

# 葡萄 microRNA 的计算识别

蔡 斌<sup>1,3</sup>, 李成慧<sup>2</sup>, 彭日荷<sup>3</sup>, 熊爱生<sup>3</sup>, 高峰<sup>3</sup>, 姚泉洪<sup>3</sup>, 章 镇<sup>1</sup>

(1. 南京农业大学 园艺学院, 江苏 南京 210095; 2. 苏州农业职业技术学院, 江苏 苏州 215008;

3. 上海市农业科学院生物技术研究所, 上海 201106)

**摘要:**利用生物信息学方法, 根据已知植物 miRNA(microRNA) 的各种特征, 设计 miRNA 预测程序-MirFinder, 从葡萄全基因组范围中预测出 146 条 miRNA, 然后利用葡萄中已知的 miRNA 对这 146 条 miRNA 进行同源检索, 发现其中 98 条与已知的 miRNA 完全相同, 另外 48 个为新发现的 miRNA。这些新发现的 miRNA 基因属于 21 家族, 其中 8 个家族之前未在葡萄中报道过, 为葡萄中新发现的 miRNA 家族, 共有 20 个 miRNA。根据植物 miRNA 和其靶基因间存在较高的序列互补性, 预测新 miRNA 家族的靶基因, 结果表明其中 6 个 miRNA 家族存在靶基因。

**关键词:** 葡萄基因组; microRNA; 生物信息学

**中图分类号:** S663.1 **文献标识码:** A **文章编号:** 1000- 7091(2008) 增刊- 0213- 04

## Computational Identification of Novel microRNAs in Grape

CAI Bin<sup>1,3</sup>, LI Cheng-hui<sup>2</sup>, PENG Ri-he<sup>3</sup>, XIONG Ai-sheng<sup>3</sup>,  
GAO Feng<sup>3</sup>, YAO Quan-hong<sup>3</sup>, ZHANG Zhen<sup>1</sup>

(1. College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China;

2. Suzhou Polytechnic Institute of Agriculture, Suzhou 215008, China; 3. Biotechnology Research

Institute of Shanghai Academy of Agriculture Sciences, Shanghai 201106, China)

**Abstract:** microRNAs (miRNAs) represent a new family of non-protein-coding small RNAs (20– 25nt) that function by degrading targeted mRNAs or repressing mRNA translation. Here we developed a program-MirFinder for prediction of grape miRNAs. The characteristic features of known plant miRNAs were used as criteria to search for miRNAs. After searching the grape genome, MirFinder identified 146 miRNAs, of which 98 miRNAs were the same as known miRNAs and 48 are new identified miRNAs. The 48 miRNAs were classified into 21 families, of which 8 families are newly identified in grape. A total of 15 potential targets were identified for 6 of the 8 new miRNA families based on the fact that miRNAs exhibit perfect or nearly perfect complementarity with their target sequences

**Key words:** Grape genome; microRNA; Bioinformatics

microRNA(miRNA)是近年来发现的一类小分子RNA,广泛存在于真核生物中。在生物体内,miRNA基因经RNA聚合酶II转录产生1 kb左右长的pri-miRNA(primary-miRNA),经多步加工后形成长度约为22碱基的miRNA<sup>[1-3]</sup>。它可通过序列互补与mRNA分子相结合从而降解靶mRNA或抑制靶基因的翻译<sup>[4]</sup>。在植物中,miRNA的主要功能是调节与植物形态发生以及抗逆反应等多种生命过程有关基因的表达。由于miRNA具有非常重要的调控功能,

miRNA的鉴定成为近几年来生物学研究的一大热点。植物miRNA的鉴定始于拟南芥。在2002年,几个研究小组分别采用类似的试验方法,从双子叶植物拟南芥里总共克隆了43个miRNA<sup>[3-5]</sup>。在此之后,植物中发现的miRNA数量进入了快速增加的时期,大量新的miRNA被克隆和鉴定。不过,这些研究主要在拟南芥和水稻中进行<sup>[6,7]</sup>。

最初绝大多数miRNA都是通过大量的克隆和测序发现的,但是这些方法无法检测低丰度或组织

收稿日期:2008- 09- 05

基金项目:上海市科委基础研究重大项目(06DZ19103);国家自然科学基金(30670179)项目;国家“863”项目(2006AA06Z358)

作者简介:蔡 斌(1977- ),男,在读博士,主要从事园艺植物基因工程与生物信息学研究。

通讯作者:章 镇(1947- ),男,教授,主要从事果树生物技术与分子生物学研究。

特异性的 miRNA。因此除了试验克隆 miRNA 的手段之外,生物信息学方法逐渐发展起来。生物信息学方法根据在不同的物种中,成熟的 miRNA 具有较大的同源性以及前体的茎环结构具有相当大的保守性这些特征搜索新的 miRNA<sup>[8,9]</sup>。

葡萄是重要的果树树种,在世界果树生产中占据着重要位置。葡萄也因此成为人类完成全基因组测序的第一种水果作物和第四种开花植物<sup>[10]</sup>。在葡萄大规模测序已经完成的情况下,随着生物信息学方法的完善和成熟,使得人们可以在基因组水平上对具有重要功能的基因进行发掘和分析。由于 miRNA 具有重要的生物学功能,本试验采用生物信息学方法从葡萄基因组中鉴定和分析 miRNA,这有助于葡萄功能基因组学的研究。

## 1 材料和方法

### 1.1 材料

欧亚种葡萄(*Vitis vinifera* L.)全基因组序列由法国 Genoscope 国家基因测序中心的 FTP 服务器下载,更新时间为 2007 年 9 月 4 日。1 328 个植物 miRNA 均来自于 Rfam 数据库(版本号为 10.1)。

所用到的计算工具和软件:NCBI-BLAST 软件包<sup>[11]</sup>、Perl 脚本语言、Mfold<sup>[12]</sup>。

### 1.2 方法

1.2.1 葡萄 miRNA 的预测 根据 miRNA 前体的特征,利用 Perl 语言设计出用于探寻葡萄 miRNA 的程序—MirFinder,其工作流程见图 1。以 200 nt 作为窗口长度,5 nt 作为步长,从葡萄每条染色体的 5' 端开始滑动窗口,得到的序列片段跟葡萄基因组和蛋白质序列作 BLAST 比对,去除编码区序列。调用 RNAfold 预测片段的二级结构,判断其是否有发夹结构,保留有发夹结构的片段。将每个片段和植物中已知的 miRNA 进行 BLAST 比对(字段长度设为 7,期望值设为 10)。将和已知 miRNA 序列相差< 4 bp 的匹配片段作进一步的分析。满足以下条件的候选序列被认为是 miRNA 前体。①成熟 miRNA 的候选序列位于发夹结构的一条臂上,而对应的 miRNA\* 位于另一条臂上;②成熟 miRNA 和 miRNA\* 的非匹配碱基数不能超过 6 个,miRNA\* 区不能形成发夹环;③成熟 miRNA 上的单链区只能是内环和凸环,不能是多分支环;④预测的 miRNA 前体二级结构的自由能小于-20 kcal/mol,并且 AU 含量为 30%~70%。然后向下滑动 5 nt,重复上述步骤。

1.2.2 MiRNA 靶标的预测 将得到的 miRNA 序列对葡萄 mRNA 序列做 BLAST 比对。在 miRNA 和其

靶基因互补区内的非匹配碱基不多于 4 个,并且不允许有空位存在。

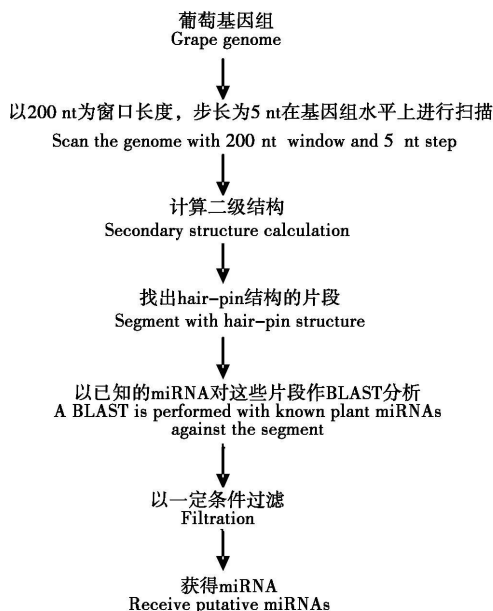


图 1 MirFinder 工作流程图

Fig. 1 Flowchart of MirFinder

## 2 结果与分析

### 2.1 葡萄中的 miRNA

将葡萄全基因组序列经过 MirFinder 程序分析后,得到 146 条成熟 miRNA 及其前体序列。然后再用已知的葡萄 miRNA 对这 146 条片段进行检索,以去除已发表的 miRNA,最后得到 48 个新的 miRNA。用 Mfold 对这 48 条进行重折叠,去除序列中的多余碱基。将前体序列对全基因组序列作 BLAST 分析,根据 BLAST 结果提取前体在染色体上的位置。对 miRNA 和其前体序列的长度、AU 含量作统计分析,结果见表 1。48 条 miRNA 的长度为 18~24 nt, AU 含量在 38%~59% 之间。

新发现的 48 个 miRNA 基因属于 21 个家族,其中 13 个家族的其他成员在葡萄中已有报道,本研究为这 13 个家族新增了 28 个 miRNA 基因成员。MiRNA169 为植物中最大的 miRNA 家族,本研究中新发现 9 个 miRNA 169,加上已报道的 15 个,在葡萄中共有 24 个。另外 8 个家族之前未见有文献报道,为葡萄中新发现的 miRNA 家族。这些家族为 miR403、miR477、miR482、miR529、miR827、miR828、miR845 和 miR1030。miR403 在葡萄中发现 3 个,在拟南芥(1 个)、杨树(3 个)中都存在,但在水稻中并未发现。miR403 可能是双子叶植物所特有的。miR477 在葡萄和杨树中存在,但在拟南芥中未见报道,推测 miR477 是类真蔷薇分支 I (Eurosids I) 所特有。

miR828 和 miR845 在葡萄和拟南芥中存在, 在杨树  中还未发现。

表 1  葡萄中新发现的 miRNA  
Tab.1  New miRNAs in grape

miRNA基因 New miRNA	序列 Sequence	长度/nt Length	AU含量/% A+U	染色体/位置(编码链) Chromosome/coordinate(Strand)	最小自由能 MFE
vvi-miR159d	UUUGGACUGAAGGGAGCUCCU	21	45	chr11/4274231-4274317(1)	-35.1
vvi-miR159e	AUUGGAGUGAAGGGAGCUCUA	21	50	chr15_random/3457753-3457926(1)	-87.4
vvi-miR169b	AAGCCAAGGAUGAAUUGCCGG	21	45	chr11_random/857467-857536(-1)	-31.9
vvi-miR169h	GAGCCAAGGAUGACUGGCCGU	21	38	chr11_random/1165413-1165494(-1)	-35.5
vvi-miR169i	GAGCCAAGGAUGGCUUGCCGU	21	38	chr11_random/1171586-1171667(-1)	-29.8
vvi-miR169n	GAGCCAAGGAUGACUUGCCGG	21	38	chr11_random/939429-939501(-1)	-36.4
vvi-miR169o	UGAGCCAAGGAUGACUUGCCG	21	43	chr11_random/1138138-1138218(-1)	-39.7
vvi-miR169q	UGAGCCAAGGAUGGCUUGCCG	21	38	chr11_random/1058477-1058559(-1)	-47.6
vvi-miR169v	UGAGCCAAGGAUGACUUGCCG	21	43	chr11_random/1133098-1133178(-1)	-41.4
vvi-miR169w	GAGCCAAGGAUGACUUGCCGG	21	38	chr11_random/943668-943745(-1)	-32.7
vvi-miR169x	UAGCCAAGGAUGACUUGCCUA	21	52	chr17/351774-351861(-1)	-31.2
vvi-miR171j	UUGAUUGAGCCGUGCCAAUAUC	22	52	chr18/1475185-1475286(1)	-33.83
vvi-miR171k	UUGAUUGAGCCGUGCCAAUAUC	22	52	chrUn_random/8941080-8941175(-1)	-40.95
vvi-miR172e	GGAAUUCUUGAUGACUUGCCAU	21	57	chr6/5961983-5962094(1)	-44.37
vvi-miR390b	AAGCUCAGGAGGGAUAGCGCC	21	38	chr8/8483680-8483780(1)	-44.4
vvi-miR393b	UCCAAAGGAUCGCAUUGAUCCC	23	50	chrUn_random/54807509-54807613(-1)	-38.3
vvi-miR394c	UUGGCAUUCUGUCCACCUCU	20	45	chr18/3503693-3503772(-1)	-41.3
Vvi-miR396e	UUCCACGGUUCUUGAACUU	21	57	chr1/1942589-1942748(1)	-48.9
vvi-miR396f	UUCCACAGGUUCUUGAACUG	21	57	chr4/5114333-5114418(-1)	-40.8
vvi-miR397	UCAUUGAGUCAGCGUUGAUG	21	52	chrUn_random/85971406-85971509(1)	-42.7
vvi-miR398b	UGUGUUCUCCAGGUCGCCCCUG	21	38	chr6/20660668-20660737(-1)	-35.9
vvi-miR398c	UGUGUUCUCCAGGUCGCCCCUG	21	38	chr6/14851890-14851959(1)	-37
vvi-miR399i	UGCCAAAGGAGAGAUUGCCUG	21	43	chr15/5804601-5804664(1)	-29.8
vvi-miR399j	UGCCAAAGGAGAUUUGCUC	19	53	chr10/1852044-1852126(-1)	-44.1
vvi-miR403a	UUAGAUUCACGCACAAACUCG	21	57	chr5/1740171-1740238(-1)	-35.24
vvi-miR403b	UUAGAUUCACGCACAAACUCG	21	57	chr5/1738525-1738619(-1)	-39.01
vvi-miR403c	UUAGAUUCACGCACAAACUCG	21	57	chr5/1304555-1304622(-1)	-38.44
vvi-miR477-3p	GUUGGAGGCCUUCUGGGACA	21	37	chrUn_random/77420884-77420946(-1)	-21.9
vvi-miR477a	CUCCCUCAAAGGCUUCCA	18	44	chrUn_random/53586956-53587022(1)	-37.6
vvi-miR477b	CUCCCUCAAAGGCUUCCA	18	44	chr19/13463692-13463759(1)	-35.2
vvi-miR477c	CUCCCUCAAAGGCUUCCA	18	44	chr19/13455047-13455113(1)	-40
vvi-miR477d	UCCCUCAAAGGCUUCCAA	18	50	chr12/10496324-10496398(1)	-32.4
vvi-miR477f	CUCCCUCAAAGGCUUCCA	18	44	chr6/24122600-24122666(-1)	-35
vvi-miR477g	CUCCCUCAAAGGCUUCCA	18	44	chr19/13657901-13657976(-1)	-37.1
vvi-miR482	UCUUCUUAACUCCUCCAUUCC	22	50	chr17/5579384-5579482(1)	-48.4
vvi-miR529	AGAAGAGAGAGAUACAGCU	20	50	chr5/17354592-17354661(-1)	-40.5
vvi-miR535f	UGACAAAGAGAGAGAGCAUGC	21	50	chrUn_random/71249809-71249887(1)	-41.8
vvi-miR535g	UGACAAAGAGAGAGAGCACAC	21	53	chrUn_random/71946360-71946439(-1)	-47.1
vvi-miR535h	UGACAAAGAGAGAGAGCACAC	21	53	chrUn_random/82215132-82215211(-1)	-48
vvi-miR535i	UGACAGCGAGAGAGAGCACAC	21	42	chrUn_random/71999223-71999302(1)	-41.2
vvi-miR827	UUAGAUCAUCAACAACAA	20	70	chr5/23154666-23154741(1)	-34.8
vvi-miR828a	UCUUGCUCAAAUGAGCAUCCA	22	59	chrUn_random/51301374-51301511(1)	-56.4
vvi-miR828b	UCUUGCUCAAAUGAGAUUCCA	22	64	chr16/7841231-7841330(1)	-39
vvi-miR845-3p	AGGCUCUGAUACCAAUUGAUG	21	55	chr7/503537-503608(-1)	-23.5
vvi-miR845a	UGCAUGCUCUGAUACCAAUUGAUG	24	58	chr5/2367971-2368113(-1)	-72.84
vvi-miR845b	GCAUGCUCUGAUACCAAUUGAUG	23	58	chr5/1915753-1915896(1)	-81.54
vvi-miR1030a	UCUGCUUUUGCAACCGCACCU	21	45	chr8/17999582-17999726(-1)	-52.8
vvi-miR1030b	UCUGCAUUUGCAACCGCACCU	21	45	chr6/22063243-22063369(1)	-53.02

表 2  葡萄中新 miRNA 家族的靶基因  
Tab.2  Potential targets of the new miRNA families in grape

miRNA 家族 miRNA families	靶基因 Targeted genes	功能 Target function	蛋白 Target proteins
477	GSVIVT00035115001, GSVIVT00010476001	转录因子	Scarecrow-like(GRAS 结构域) 蛋白(SCL)
529	GSVIVT00025360001, GSVIVT00036375001	转录因子	SQUAMOSA-启动子结合蛋白(SBP)
828	GSVIVT00034095001, GSVIVT00009522001, GSVIVT00032272001, GSVIVT00020733001, GSVIVT00016076001, GSVIVT00016944001	转录因子	MYB 蛋白(MYB)
845	GSVIVT00027557001, GSVIVT00022861001, GSVIVT00019911001		信号肽酶亚基蛋白 (Signal peptidase subunit protein)
1030	GSVIVT00016601001		ACC 合成酶(ACC synthase)
403	GSVIVT00028137001	未知	

2.2  葡萄 miRNA 的靶基因  提供的基因中搜索它们的靶基因, 结果发现其中 6 个以 8 个新发现的 miRNA 家族, 在 Genoscope 提 miRNA 家族存在靶基因, 共有 15 个(表 2)。转录因

子是许多 miRNA 的靶基因,它们往往调控植物的生长和形态发生。Scarecrow-like (GRAS 结构域) (SCL)、SQUAMOSA-启动子结合蛋白 (SBP) 和 MYB 分别是 miRNA 477、miRNA529 和 miRNA828 的靶基因产物。

### 3 讨论

miRNA 在基因调控中起着非常重要的作用。有研究认为,人类有 1/3 的基因由 miRNA 所调控<sup>[13]</sup>。因此,尽快找出所有的 miRNA 并研究其功能,对进一步了解基因表达调控具有重要意义。但由于试验的方法无法满足快速识别 miRNA 的要求,并且某些 miRNAs 的表达水平很低以及具有组织特异性或表达诱导特异性,很难用直接克隆法和遗传学方法获得,因此,利用生物信息学的方法来预测 miRNA 意义深远。

本试验利用 Perl 语言编写的程序—MirFinder 扫描整个葡萄基因组,结合与已知 miRNA 同源检索的方法是一种高效、准确的方法。通过对满足 miRNA 特征的序列进行严格筛选,大大降低了假阳性序列。为下一步的试验验证奠定了良好的基础。本研究中共识别出 146 个葡萄 miRNA,用已知的葡萄 miRNA 进行筛选后剩下 48 条,说明其中包含了大量的已知葡萄 miRNA,这也从另一个侧面反映了 MirFinder 具有较高的识别率。

和动物相比,植物 miRNA 和其靶基因间存在较高的互补性,这有利于通过生物信息学的方法来预测植物靶基因<sup>[14]</sup>。目前已经确定的 miRNA 所调控的靶标大部分为转录因子。在本研究中,miRNA477 的靶基因属于 SCL 家族。SCL 在植物信号转导、激素的调节以及分生组织的保持中起关键性作用。SBP 是 miRNA 529 的靶基因,主要参与植物生长、发育的多个方面,比如花器官的形成等。miRNA828 的靶标则是 MYB 类基因。MYB 在二次代谢、细胞的形态建成、对环境的应答和细胞周期中都起到一定的调控作用。

### 参考文献:

- [1] Bartel D P. MicroRNAs: genomics, biogenesis, mechanism, and function[J]. Cell, 2004, 116: 281–297.
- [2] Zhang B H, Pan X P, Cobb G P, et al. Plant microRNA: a small regulatory molecule with big impact[J]. Developmental Biology, 2006, 289: 3–16.
- [3] Park W, Li J, Song R, et al. Carpel Factory, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*[J]. Curr Biol, 2002, 12: 1484–1495.
- [4] Llave C, Kasschau K D, Rector M A, et al. Endogenous and silencing-associated small RNAs in plants[J]. Plant Cell, 2002, 14: 1605–1619.
- [5] Reinhart B J, Weinstein E G, Rhoades M W, et al. MicroRNAs in plants[J]. Genes Dev, 2002, 16: 1616–1626.
- [6] Griffiths-Jones S. The microRNA registry[J]. Nucleic Acids Research, 2004, 32: 109–111.
- [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] Weber M J. New human and mouse microRNA genes found by homology search[J]. FEBS J, 2005, 272: 59–73.
- [9] 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: 336–360.
- [10] Jaillon O, Aury J M, Noel B, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla[J]. Nature, 2007, 449: 463–468.
- [11] Altschul S, Madden T, Schaffer A, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs[J]. Nucleic Acids Research, 1997, 25: 3389–3402.
- [12] Zuker M. Fold web server for nucleic acid folding and hybridization prediction[J]. Nucleic Acids Research, 2003, 31: 3406–3415.
- [13] Lewis B P, Burge C B, Bartel D P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets[J]. Cell, 2005, 120: 15–20.
- [14] Rhoades M W, Reinhart B J, Lim L P, et al. Prediction of Plant microRNA Targets[J]. Cell, 2002, 110: 513–520.