般方法是从 miRBase 数据库下载收录的已知植物 miRNA 序列,去除其中的重复序列,以此为探针搜索目标植物的 EST 和 GSS 同源序列。选取低于 4 个碱基错配的同源序列,经比对去除冗余序列及蛋白质编码序列后,作为候选 miRNA 序列。最后,用 Mfold 3.2 软件预测候选 miRNA 序列的 2 级结构,并分析其稳定性<sup>[12]</sup>。miRNA 的预测标准:Bonnet 等<sup>[13]</sup>提出了预测 miRNA 及分析其 2 级结构稳定性的标准为:(1)预测得到的 miRNA 与已知 miRNA 序列碱基错配数小于 4;(2)miRNA 的前体能折叠形成发夹型 2 级结构;(3)miRNA 前体折叠自由能最小(Negative minimal free energies, MFEs),折叠自由能系数最小(Minimal free energy index, MFEIs);(4)miRNA 中(A+U)含量在 30%~70%左右;(5)miRNA 互补区 2 条单链之间只存在小于 6 的错配碱基;(6)miRNA 互补区 2 条单链不能存在额外环或缺口。

**1.4 香蕉 miRNA 靶基因的预测** miRNA 主要通过和靶基因完全互补或者不完全互补的方式识别引起靶基因的降解或者翻译抑制<sup>[1]</sup>。利用这一机理,可以通过同源性搜索的方法,在植物中预测 miRNA 的靶基因。目前,通过实验手段已经验证了大部分预测得到的靶基因。Zhang B H 等<sup>[14]</sup>利用在线软件 miRU(<http://bioinfo3.noble.org/miRU.htm>)在各种植物中预测 miRNA 靶基因,取得了较好的结果。本研究利用植物在线软件 psRNATarget(<http://plantgrn.noble.org/psRNATarget/>)在香蕉基因组数据库([http://banana-genome.cirad.fr/download/musa\\_cds.fna.gz](http://banana-genome.cirad.fr/download/musa_cds.fna.gz))中预测香蕉 miRNA 的靶基因。成熟的 miRNA 与其靶基因 mRNA 之间只允许不超过 4 个错配碱基,并且不允许存在缺口,“miRNA-靶基因”配对的最低自由能要低于  $-28.2 \text{ kcal} \cdot \text{mol}^{-1}$ 。

## 2 结果与分析

**2.1 植物 miRNA 的保守性分析** 大部分植物 miRNA 存在着序列和结构保守性。从 miRBase database (Release 10.1)中下载已收录的植物 miRNA 及其前体序列,这些 miRNA 序列来源于拟南芥、水稻、玉米、烟草、大豆等多种植物。筛选笔者下载的植物 miRNA,得到存在于 2 个物种以上的共 42 个 miRNA 家族(表 1)。其中,有 8 个 miRNA 家族保守性较高,存在于 10 个以上的物种中。另外,有 12 个 miRNA 家族存在于 6~10 个物种中。剩余的 22 个 miRNA 家族存在于 2~5 个物种中。

**2.2 改进植物 miRNA 的预测方法** 植物成熟 miRNA 序列具有高度的进化保守性,所以可以利用这种保守性在其他植物的相关核酸数据库中采用同源搜索法预测得到新的 miRNA。传统的同源搜索法是利用拟南芥、水稻或者其他 miRBase 数据库中已知植物的 miRNA 作为“word sequence”搜索新的 miRNA。工作量较大,重复性较高,且搜索范围大部分局限在全基因组或表达序列标签 EST 数据库。在本研究改进的植物 miRNA 计算机预测方法中,笔者把 42 个保守 miRNA 家族的序列作为“word sequence”,搜索范

围包括了 EST 数据库和 GSS 数据库。在完成 BLAST 程序后,将碱基错配数低于 4 个的序列作为运行 BLASTX 的候选序列。

2.3 预测的香蕉潜在 miRNA 按照上述步骤,经过一系列的筛选以及 2 级结构折叠,在严格控制参数的情况下得到 16 条潜在的香蕉 miRNA(表 2),它们分属于 9 个家族,其中 miR156,miR396 家族各有 2 个成员,miR535 家族有 6 个成员,其余家族只有 1 名成员。预测的 miRNA 来源分别为 3 条 miR156a,miR156b,miR166 来源于 EST 数据库,其余 13 条 miRNA 来源于 GSS 数据库。

表 1 miRBase 数据库( release 10.1) 中植物的保守 miRNA 家族  
Tab.1 The conserved miRNA families of plants in miRBase database ( release 10.1)

| 植物物种数目<br>Plant species | 保守的 miRNA 家族<br>Conserved miRNA's family                                                                                                                     |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 11 ~ 17                 | miR156/157 miR159 miR160 miR166 miR167 miR171 miR319 miR396                                                                                                  |
| 6 ~ 10                  | miR162 miR164 miR168 miR169 miR172 miR390 miR393 miR394 miR395 miR398 miR399 miR408                                                                          |
| 2 ~ 5                   | miR161 miR397 miR403 miR413 miR414 miR415 miR416 miR417 miR418 miR419 miR420 miR426<br>miR444 miR472 miR477 miR479 miR482 miR529 miR535 miR536 miR783 miR827 |

表 2 新预测的香蕉 miRNA 家族及其成员特征  
Tab. 2 The predicted banana miRNA families and numbers

| miRNA   | 基因编号<br>Gene ID | 成熟序列<br>Mature seq    | 位点 site | NM | LM | LP  | ( A + U ) / % | MFE   | MFEI |
|---------|-----------------|-----------------------|---------|----|----|-----|---------------|-------|------|
| miR156a | ES434836        | UGACAGAAGAGAGUGAGCAC  | EST5'   | 0  | 20 | 84  | 45.24         | 47.90 | 1.04 |
| miR156b | JK538379        | UGACAGAAGAUAGAGAGCAC  | EST5'   | 2  | 20 | 80  | 52.50         | 45.10 | 1.19 |
| miR166  | FF558732        | UCGGACCAGGCUUCAUCCCC  | EST3'   | 0  | 21 | 77  | 41.56         | 41.40 | 0.92 |
| miR160  | HN248011        | UGCCUGGCUCCUGUAUGCCA  | GSS5'   | 0  | 21 | 82  | 46.34         | 41.40 | 0.94 |
| miR168  | HN253213        | UCGCUUGGUGCAGGUCGGGAA | GSS3'   | 0  | 21 | 70  | 30.00         | 40.20 | 0.82 |
| miR396a | DX457510        | UUCCACAGCUUUCUUGAACUG | GSS5'   | 0  | 21 | 80  | 55.00         | 35.30 | 0.98 |
| miR396b | HN260136        | UUCCACAGCUUUCUUGAACU  | GSS3'   | 1  | 20 | 66  | 60.61         | 26.10 | 1.00 |
| miR162  | HN247504        | UCGAUAAACCGCUGCGUCCAG | GSS3'   | 2  | 21 | 127 | 49.61         | 50.90 | 0.80 |
| miR164  | HN246671        | UGGAGAAGCAGGGCACGUGCA | GSS5'   | 0  | 21 | 74  | 51.35         | 32.60 | 0.91 |
| miR399  | HN253203        | UGCCAAAGGAGACUUGCCCU  | GSS5'   | 2  | 20 | 110 | 47.27         | 54.80 | 0.94 |
| miR535a | HN242296        | UGACAACGAGAGAGACACGC  | GSS5'   | 0  | 21 | 82  | 40.24         | 51.10 | 1.04 |
| miR535b | HN242296        | UGACGACGAGAAAGAGCACGC | GSS3'   | 2  | 21 | 82  | 40.24         | 51.10 | 1.04 |
| miR535c | HN241873        | UGACAACGAGAGAGACACGC  | GSS5'   | 0  | 21 | 80  | 50.00         | 41.80 | 1.05 |
| miR535d | HN241873        | UGACAACGAUAGAAAGCACGC | GSS3'   | 2  | 21 | 80  | 50.00         | 41.80 | 1.05 |
| miR535e | HN260441        | UGACAACGAGAAAGAGCACGC | GSS5'   | 1  | 21 | 82  | 39.02         | 48.40 | 0.97 |
| miR535f | HN260441        | UGACGACGAGAGAGACACGC  | GSS3'   | 1  | 21 | 82  | 39.02         | 48.40 | 0.97 |

注: NM 为错配碱基数; LM 为成熟 miRNA 的长度; LP 为前体长度; ( A + U ) % 为前体的 A + U 碱基含量; MFE 为最小折叠自由能; MFEI 为最小折叠自由能指数

Note: NM: number of mispairing; LM: length of mature miRNA; LP: length of precursor; ( A + U ) % : the content of precursor; MFE: minimal folding free energies; MFEI: minimal folding free energies index

2.4 香蕉 miRNAs 的特性分析 本研究在香蕉中预测得到 16 条成熟 miRNA 长度为 20 ~ 21 nt,符合 miRNA 的长度要求。而 miRNA 前体则表现出多样性,本实验预测的 16 条 miRNA 前体中,有 15 条长度在 70 ~ 130 nt 之间,这与在棉花、玉米、烟草以及大豆等植物中的报道一致;仅 1 条 miR396b 的前体长度为 66 nt。由图 1 可见,虽然 miRNA 前体的长度和 2 级结构表现出了一定的多样性,但都能折叠形成稳定的

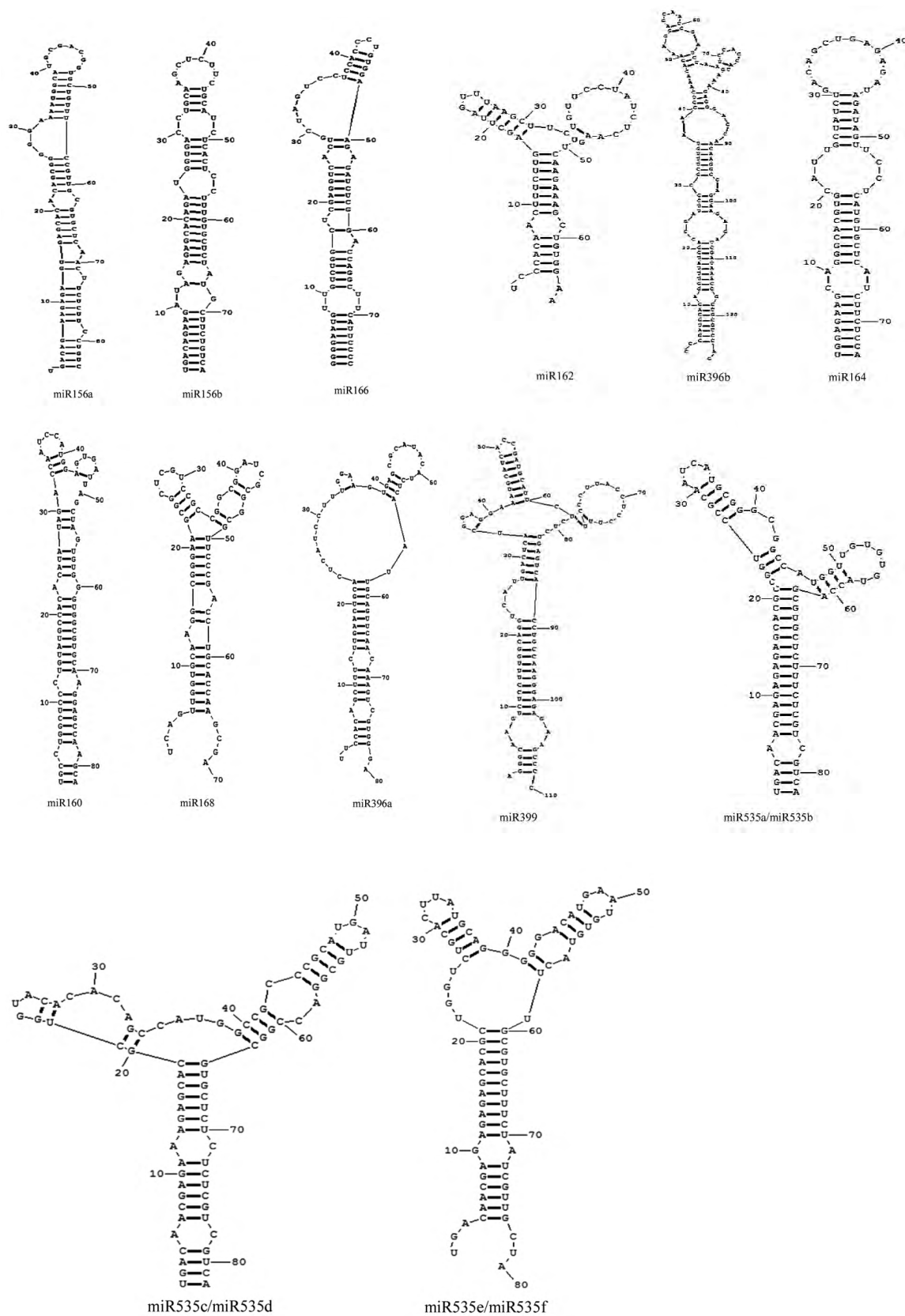


图 1 预测的 16 条香蕉 miRNA 前体的 2 级结构图  
Fig. 1 The secondary structures of predicted 16 banana miRNA precursors

2 级茎环结构。9 条 miRNA 位于其 miRNA 前体序列的 5' 端, 7 条位于其前体序列的 3' 端, 可见成熟 miRNA 出现在 2 条臂上的概率是相同的。在研究过程中, 发现一个有趣的现象: miRNA535a 和 miRNA535b, miRNA535c 和 miRNA535d, 以及 miRNA535e 和 miRNA535f 这 3 对序列的位置出现在 2 级结构对应的臂上, 且是严格配对的。这是由于其 miRNA 的序列特性决定的, 在允许出现少数错配的情况下, 2 条序列都能与“word sequence”类比匹配。后续靶基因的研究表明, 2 条序列均能找到相对应的靶基因, 且靶基因功能和性质是不同的。

研究结果表明, MFE 和 MFEI 是区分 miRNA 与其他小分子 RNA 的重要参数<sup>[15]</sup>。本研究预测得到的香蕉 miRNA 前体均有较高的 MFE 值和 MFEI 值(表 2), miRNA 前体的 MFE 为 26.10 ~ 54.80 kcal · mol<sup>-1</sup>, 平均值为 43.64 kcal · mol<sup>-1</sup>; MFEI 为 0.80 ~ 1.19, 平均值为 0.98, 显著高于其他 RNA 分子的 MFEI, 例如 tRNA (0.64), rRNA (0.59), mRNA (0.65), 符合先前报道的 miRNA 前体的 MFEI 平均值在 1.0 左右<sup>[15]</sup>。Zhang B H 等<sup>[15]</sup>报道了 miRNA 前体的 (A + U) 含量高于 (G + C) 含量, 能够使 miRNA 前体更加稳定。然而本研究发现的 miRNA 前体的 (A + U) 含量在 30.00% ~ 60.61% 之间, 平均值为 46.13%, 并不显著高于 (G + C) 含量, 但仍在报道的 30% ~ 70% 范围内, 因此是符合要求的。

**2.5 香蕉 miRNAs 靶基因的预测** 以新发现的 miRNA 序列作为探针, 用植物小分子 RNA 靶基因搜索软件 psRNATarget 从香蕉基因组数据库 ([http://banana-genome.cirad.fr/download/musa\\_cds.fna.gz](http://banana-genome.cirad.fr/download/musa_cds.fna.gz)) 中找到 16 条 miRNA 的靶基因(表 3)。本实验共发现 16 个 miRNA 的 136 个靶基因, 每个 miRNA 均对应有不同数目的靶基因。其中 miR156a 和 miR156b 的靶基因数目最多为 24 个; 其次, miR396a, miR396b 和 miR164 的靶基因数目分别为 14, 19, 18 个; 靶基因数目最少的是 miR535e 只有 1 个靶基因。尽管其中一部分基因编码的是功能未知的蛋白, 但部分靶基因编码蛋白是清楚的, 这些靶基因编码的蛋白多种多样, 如生长调节因子、启动子结合蛋白、叶绿体线粒体中的生化酶、疾病抗性蛋白等, 参与生长发育、形态构建、生理生化、病理反应等过程。

### 3 讨 论

miRNA 是一种新发现的非编码的基因表达调控因子, 参与真核生物的基因表达调控, 并在调控中起重要作用, 它的发现丰富了人们对基因表达调控的认识。对于 miRNA 的发现与研究主要集中在大豆、玉米、棉花和烟草等经济作物中, 其他物种中 miRNA 的研究相对较少。香蕉作为一种重要的经济作物, 其 miRNA 的研究比较滞后。本研究预测了 miRNA 的序列及潜在的靶基因, 为进一步研究香蕉 miRNA 的调控机理奠定了基础。

植物 miRNA 靶基因具有倾向于靶定转录因子的特点<sup>[15]</sup>。本研究预测得到的香蕉 miRNA 靶基因中, 有一部分负责编码转录因子。这些转录因子对香蕉生长发育和其他的生理代谢过程具有重要调控作用。比如本研究预测到 NAC 转录因子家族是香蕉 miR164 的靶基因, 在野生拟南芥中 miR164 的过量表达会导致花组织的融合和子叶的融合, 这说明由 miR164 调控造成 NAC 转录因子的降解会影响植物花的正常发育。除了靶定转录因子外, 香蕉 miRNA 的靶基因还编码一些参与新陈代谢、胁迫反应等的蛋白。比如香蕉 miR535a 的靶基因编码钾离子运输蛋白。

Tab. 3 The target genes and their functions of potential miRNAs in banana

| miRNA   | 靶基因位点<br>The site of target genes                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | 靶蛋白<br>Target protein                                                                                                                                                                                                                                                                    |
|---------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| miR156a | GSMUA_Achr11T19580_001、GSMUA_Achr6T22280_001、<br>GSMUA_Achr5T22630_001、GSMUA_Achr2T06150_001、<br>GSMUA_Achr3T11170_001、GSMUA_Achr10T02970_001、<br>GSMUA_Achr10T23280_001、GSMUA_Achr3T27750_001、<br>GSMUA_Achr9T23140_001、GSMUA_Achr3T05360_001、<br>GSMUA_Achr9T14160_001、GSMUA_Achr9T27660_001、<br>GSMUA_Achr6T36010_001、GSMUA_Achr5T21990_001、<br>GSMUA_Achr5T25090_001、GSMUA_AchrUn_randomT08380_001、<br>GSMUA_AchrUn_randomT21900_001、<br>GSMUA_AchrUn_randomT05850_001<br>GSMUA_Achr9T29330_001、GSMUA_Achr9T01380_001<br>GSMUA_Achr8T23820_001<br>GSMUA_Achr6T06030_001<br>GSMUA_Achr4T12470_001<br>GSMUA_Achr4T01450_001 | Squamosa promoter-binding-like protein<br><br><br><br><br><br><br><br><br><br>OsSPL16-SBP-box gene family member<br>Hypothetical protein<br>uncharacterized protein<br>uncharacterized protein Sb04g004940<br>Whole genome shotgun sequence of line<br>PN40024 , scaffold_7. assembly12x |
| miR156b | GSMUA_Achr3T11170_001、GSMUA_Achr10T02970_001、<br>GSMUA_Achr10T23280_001、GSMUA_Achr3T27750_001、<br>GSMUA_Achr3T05360_001、GSMUA_Achr9T27660_001、<br>GSMUA_Achr9T14160_001、GSMUA_Achr5T21990_001、<br>GSMUA_Achr9T23140_001、GSMUA_Achr4T05720_001、<br>GSMUA_Achr5T25090_001、GSMUA_Achr2T06150_001、<br>GSMUA_Achr6T22280_001、GSMUA_Achr6T36010_001、<br>GSMUA_AchrUn_randomT08380_001、<br>GSMUA_AchrUn_randomT21900_001、<br>GSMUA_AchrUn_randomT05850_001<br>GSMUA_Achr9T29330_001、GSMUA_Achr9T01380_001<br>GSMUA_Achr8T23820_001、GSMUA_Achr4T26940_001<br>GSMUA_Achr6T06030_001<br>GSMUA_Achr4T12470_001<br>GSMUA_Achr4T01450_001  | Squamosa promoter-binding-like protein<br><br><br><br><br><br><br><br><br><br>OsSPL16-SBP-box gene family member<br>Hypothetical protein<br>uncharacterized protein<br>uncharacterized protein Sb04g004940<br>Whole genome shotgun sequence of line<br>PN40024 , scaffold_7. assembly12x |
| miR166  | GSMUA_Achr8T02960_001、GSMUA_Achr4T31660_001、<br>GSMUA_Achr2T13050_001、GSMUA_Achr7T01060_001、<br>GSMUA_Achr5T02420_001<br>GSMUA_Achr5T18530_001<br>GSMUA_Achr8T20090_001<br>GSMUA_Achr8T02070_001                                                                                                                                                                                                                                                                                                                                                                                                                              | Homeobox-leucine zipper protein HOX32<br><br><br>Homeobox-leucine zipper protein HOX9<br>Homeobox-leucine zipper protein ATHB-15<br>2-oxoglutarate dehydrogenase , mitochondrial                                                                                                         |

续表 3 Continued Tab 3

| miRNA   | 靶基因位点<br>The site of target genes                                                                                                                                                                                                                                                                                                                                                                    | 靶蛋白<br>Target protein                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|---------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| miR160  | GSMUA_Achr6T18900_001、GSMUA_Achr5T03960_001、<br>GSMUA_Achr9T29480_001、GSMUA_Achr8T18930_001、<br>GSMUA_Achr5T18540_001、GSMUA_Achr4T18240_001、<br>GSMUA_Achr5T14630_001、GSMUA_Achr10T04600_001                                                                                                                                                                                                         | Auxin response factor                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| miR168  | GSMUA_Achr1T07120_001、GSMUA_Achr3T27070_001、<br>GSMUA_Achr1T17950_001、GSMUA_Achr3T13970_001<br><br>GSMUA_Achr9T00190_001<br><br>GSMUA_Achr11T14030_001                                                                                                                                                                                                                                               | Protein argonaute<br><br>Receptor protein kinase CLAVATA1<br>Stachyose synthase                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| miR396a | GSMUA_Achr7T00440_001<br><br>GSMUA_Achr10T18020_001<br><br>GSMUA_Achr4T09680_001<br><br>GSMUA_Achr1T21940_001<br><br>GSMUA_Achr3T22790_001<br><br>GSMUA_Achr8T30640_001<br><br>GSMUA_Achr10T06370_001、GSMUA_Achr1T23840_001、<br>GSMUA_Achr2T04650_001、GSMUA_Achr6T01270_001、<br>GSMUA_AchrUn_randomT04320_001<br><br>GSMUA_Achr6T24050_001<br><br>GSMUA_Achr8T07650_001<br><br>GSMUA_Achr4T25770_001 | DNA ( cytosine-5) -methyltransferase 1B<br>Putative Protein sel-1 homolog 2<br>CAAX amino terminal protease family protein<br>Whole genome shotgun sequence of line<br>PN40024 , scaffold_16. assembly12x<br>Whole genome shotgun sequence of line<br>PN40024 , scaffold_7. assembly12x ( Frag-<br>ment)<br>Whole genome shotgun sequence of line<br>PN40024 , scaffold_2. assembly12x<br>growth-regulating factor<br><br>Putative UPF0553 protein<br>uncharacterized protein<br>heavy metal-associated domain containing<br>protein |
| miR396b | GSMUA_Achr7T00440_001<br><br>GSMUA_Achr10T18020_001<br><br>GSMUA_Achr6T24050_001<br><br>GSMUA_Achr1T21940_001<br><br>GSMUA_Achr3T22790_001<br><br>GSMUA_Achr8T30640_001<br><br>GSMUA_Achr4T25770_001<br><br>GSMUA_Achr8T07650_001<br><br>GSMUA_Achr4T17680_001<br><br>GSMUA_AchrUn_randomT15490_001                                                                                                  | DNA ( cytosine-5) -methyltransferase 1B<br>Putative Protein sel-1 homolog 2<br>Putative UPF0553 protein<br>Whole genome shotgun sequence of line<br>PN40024 , scaffold_16. assembly12x<br>Whole genome shotgun sequence of line<br>PN40024 , scaffold_7. assembly12x<br>Whole genome shotgun sequence of line<br>PN40024 , scaffold_2. assembly12x<br>heavy metal-associated domain containing<br>protein<br>uncharacterized protein<br>nodulin<br>Predicted protein                                                                 |

续表 3 Continued Tab 3

| miRNA   | 靶基因位点<br>The site of target genes                                                                                                                                                                        | 靶蛋白<br>Target protein                                    |
|---------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------|
|         | GSMUA_Achr3T15140_001                                                                                                                                                                                    | growth regulating factor protein                         |
|         | GSMUA_Achr1T23840_001、GSMUA_Achr8T04970_001、<br>GSMUA_Achr1T27000_001、GSMUA_Achr2T04650_001、<br>GSMUA_Achr6T01270_001、GSMUA_Achr10T06370_001、<br>GSMUA_Achr7T27480_001、<br>GSMUA_AchrUn_randomT04320_001 | growth-regulating factor                                 |
| miR162  | GSMUA_Achr9T07690_001                                                                                                                                                                                    | Anthranilate synthase component I-4 , chlo-<br>roplastic |
| miR535a | GSMUA_Achr10T28970_001                                                                                                                                                                                   | Hypothetical protein                                     |
|         | GSMUA_Achr3T06140_001                                                                                                                                                                                    | Potassium transporter 6                                  |
| miR535b | GSMUA_Achr2T03720_001                                                                                                                                                                                    | Extended synaptotagmin-4                                 |
|         | GSMUA_Achr10T28970_001                                                                                                                                                                                   | Hypothetical protein                                     |
| miR535c | GSMUA_Achr10T28970_001                                                                                                                                                                                   | Hypothetical protein                                     |
|         | GSMUA_Achr3T06140_001                                                                                                                                                                                    | Potassium transporter 6                                  |
| miR535d | GSMUA_Achr10T28970_001                                                                                                                                                                                   | Hypothetical protein                                     |
|         | GSMUA_Achr11T06120_001                                                                                                                                                                                   | expressed protein                                        |
| miR535e | GSMUA_Achr10T28970_001                                                                                                                                                                                   | Hypothetical protein                                     |
| miR535f | GSMUA_Achr10T28970_001                                                                                                                                                                                   | Hypothetical protein                                     |
|         | GSMUA_Achr10T03540_001                                                                                                                                                                                   | Sodium/hydrogen exchanger                                |
|         | GSMUA_Achr8T22130_001                                                                                                                                                                                    | uncharacterized protein                                  |
| miR164  | GSMUA_Achr10T22350_001                                                                                                                                                                                   | Protein CUP-SHAPED COTYLEDON 2                           |
|         | GSMUA_Achr9T12130_001                                                                                                                                                                                    | disease resistance protein                               |
|         | GSMUA_Achr6T31350_001、GSMUA_Achr9T10210_001、<br>GSMUA_Achr6T30050_001、GSMUA_Achr5T26640_001、<br>GSMUA_Achr3T23360_001、GSMUA_Achr11T25720_001                                                             | NAC domain-containing protein                            |
|         | GSMUA_Achr9T27530_001、GSMUA_Achr5T00500_001、<br>GSMUA_Achr6T30570_001、GSMUA_Achr7T11500_001                                                                                                              | no apical meristem protein                               |
|         | GSMUA_Achr6T28900_001                                                                                                                                                                                    | GDSL esterase/lipase At1g33811                           |
|         | GSMUA_Achr10T26180_001                                                                                                                                                                                   | NO APICAL MERISTEM                                       |
|         | GSMUA_Achr9T22210_001                                                                                                                                                                                    | E3 ubiquitin-protein ligase At4g11680                    |
|         | GSMUA_Achr9T06850_001                                                                                                                                                                                    | expressed protein                                        |
|         | GSMUA_Achr7T15220_001                                                                                                                                                                                    | UDP-glucuronic acid decarboxylase 1                      |
|         | GSMUA_Achr4T26960_001                                                                                                                                                                                    | E3 ubiquitin-protein ligase BRE1-like 1                  |
| miR399  | GSMUA_Achr9T09130_001                                                                                                                                                                                    | Uncharacterized protein sll1770                          |
|         | GSMUA_Achr4T03580_001                                                                                                                                                                                    | 2-oxoglutarate/malate translocator , chloro-<br>plastic  |

## 参考文献:

- [1] Bartel D P. MicroRNAs: Genomics , biogenesis , mechanism and function [J]. Cell , 2004 , 116( 2) : 281 – 297.
- [2] Jung J H , Seo P J , Park C M. MicroRNA biogenesis and function in higher plants [J]. Plant Biotechnol Rep 2009 , 3( 2) : 111 – 126.
- [3] Chen X M. MicroRNA biogenesis and function in plants [J]. FEBS Lett , 2005 , 579( 26) : 5923 – 5931.
- [4] Taylor P F , Zhang B H. Identification of plant microRNAs using expressed sequence tag analysis [J]. Methods Mol Biol , 2011 , 678: 13 – 25.
- [5] Baker C C , Sieber P , Wellmer F , et al. The early extra petals1 mutant uncovers a role for microRNA miR164c in regulating petal number in Arabidopsis [J]. Curr Biol , 2005 , 15( 4) : 303 – 315.
- [6] 卫波 , 张荣志 , 李爱丽 , 等. 利用高通量测序技术发现植物小分子 RNA 研究进展 [J]. 中国农业科学 , 2009 , 42( 11) : 3755 – 3764.
- [7] Jones-Rhoades M W , Bartel D P. Computational identification of plant microRNAs and their targets , including a stress-induced miRNA [J]. Mol Cell , 2004 , 14( 6) : 787 – 799.
- [8] Zhang B H , Pan X P , Cannon C , et al. Conservation and divergence of plant microRNA genes [J]. Plant J , 2006 , 46( 2) : 243 – 259.
- [9] Yin Z J , Li C H , Han X L , et al. Identification of conserved microRNAs and their target genes in tomato ( *Lycopersicon esculentum* ) [J]. Gene , 2008 , 414( 1/2) : 60 – 66.
- [10] Zhang B H , Wang Q L , Wang K B , et al. Identification of cotton microRNAs and their targets [J]. Gene , 2007 , 397( 1/2) : 26 – 37.
- [11] Zhang B H , Pan X P , Anderson T A. Identification of 188 conserved maize microRNAs and their targets [J]. FEBS Lett , 2006 , 580( 15) : 3753 – 3762.
- [12] Zuker M. Mfold web server for nucleic acid folding and hybridization prediction [J]. Nucleic Acids Res. 2003 , 31( 13) : 3406 – 3415.
- [13] Bonnet E , Wuys J , Rouze P , et al. Evidence that microRNA precursors , unlike other non-coding RNAs , have lower folding free energies than random sequences [J]. Bioinformatics , 2004 , 20( 17) : 2911 – 2917.
- [14] Zhang B H , Pan X P , Wang Q L , et al. Identification and characterization of new plant microRNAs using EST analysis [J]. Cell Res. 2005 , 15( 5) : 336 – 360.
- [15] Zhang B H , Pan X P , Cox B , et al. Evidence that miRNAs are different from other RNAs [J]. Cell Mol Life Sci. 2006 , 63( 2) : 246 – 254.

## microRNA and Target Gene Prediction of Banana based on Bioinformatics

PENG Yun , LONG Haifei , HE Rui , XU Min , ZENG Canbin , LUO Juan ,  
QI Zhao , YAN Zhen , RUAN Yunze , HE Chaozu , TANG Hua

( Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresources , College of Agronomy , Hainan University , Haikou , Hainan 570228 , China)

**Abstract:** microRNAs ( miRNAs) are a class of 21 – 23 nucleotides long non-coding RNA molecules in eukaryotic organisms , which play an important role in gene regulation by translation cutting or inhibition of target mRNA at the post transcriptional level. The expressions of miRNA are more rapid than those of the coding genes , are not influenced by the translation process , and have high regulation efficiency in the expression of target mRNA , and plant miRNAs play key roles in response or tolerance to biotic and abiotic stresses , growth and epigenetic inheritance. In this study , miRNAs in banana were predicted by using the bio-informatics methods. The BLAST searching between miRNAs of known plants from miRBase database and those of banana from the EST and GSS database showed that 16 conserved miRNAs in banana were grouped into 9 miRNA families: three from ETS database , and 13 miRNAs from GSS database. The target genes were then predicted with online software psRNATarget , and 168 target genes were found encoding many proteins and involving many biological processes.

**Keywords:** banana; miRNA; target gene; bioinformatics analysis