

• 药材与资源 •

甘草 microRNA 及其靶基因的生物信息学预测

李华云, 王 宁, 韦春香, 张 彤, 付晨熙, 高 飞*, 周宜君

中央民族大学生命与环境科学学院, 北京 100081

摘要: 目的 利用生物信息学手段预测甘草 *Glycyrrhiza uralensis* 的 microRNA (miRNA), 并实验验证其存在, 确定相应的靶基因, 为理解甘草 miRNA 的生物学功能奠定基础。方法 下载公共核酸数据库中的甘草高通量测序序列和表达序列、标签序列, 使用 Trinity 和 Assembler 等软件拼接以获得转录本数据。基于 miRNA 前体序列在植物物种间的保守性, 将 miRBase 数据库中的已知植物 miRNA 前体与甘草转录组序列进行 Blast 比对, 按照 miRNA 前体应具备的标准进行筛选。通过 stem-loop qRT-PCR 实验验证 miRNA。使用 psRNATarget 软件进行 miRNA 的靶基因预测, 并对靶基因进行功能分析。结果 序列拼接获得 88 263 条甘草转录组数据。使用这些数据预测到分属于 17 个家族的 49 个 miRNA 序列, 以及相应的 30 个茎环结构前体, 随机选取其中 12 个, 经实验验证, 确实存在。这些 miRNA 靶向的 172 个基因主要参与基因转录调控、信号转导、发育调控和防御反应等生物学过程。结论 新预测的甘草 miRNA 及其靶基因为进一步研究 miRNA 在甘草中的生物学功能奠定了基础。

关键词: 甘草; miRNA; 靶基因; 生物信息学; 转录组

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2016)14-2506-09

DOI: 10.7501/j.issn.0253-2670.2016.14.020

Bioinformatic prediction of microRNAs in *Glycyrrhiza uralensis* and their target genes

LI Hua-yun, WANG Ning, WEI Chun-xiang, ZHANG Tong, FU Chen-xi, GAO Fei, ZHOU Yi-jun

College of Life and Environmental Sciences, Minzu University of China, Beijing 100081, China

Abstract: Objective To provide a good start for understanding the roles of microRNA in *Glycyrrhiza uralensis*, the microRNAs and their target genes were predicted using bioinformatic approach. **Methods** The deep sequencing data and EST sequences downloaded from public database were assembled using the Trinity and Assembler softwares to establish the transcriptome database. Since most of the plant miRNAs were conserved in plant species, all plant miRNAs stem-loop precursors were aligned to the assembled transcriptome database and the putative miRNA precursors were identified according to a rigorous criterion. psRNATarget was employed to predicted the targets of miRNA. **Results** A general transcriptome database with 88 and 263 sequences was obtained. Based on the transcriptomic sequences, 49 miRNA, classified into 17 families, arising from 30 stem-loop precursors, were identified. A total of 172 genes were predicted to be regulated by these miRNA, and these genes were involved into diversified biological processes including gene transcription regulation, signal transduction, development regulation, and defense response. **Conclusion** The miRNAs and the corresponding target genes identified in this study will provide a solid basis for understanding their biological functions in *G. uralensis*.

Key words: *Glycyrrhiza uralensis* Fisch.; microRNA; target gene; bioinformatics; transcriptome

甘草 *Glycyrrhiza uralensis* Fisch. 为豆科 (Leguminosae) 甘草属 *Glycyrrhiza* L. 多年生草本植物, 是中国传统大宗补益中草药, 药用部位是根及根茎。甘草具有祛痰止咳、清热解毒、抗癌和免疫

调节等诸多功能, 已被列入多国药典 (《中国药典》2010 年版、《日本药典》JP16、《美国药典》USP34 和《欧洲药典》Ph. Eur. 7) [1]。甘草主要分布于新疆、宁夏、内蒙古和甘肃等省区的干旱、半干旱的

收稿日期: 2016-01-05

基金项目: 国家自然科学基金资助 (31370356); 中央民族大学学术团队建设项目 (2015C8); 国家大学生创新创业训练计划 (GCCX2014110045)

作者简介: 李华云 (1990—), 男, 硕士在读, 研究方向为生物化学与分子生物学研究。Tel: (010)68932633 E-mail: lihuayun0331@163.com

*通信作者 高 飞, 男, 博士, 副教授, 硕士生导师, 主要从事生物化学与分子生物学研究。Tel: (010)68932633 E-mail: gaofei@muc.edu.cn

荒漠草原、沙漠边缘和黄土丘陵地带,根系发达,耐逆性强,具有防风固沙的功能,对于我国西北部荒漠地区生态环境的保护和改善具有重要意义^[2]。

microRNA (miRNA) 是一类在真核生物中发现的长度约为 21 nt 的单链非编码小分子 RNA。最早发现的 miRNA 是秀丽隐杆线虫 *Caenorhabditis elegans* 的 lin-4 和 let-7^[3-4]。miRNA 的成熟形式主要在细胞质中通过与靶基因的碱基互补配对,进而介导靶基因的降解或翻译抑制,在转录后水平对基因表达进行调控^[5]。在植物中,miRNA 通过负调控转录因子和其他重要的功能基因,广泛参与植物器官的生长发育、形态建成、逆境胁迫应答和次生代谢调控^[6]。

目前已经鉴定了超过 72 种植物的 7 385 种 miRNA,在每种植物中发现的 miRNA 种类数介于几十个到几百个之间^[7]。整体说来,对于拟南芥、水稻、毛果杨等模式植物的 miRNA 研究较多,对于中草药等植物的研究相对较少。中草药在中国传统医药中具有重要地位。miRNA 在植物体内发挥重要的调控作用,参与植物环境逆境应答以及次生代谢产物的生物合成,研究药用植物 miRNA 的生物学功能对于理解药用成分形成和积累的生态调控机制,培育高产优质的中药材具有重要意义。近年来,对药用植物 miRNA 的研究正在迅速开展起来,研究人员利用高通量测序、miRNA 芯片、生物信息分析等方法研究了人参、毛地黄、半夏、千里光等中草药的 miRNA^[8-9]。但迄今为止,未见甘草 miRNA 的相关研究。本研究拟应用生物信息学工具,基于甘草转录本序列鉴定 miRNA,并确定其靶基因,为进一步研究 miRNA 在甘草逆境应答和次生代谢途径调控中的作用奠定基础。

1 材料

甘草种子来自内蒙古自治区鄂尔多斯市,经中央民族大学生命与环境科学学院刘博博士鉴定为甘草 *Glycyrrhiza uralensis* Fisch. 的根及根茎。

Trizol 试剂购自美国英杰生命技术有限公司 (Invitrogen),逆转录试剂和荧光定量 PCR 试剂盒购自天根生化科技(北京)有限公司,引物由生工生物工程(上海)股份有限公司合成。其余试剂均为分析纯。

2 方法

2.1 甘草核酸数据的获取和转录本数据拼接

在 NCBI 网站 (<http://www.ncbi.nlm.nih.gov>) 下

载甘草的高通量测序数据和 EST 序列。高通量数据登录号为 DRR006519、DRR006520、DRR006521、DRR006522、DRR006523、DRR006524、DRR006525 和 DRR006526。表达序列标签 (EST) 数据有 55 942 条。数据过滤软件使用 Trimmomatic (<http://www.usadellab.org/cms/index.php?page=trimmomatic>)^[10] 和 solexaQA (<http://solexaqa.sourceforge.net/>)^[11]。拼接软件使用 Trinity (<http://trinityrnaseq.github.io/>)^[12] 和 iAssembler (v1.3.2, <http://bioinfo.bti.cornell.edu/tool/iAssembler/>)^[13],拼接采用默认参数进行。聚类去冗余软件使用 CD-HIT (<http://weizhong-lab.ucsd.edu/cd-hit/>)^[14]。

2.2 甘草 miRNA 的生物信息学预测

整体预测流程参考文献方法^[15]进行。简言之,首先从植物 miRNA 数据库 (<http://bioinformatics.cau.edu.cn/PMRD>) 中下载 10 898 个植物来源的 miRNA 茎环结构前体,去除不同物种 miRNA 之间的重复序列以及不可靠的数据后,剩余的 4 835 条 miRNA 前体序列作为预测甘草 miRNA 前体的参照序列与拼接得到的甘草转录本序列进行同源比对分析 (Blastn),筛选出 $E < 1e-6$ 的转录本序列。这些转录本序列去除冗余后,用 Mfold 软件 (<http://mfold.rna.albany.edu/q=mfold/RNA-Folding-Form>)^[16] 进行折叠,预测候选 miRNA 前体序列的二级结构,分析是否有 miRNA 成熟形式存在茎环结构上茎的部分,评价茎环结构的稳定性。miRNA 前体的二级结构筛选标准参照 Zhang 等^[17]的方法。

2.3 甘草根中 miRNA 的检测

基于甘草的主要药用部位是根茎,选取培养了 8 周的甘草幼苗根部,用 Trizol 法提取根中的总 RNA,并分别用 Nano DropTM2000 分光光度计、琼脂糖凝胶电泳验证总 RNA 的纯度、浓度、完整性。

随机选取生物信息学预测中的 12 条 miRNA,并分别设计逆转录茎环引物及荧光定量 PCR 引物 (表 1),利用 stem-loop 荧光定量 PCR 验证 miRNA。

2.4 靶基因的预测和 GO 分析

将全部靶基因序列比对到拟南芥蛋白数据,用拟南芥蛋白注释靶基因,以了解靶基因的生物学功能,截取标准为 $E < 1e-5$ 。使用 agriGO (<http://bioinfo.cau.edu.cn/agriGO/analysis.php>) 中的 GO 分析工具对靶基因进行功能分类和富集分析。

表 1 待验证的 miRNAs 及其对应引物

Table 1 Selected miRNAs and their corresponding primers

miRNAs	miRNA 序列 (5'-3')	逆转录引物 (5'-3')	正向引物 (5'-3')	反向引物 (5'-3')
gur-miR159e-3p	UUUGGAUUGAAGG	GTTGGCTCTGGTGCAGGGTCCGAGG	GTGGGTTTGGAT	GTGCAGGGTCCG
	GAGCUCUA	TATTCGCACCAGAGCCAACAGAGCT	TGAAGGG	AGGT
gur-miR390-5p	AAGCUCAGGAGGG	GTTGGCTCTGGTGCAGGGTCCGAGG	GTTGAAGCTCAG	GTGCAGGGTCCG
	AUAGCG	TATTCGCACCAGAGCCAACGGCGCT	GAGGGAT	AGGT
gur-miR393b-5p	UCCAAAGGGAUCG	GTTGGCTCTGGTGCAGGGTCCGAGG	GTGTTCCAAAGG	GTGCAGGGTCCG
	CAUUGAUC	TATTCGCACCAGAGCCAACAATCAA	GATCGCA	AGGT
gur-miR482a-3p	UUCCCAAAGCCGC	GTTGGCTCTGGTGCAGGGTCCGAGG	TTGTTCCCAAAG	GTGCAGGGTCCG
	CCAUUCCGA	TATTCGCACCAGAGCCAACCTCGGAA	CCGCC	AGGT
gur-miR164a-5p	UGGAGAAGCAGGG	GTTGGCTCTGGTGCAGGGTCCGAGG	GTTTTGGAGAAG	GTGCAGGGTCCG
	CACGUGCA	TATTCGCACCAGAGCCAACCTGCACG	CAGGGCA	AGGT
gur-miR396i-5p	UUCCACAGCUUUC	GTTGGCTCTGGTGCAGGGTCCGAGG	GGGGTTCCACAG	GTGCAGGGTCCG
	UUGAACUG	TATTCGCACCAGAGCCAACCAGTTC	CTTTCTT	AGGT
gur-miR156b-5p	UUGACAGAAGAUUA	GTCGTATCCAGTGCAGGGTCCGAGG	CCGCGTTGACAG	CAGTGCAGGGTC
	GAGAGCAC	TATTCGCACTGGATACGACGTGCTC	AAGATAGAGA	CGAGGTA
gur-miR394b-5p	UUUGGAUUGAAGG	GTTGGCTCTGGTGCAGGGTCCGAGG	GTTGGTTGGCATT	GTGCAGGGTCCG
	GAGCUCAA	TATTCGCACCAGAGCCAACGGAGGT	CTGTCC	AGGT
gur-miR166g-3p	UCGGACCAGGCUU	GTCGTATCCAGTGCAGGGTCCGAGG	CGGTCGGACCAG	CCAGTGCAGGGT
	CAUUCCCC	TATTCGCACTGGATACGACGGGGAA	GCTTCATT	CCGAGGTATTC
gur-miR167b-5p	UGAAGCUGCCAGC	GCGTGGTCCACACCACCTGAGCCGC	GAGCCGTGAAGC	TCCACACCACCT
	AUGAUCUGA	CACGACCACGCTCAGATCA	TGCCAG	GAGCCG
gur-miR168a-5p	UCGCUUGGUGCAG	CTCAGCGGCTGTCGTGGACTGGGTG	CGGTGTGTCGCT	GGCTGTCGTGGA
	GUCGGGAA	CTGCCGCTGAGTTCCCGACTTCCCG	TGGTGC	CCTGGGTG
gur-miR2118-3p	UUGCCGAUUCAC	GTCGTATCCAGTGCAGGGTCCGAGG	AAGGCGTTTCCG	CAGTGCAGGGTC
	CCAUUCCU	TATTCGCACTGGATACGACTAGGAA	ATTCCA	CGAGGTAT

3 结果与分析

3.1 甘草转录本数据的拼接

首先用 Trimmomatic 软件对 Illumina HiSeq2000 测序所得数据进行滤过；用 solexaQA 对 EST 数据进行滤过，获得去除测序接头、去除低质量数据、去除冗余序列的可用于拼接的数据。然后分别用 Trinity 和 iAssembler 混合拼接 Illumina Heseq 2000 和 EST 数据。最后将上述 2 部分拼接好的数据用 CD-HIT 合并聚类。

本研究拼接获得的 unigene 的平均长度为 896 bp，略小于 Ramilowski 等^[18]的拼接结果，但远大于 Li 等^[19]的拼接结果。本研究的 unigene 的 N50(1 493 bp) 和拼接总长度 (67 122 825 bp) 明显大于其他 2 个拼接结果。总的来看，本研究获得的转录本数据覆盖度最高，质量较好，可以用于 miRNA 的生物

信息学预测等分析。

3.2 预测到的甘草 miRNA、前体及其二级结构

以已知植物 miRNA 前体序列为探针与拼接获得的 88 263 条甘草转录本序列进行 BLASTN 比对分析，筛选出 685 条候选 miRNA 前体序列。去除其中的蛋白编码基因后，进一步对这些转录本进行手动筛选，确定了其中能够用 Mfold 软件折叠形成稳定的茎环结构，并符合相应筛选标准的核酸序列共 50 条，代表 30 个茎环前体序列 (表 2 和图 1)。在这些前体序列上分布着分属于 17 个不同家族的 49 个 miRNA 的成熟序列 (表 3)。其中，家族成员最多的是 miR156 家族，有 8 个成员；其次为 miR159 家族，有 7 个成员。

在甘草中预测到的 49 条成熟 miRNA 的长度范围为 20~24 nt，其中 21 nt 长度的 miRNA 最多 (31/49, 63%)，其次为 22 nt 的 miRNA (11/49)，

其他长度的 miRNA 数量较少。预测的 30 个 stem-loop 前体的序列长度介于 82~187 nt, 这些前体都能折叠成稳定的二级结构, 最小自由能 MFE 介于 -101.9~-30.4 kcal/mol, 最小自由能指数 MFEI 介于 0.67~1.03 (表 2)。

本实验考察了 49 个预测的甘草 miRNA 与其他植物来源的同源 miRNA 的序列相似性(表 3)。71% (35/49) 的甘草 miRNA 与其他植物中对应 miRNA 的序列完全相同, 10% (5/49) 的甘草 miRNA 出现

1 个差异碱基, 出现多于 1 个差异碱基的甘草 miRNA 共有 8 个, 占 16%。这说明本研究预测的 miRNA 多数为保守 miRNA。还考察了 31 个预测的甘草 miRNA 前体与其他植物来源的 miRNA 前体的序列相似性(表 3)。发现所有前体与其他植物来源的 miRNA 前体都有程度不同的相似性, E 值最大为 $3e-12$, 最小为 $7e-42$ 。这一结果说明, 与蛋白编码基因类似, 同源 miRNA 的成熟形式和茎环结构前体在序列上也具有保守性。

表 2 甘草中新发现的 miRNA 前体

Table 2 Precursors of miRNA newly discovered from *G. uralensis*

名称	茎环前体序列	前体同源物	E 值	MFE/(kcal·mol ⁻¹)	MFEI
gur-MIR156a	gagagacUGACAGAAGAGAGUGAGCACAugcugcuagugauuguaugaa ggguauacaauugcgggugcguGCUCACUUCUCUUUCUGUCAGCuucuc	gma-MIR156a	2×10^{-25}	52.80	1.12
gur-MIR156b	ggugaucugUUGACAGAAGAUAGAGAGCACagaugaugaaugcauga aagggauggcacucacuccuuuguGCUCUCUAUACUUCUGUCAUCacc	gma-MIR156e	1×10^{-35}	47.90	1.09
gur-MIR156a	ggugaucugCUGACAGAAGAUAGAGAGCACagaugaugaaugcauga aagggauggcacucacuccuuuGUGCUCUCUAUUCUUCUGUCAucacc	gma-MIR156e	2×10^{-34}	45.70	1.04
gur-MIR156g	uugggacauagaaaUGACAGAAGAGAGUGAGCACacagaggcacugca uauagcuauuacugcuuuuguguGCUCACUUCUCUCUCUGUCACC uuccagugccuga	tcc-MIR156g	1×10^{-29}	55.50	1.11
gur-MIR159a	uggagcucuuuccuccauuucugaaacgaggugauaguagaccuagcugcuaguuca uggauaccucuggcucugcuagcaaauggaccuacuuuugcgaaggucaggucugca ugaacuuggagauaggguaaaccuucaccguuuugguAUUGGAGUGAAGG GAGCUCCA	gma-MIR159f	2×10^{-21}	82.30	0.97
gur-MIR159e	guuauaggagugagcuccuugaaguccaaacgaggauugcuggguuugauagAGCU GCUGAGCUAUGGAUCCCAuagugcuacccuuuaucaugugguaguccugc ggcuuccauauccegggagcuuacccuuuugcucuuauccuucUUUGGAUUGA AGGGAGCUCUAcaccuuaau	gma-MIR159e	4×10^{-37}	83.50	0.93
gur-MIR162	ggaaggugaagucacUGGAGGCAGCGGUUCAUCGAUCucuuccugaguuu ggauguggaagaacacaaagcaugaauccgUCGAUAAACCUCUGCAUCCAG cgucacuuugcc	vun-MIR162	1×10^{-38}	44.40	0.77
gur-MIR164k	ggucuuguUGGAGAAGCAGGGCACGUGCAacacaucaagaucucucuaa acuaguaguugucugauuuugcagugcuccccuuuuccaacaugauu	gma-MIR164k	3×10^{-12}	43.30	0.96
gur-MIR164a	guUGGAGAAGCAGGGCACGUGCAaaauuaaauccucugguucucuaaag ccaauagcaaaugcuUGCACGUGCUCUUUUUCUCCAAc	tcc-MIR164a	2×10^{-13}	43.00	1.00
gur-MIR166g	guugaggggaugcugcugguucgagaucauuaaauagucuaucucagguuaac ucuucacauagagauuUCGGACCAGGCUUCAUCCCCCucagc	gma-MIR166g	3×10^{-21}	55.60	1.18
gur-MIR166e	ggaagcuuuuuuuugaggGGAUGUUGUCUGGCUCGAGGacccuucu ccuugauccagugauuuuuccucucuaugauagagagauagaggauugcugugua cuacuugggucaaggaaugauaguguugUCGGACCAGGCUUCAUCCCCC ccaauuaauugcuucc	csi-MIR166e	2×10^{-18}	76.30	0.95
gur-MIR167b	uauccacuuacaguaguUGAAGCUGCCAGCAUGAUCUGAagcuaccuauu ucuauuucuaauggugagaauagaugaugugguagcuuaccuucuauguaagcaua	lja-MIR167	3×10^{-23}	49.80	1.16
gur-MIR167a	ggucacaaagaaaaagUGAAGCUGCCAGCAUGAUCUAgcuuugguuaga gagcgcugcgcuaagcugaauaaggaguguuagcccgaaugaaccugacuaGGUCA UGCUGUGACAGCCUCACUccuuccuauuuuggggacc	lja-MIR167a	2×10^{-22}	58.90	0.81
gur-MIR168a	cgcggucucuaauUCGCUUGGUGCAGGUCGGGAAccgcuaucugcugcgau uuucgcgcguauagacggagcggucggcgcgugaaauuggauCCCGCCUUGCAU CAACUGAAUcggagccgcg	gma-MIR168a	1×10^{-27}	72.20	0.94

名称

名称	茎环前体序列	前体同源物	<i>E</i> 值	MFE/(kcal·mol ⁻¹)	MFEI
gur-MIR171c	GAUAUUGGUGCGGUUCAUCAgaaggcgguccuuuaaucaaaaaagc ucuguguuauuUGAUUGAGCCGUGCCAAUAUC	gma-MIR171c	3×10^{-25}	39.10	1.15
gur-MIR171b	agcaaGGUAUUGGCGCGUCUCAUUUgaagacauggcuauuuaucaugag aauccagccauguaguuUGAUUGAGCCGCGUCAUAUAUCuuguu	lja-MIR171b	3×10^{-22}	47.70	1.22
gur-MIR172	agucgcuguuugccgauGGAGCACCACAAGAUUCACAaguuuuugggu uuuuuuuuuugguuuugggguggucccccuaauugcuguccaaaaucuuuagcccuug auaUGAGAAUCUUGAUGAUGCUGCAgcggcauaaaugacu	gma-MIR172e	1×10^{-28}	62.20	1.22
gur-MIR2118	gagagagaaagggaaaggggaagagcuugggggaaguuaugGGAGAUGGGAGG UUCGGCAAAAcagaucugaauccuucacaaagaauucuguuuuuUUGCCG AUUCCACCCAUCCUAuguuuuccucugguuccucuuuccacucuccucuc	mtr-MIR2118	5×10^{-34}	79.00	1.07
gur-MIR319a	ggauuuauUAGCUGCCGACUCAUUAUCCAaanguugagucauuuag auccauacucaccaaauccgugaauaugcgggagacaaaauugaauu	gma-MIR319a	9×10^{-23}	30.70	0.81
gur-MIR319e	gauugaaggggcucacucaguccagucgagcgcgguugaagugcugagcacagcugc ugacucguugguugcagagcucaccaugaauuagccucuuuuuuuuggugucucugau ccaacgaugcgggagcugccacucagucguugcugucaugucuuggacugaaggggccucu ucauuc	mes-MIR319e	3×10^{-22}	101.90	1.03
gur-MIR390	aguaugggagaauucuguaAAGCUCAGGAGGGAUAGCGccauggauauugcu ucuccauuuugaugggauuuugcaauuuauaaucauuauucugaucuucucuauagCG CUAUCCAUCCUGAGUUUCAuggguuucucuauc	mes-MIR390	3×10^{-15}	61.45	1.10
gur-MIR393b	aggcaUCCAAAGGGAUCGCAUUGAUCCaaauucaguaacccuuauuuu ucucucuguuuuuccacuuauuuccuuauuucucacaaauuugggAUCAUGCUAU CCCUUUGGAUuccu	mtr-MIR393b	3×10^{-17}	37.52	0.78
gur-MIR394b	ucaugggggcuacaaaagguagcuuacagaguuuuUUGGCAUUCUGUCCAC CUCCcucccucccuccacucaauuuuauuuccaaugucucuauagcucuguugcgcg ccgcguggAGGUGGGCAUACUGCCAACUcagcucucuuggcuucucuuuugu aaagccuuccga	mdm-MIR394b	8×10^{-21}	63.46	0.71
gur-MIR396i	ggccucuuuguauuucUCCACAGCUUUCUUGAACUGcauucaaagaguuc uuugcaugcaugccauggcacucuugcuccaccacacuuuuugcaGUUCAUA AAGCUGUGGGAAuacagauagguc	gma-MIR396i	7×10^{-42}	60.49	0.96
gur-MIR396g	gucaugcuUCCACAGCUUUCUUGAACUucuuuuuugcaucuuuuuagc ucucuacaggaucauuuuuagccuagguagcuacuagAAGCUCAAGAAAGCU GUGGGAGaauaugc	gma-MIR396g	3×10^{-17}	43.52	0.89
gur-MIR399a	uaagcaauacagcuaUAGGGCUUCUCUUCUUGGcaggaauuuaucaugacca uuuccauggugugucuuGCCAAAGGAGAGUUGCCCUGuugcugcuuuagcu ua	pvu-MIR399a	1×10^{-29}	55.30	1.20
gur-MIR482a	ggauggggagauuggagcuacagaaguugGGAAUGGGCGGUUGGGAA GAaaucagaaauccuuuuuuuuuaguuuuuUCCCCAAAGCC GCCCAUCCGAuguuuuugguuuccccuuucc	mdm-MIR482a	2×10^{-18}	73.50	1.23
gur-MIR482b	gggauggggagauuggagaaauuagaguugGGAAUGGGUUGGUUGGGA UGAaagauuuacuccaacaacaauaacguuguguguuuUCUCCCCAAUUC CGCCCAUCCUAuuuuuuuauuuuuccuccuuucc	mdm-MIR482a	1×10^{-20}	69.30	1.16
gur-MIR1507b	auguuuugguAGAGGCGUAUGGAGUGAGAGAUuggaaagcucuuccgauu ucuagucauuuauuuuccuUUUCAUCCAUUGUCAUCUAACgaacgu	mtr-MIR1507	4×10^{-14}	43.00	1.08
gur-MIR1507a	auguuuuggcagaggcauuggagugagagcuuagaagcauuuuuccgauuucuuugcauu gaaccuuucccuUCUCGUUCCAACGUCGUCUAACgaacau	gma-MIR1507a	1×10^{-13}	30.40	0.67

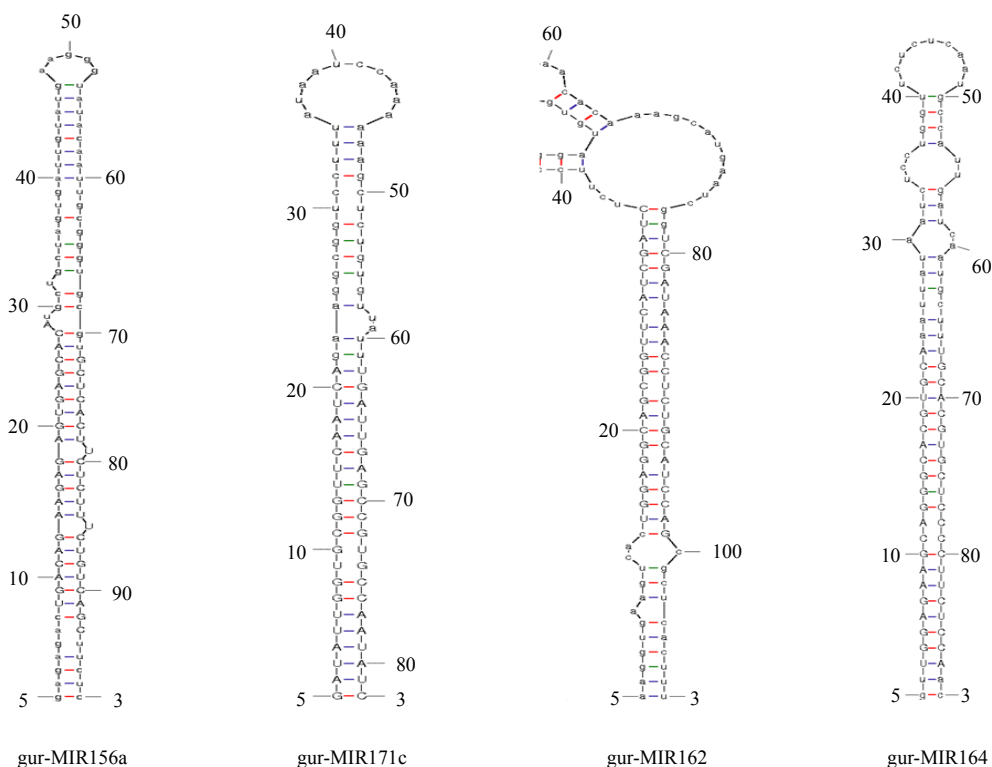


图 1 部分 miRNA 前体的二级结构

Fig. 1 Secondary structures of some precursors of miRNA

表 3 甘草中新发现的 miRNA

Table 3 miRNAs newly discovered from *G. uralensis*

序号	家族	名称	序列	长度/nt	同源 miRNA	与同源 miRNA 的差异碱基数
1	156	gur-miR156a-5p	UGACAGAAGAGAGUGAGCACA	21	bna-miR156a	0
2		gur-miR156a-3p	GCUCACUUCUCUUUCUGUCAGC	22	zma-miR156d-3p	0
3		gur-miR156b-5p	UUGACAGAAGAUAGAGAGCAC	21	ath-miR157a-5p	0
4		gur-miR156b-3p	GCUCUCUAUACUUCUGUCAUC	21	aly-miR157c-3p	1
5		gur-miR156e-5p	CUGACAGAAGAUAGAGAGCAC	21	smo-miR156b	0
6		gur-miR156e-3p	GUGCUCUCUAUUCUUCUGUCA	21	aly-miR157c-3p	3
7		gur-miR156g-5p	UGACAGAAGAGAGUGAGCACA	21	bna-miR156a	0
8		gur-miR156g-3p	GCUCACUUCUCUCUCUGUCACC	22	bdi-miR156b-3p	0
9	159	gur-miR159a-3p	AUUGGAGUGAAGGGAGCUCCA	21	gma-miR159b-3p	0
10		gur-miR159e-3p	UUUGGAUUGAAGGGAGCUCUA	21	ath-miR159a	0
11		gur-miR159e-5p	AGCUGCUGAGCUAUGGAUCCC	21	gma-miR159d	1
12	162	gur-miR162-5p	UGGAGGCAGCGGUUCAUCGAUC	22	csi-miR162-5p	0
13		gur-miR162-3p	UCGAUAAACCUCUGCAUCCAG	21	ath-miR162a-3p	0
14	164	gur-miR164k-5p	UGGAGAAGCAGGGCAGUGCA	21	ath-miR164a	0
15		gur-miR164a-5p	UGGAGAAGCAGGGCAGUGCA	21	ath-miR164a	0
16	166	gur-miR166g-3p	UCGGACCAGGCUUCAUCCCC	21	pvu-miR166a	0
17		gur-miR166e-3p	UCGGACCAGGCUUCAUCCCC	21	pvu-miR166a	0
18	167	gur-miR167b-5p	UGAAGCUGCCAGCAUGAUCUGA	22	ccl-miR167a	0
19		gur-miR167a-5p	UGAAGCUGCCAGCAUGAUCUA	21	ath-miR167a-5p	0
20		gur-miR167a-3p	GGUCAUGCUGAGACGCCUCACU	23	aly-miR167b-3p	1

续表 3

序号	家族	名称	序列	长度/nt	同源 miRNA	与同源 miRNA 的差异碱基数
21	168	gur-miR168a-5p	UCGCUUGGUGCAGGUCGGGAA	21	ath-miR168a-5p	0
22	171	gur-miR171c-5p	AGAUUUGGUGCGGUUCAAUC	21	gma-miR171c-5p	0
23		gur-miR171c-3p	UGAUUGAGCCGUGCCAAUAUC	21	gma-miR171e	0
24		gur-miR171b-3p	UGAUUGAGCCGCGUCAAUAUC	21	vvi-miR171b	0
25	172	gur-miR172e-3p	UGAGAAUCUUGAUGAUGCUGCA	22	vvi-miR172d	0
26	2118	gur-miR2118-5p	GGAGAUUGGAGGUUCGGCAAA	21	gma-miR2118a-5p	2
27		gur-miR2118-3p	UUGCCGAUCCACCCAUUCCU	21	gma-miR2118a-3p	0
28	159	gur-miR319a-5p	UAGCUGCCGACUCAUUAUCCA	22	ppe-miR319b	1
29		gur-miR319e-5p	AGCUGCUGACUCGUUGGUUC	20	mtr-miR4414a-3p	0
30		gur-miR319e-3p	AUCCAACGAUGCGGAGCUGC	21	mtr-miR4414a-3p	0
31		gur-miR319b-3p	UUGGACUGAAGGGGCCUCU	20	gma-miR319f	0
32	390	gur-miR390-5p	AAGCUCAGGAGGGAUAGCG	21	ath-miR390a-5p	0
33		gur-miR390-3p	CGCUAUCCAUCUGAGUUUCA	21	aly-miR390a-3p	0
34	393	gur-miR393b-5p	UCCAAAGGGAUCGCAUUGAUC	21	mtr-miR393b-5p	0
35		gur-miR393b-3p	AUCAUGCUAUCCCUUUGGAU	21	mtr-miR393b-3p	0
36	394	gur-miR394b-5p	UUGGCAUUCUGUCCACCUC	20	gma-miR394c-5p	0
37		gur-miR394b-3p	AGGUGGGCAUACUGCCAACU	20	gma-miR394b-3p	1
38	396	gur-miR396i-5p	UUCCACAGCUUUCUUGAACUG	21	gma-miR396a-5p	0
39		gur-miR396i-3p	GUUCAUAAAGCUGUGGGAAG	21	gma-miR396i-3p	0
40		gur-miR396g-5p	UUCCACAGCUUUCUUGAACU	21	gma-miR396b-5p	0
41		gur-miR396g-3p	AAGAAAGCUGUGGGAGAAUAUGGC	24	gma-miR396d	0
42	399	gur-miR399a-3p	UGCCAAAGGAGAGUUGCCUG	21	pvu-miR399a	0
43	482	gur-miR482a-5p	GGGAAUGGGCGGUUGGGAAGA	22	mdm-miR482a-5p	2
44		gur-miR482a-3p	UUCCCAAAGCCGCCAUUCCGA	21	mdm-miR482a-3p	3
45		gur-miR482b-5p	GGGAAUGGGUUGGUUGGGAUGA	22	mdm-miR482a-5p	3
46		gur-miR482b-3p	UCUCCCCAAUCCGCCCAUUCCUA	24	gma-miR482a-3p	0
47	1507	gur-miR1507b-5p	AGAGCGUAUGGAGUGAGAGAU	22	mtr-miR1507-5p	5
48		gur-miR1507b-3p	UUUCAUCCAUAUGUCAUCUAA	22	mtr-miR1507-3p	4
49		gur-miR1507a-3p	UCUCGUCCAUAACGUCGUCUAA	22	gma-miR1507a	3

3.3 甘草根中 miRNA 的验证

为了确定生物信息学预测中的 miRNA 真实存在,本研究随机选取了 12 条新发现的 miRNA,通过 stem-loop qRT-PCR 实验进行验证。实验结果显示,12 条 miRNA 荧光定量扩增产物的熔解曲线均为单峰,将 12 条 miRNA 荧光定量 PCR 产物进行琼脂糖凝胶电泳,均出现明显单一条带,大小约 60 bp,与荧光定量 PCR 扩增产物的实际大小一致(图 2),证明本研究通过生物信息学预测的 miRNA 在甘草中确实存在。

3.4 甘草 miRNA 靶基因的预测及功能分析

根据预测的 49 条甘草 miRNA 序列以及拼接获得的转录本序列,通过在线软件 psRNATarget 对其靶基因进行了预测,共预测到 32 条甘草 miRNA 的 273 个靶基因,其中有 172 个靶基因可以用拟南芥蛋白序列进行注释。靶基因编码的蛋白具体包括以下几类:(1) 转录因子,包括 SBP 类、myb 类、GRAS 类、bHLH 类、NAC 类、bZIP 类、TCP 类等

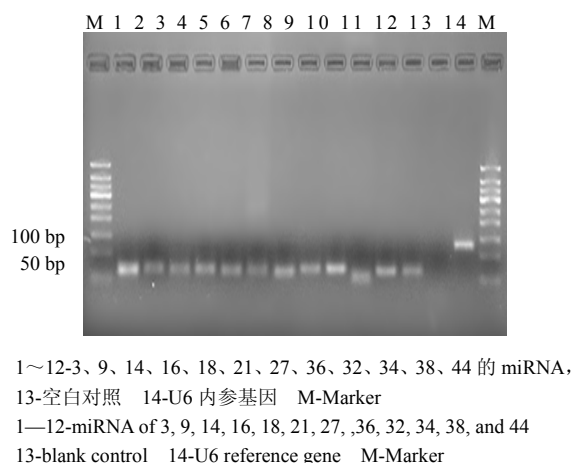


图 2 miRNA stem-loop qRT-PCR 产物电泳图

Fig. 2 Electrophoresis of miRNA stem-loop qRT-PCR products

转录因子;(2) 激酶和磷酸酶,包括 MAPKKK5、富含亮氨酸重复的蛋白激酶 Leucine-rich repeat protein kinase、U-box domain-containing protein kinase、protein phosphatase 2CA 等;(3) 激素信

号转导成员,包括 ARF (生长素响应因子)、生长调节因子 5 (GRF5 | growth-regulating factor 5)、AFB2 | auxin signaling F-box 2 等;(4) 抗病蛋白,如 NB-ARC domain-containing disease resistance protein、LRR and NB-ARC domains-containing disease resistance protein、Disease resistance protein (TIR-NBS-LRR class) family;(5) 其他类别的酶等。GO 分类和富集分析结果(图 3)表

明被富集的生物学过程为 DNA 依赖的转录调控(GO: 0006355)、器官发育(GO: 0048513)和防御反应(GO: 0006952)。被富集的分子功能包括转录因子活性(GO: 0003700)和泛素蛋白连接酶活性(GO: 0004842)。以上分析结果表明,甘草 miRNA 主要通过调控转录因子等靶基因参与基因转录调控、信号转导、胁迫应答、蛋白翻译后修饰等生物学过程。

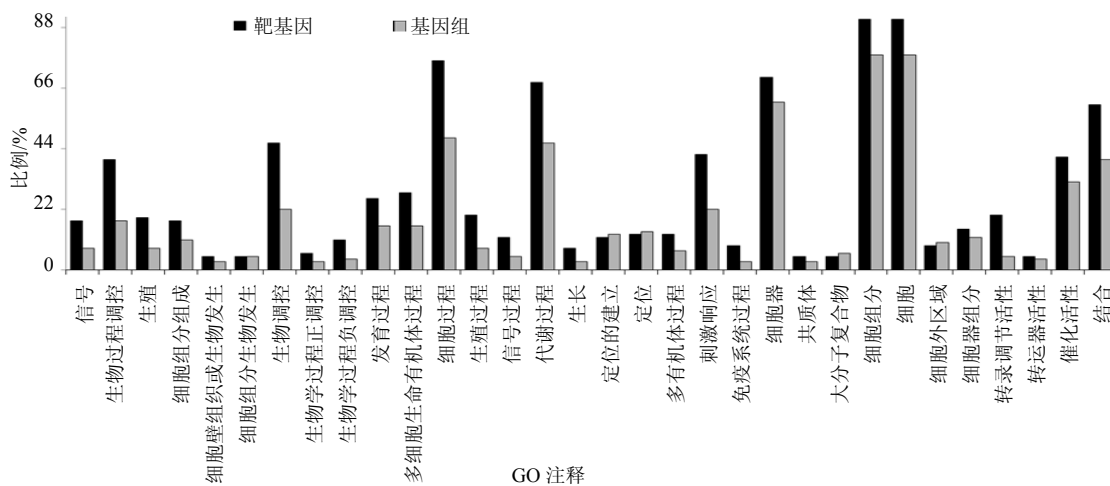


图 3 甘草 miRNA 靶基因编码蛋白的 GO 分类情况

Fig. 3 GO classification of predicted miRNA target genes in *G. uralensis*

4 讨论

与动物 miRNA 相比,植物 miRNA 的研究相对较晚。尽管如此,随着第二代测序技术的飞速发展,研究人员已从数十种植物中鉴定出少到几十个、多至几百个的 miRNA,并对部分植物的 miRNA 进行了进一步的功能分析。这些研究表明 miRNA 广泛参与调控植物的生长发育、形态建成、次生代谢物生成、信号转导以及环境因子应答等^[20]。甘草作为我国传统大宗中草药,同时也是维持我国西北部干旱区域生态系统健康的重要植物^[1-2],其次生代谢产物的调控、积累机制和干旱等逆境适应机制都与 miRNA 存在密切的内在联系,鉴定甘草 miRNA、研究 miRNA 的生物学功能将促进对其相关分子机制的了解。

目前对于中草药等非模式植物的 miRNA 的鉴定主要依赖于转录组序列。转录组序列有两类来源:表达序列标签(EST)和高通量转录组测序数据。目前随着高通量测序费用的不断降低,对中草药植物进行的高通量转录组测序研究不断增加^[21],为 miRNA 的鉴定提供了数据基础。miRNA 鉴定的关

键是找到 miRNA 的茎环结构前体,因此获得高覆盖度的转录本数据是鉴定 miRNA 前体的前提。在 GenBank 中的甘草核酸数据中,既有 5 万余条来自 Sanger 测序的 EST 序列,也有若干个高通量测序数据。为了获得高质量的转录本序列数据,应用 Trinity 等常用拼接软件,将不同来源的转录本数据拼接在一起,拼接得到的转录本数据覆盖度较好,这一高质量的转录本数据将有助于后续的 miRNA 的生物信息学鉴定和靶基因预测研究。

在 miRNA 的诸多鉴定方法中,生物信息学预测无须进行测序,简便快捷,因而得到较多的应用^[15,17,22-23]。在这些研究中,主要采用 2 种方法进行 miRNA 的生物信息学鉴定。第 1 种鉴定方法是,将相关物种的 miRNA 成熟形式比对到转录本上,允许 3~4 个碱基错配,然后取一定长度的侧翼序列(如 200 nt),用 mfold 等 RNA 折叠软件进行分析^[17]。第 2 种方法是本研究采用的方法^[15]。这种做法与第一种方法的主要区别在于,比对到转录本序列上的是相关物种的已知 miRNA 的茎环前体序列,而不是成熟形式的序列。这 2 种方法各有特点,也都有

一定的应用。本研究在预试验中比较了这 2 种鉴定方法,发现第 2 种方法更方便易行,人工验证部分工作量相对较少,因而采用了这个方法。

总之,本研究应用生物信息学手段,在转录本序列拼接的基础上,预测了甘草 miRNA,确定了一批甘草 miRNA 的靶基因,为认识甘草 miRNA 的生物学功能提供了初步数据。

参考文献

- [1] 刘洋洋,刘春生,曾斌芳,等.甘草种质资源研究进展[J].中草药,2013,44(24):3593-3598.
- [2] 李学斌,陈林.中国甘草资源的生态分布及其繁殖技术研究[J].生态环境学报,2013,22(4):718-722.
- [3] Lee R C, Feinbaum R L, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14* [J]. *Cell*, 1993, 75(5): 843-854.
- [4] Reinhart B J, Slack F J, Basson M, et al. The 21 nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans* [J]. *Nature*, 2000, 403(6772): 901-906.
- [5] Qin Z, Li C, Mao L, et al. Novel insights from non-conserved microRNAs in plants [J]. *Front Plant Sci*, 2014, 5: 586-589.
- [6] Budak H, Akpinar B A. Plant miRNAs: biogenesis, organization and origins [J]. *Funct Integr Genomics*, 2015, DOI10.1007/s10142-015-0451-2.
- [7] Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data [J]. *Nucl Acids Res*, 2014, 42: D68-D73.
- [8] 曾奇峰,曾光尧,周应军. MicroRNA 与药用植物 [J]. 中国生物化学与分子生物学报, 2015, 31(4): 346-351.
- [9] 周芳名,白志川,卢善发. 药用植物 microRNA [J]. 中草药, 2013, 44(2): 232-237.
- [10] Bolger A M, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data [J]. *Bioinformatics*, 2014, 30(15): 2114-2120.
- [11] Cox M P, Peterson D A, Biggs P J. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data [J]. *BMC Bioinformatics*, 2010, 11: 485-489.
- [12] Grabherr M G, Haas B J, Yassour M, et al. Full-length transcriptome assembly from RNA-seq data without a reference genome [J]. *Nat Biotechnol*, 2011, 29(7): 644-652.
- [13] Zheng Y, Zhao L, Gao J, et al. iAssembler: a package for de novo assembly of Roche-454/Sanger transcriptome sequences [J]. *BMC Bioinformatics*, 2011, 12: 453-456.
- [14] Fu L M, Niu B F, Zhu Z W, et al. CD-HIT: accelerated for clustering the next generation sequencing data [J]. *Bioinformatics*, 2012, 28(23): 3150-3152.
- [15] Khan B M Y, Irfan M, Yousaf R, et al. Identification of micro-RNAs in cotton [J]. *Plant Physiol Biochem*, 2008, 46(8/9): 739-751.
- [16] Zuker M. Mfold web server for nucleic acid folding and hybridization prediction [J]. *Nucl Acids Res*, 2003, 31(13): 3406-3415.
- [17] Zhang B H, Pan X P, WU Q L, et al. Identification and characterization of new plant microRNAs using EST analysis [J]. *Cell Res*, 2005, 15(5): 336-360.
- [18] Ramilowski J A, Sawai S, Seki H, et al. *Glycyrrhiza uralensis* transcriptome landscape and study of phytochemicals [J]. *Plant Cell Physiol*, 2013, 54(5): 697-710.
- [19] Li Y, Luo H M, Sun C, et al. EST analysis reveals putative genes involved in glycyrrhizin biosynthesis [J]. *BMC Genomics*, 2010, 11: 268-275.
- [20] Voinnet O. Post-transcriptional RNA silencing in plant-microbe interactions: A touch of robustness and versatility [J]. *Curr Opin Plant Biol*, 2008, 11(4): 464-470.
- [21] 赵春丽,李先恩,都晓伟,等.地黄 microRNAs 和靶基因的生物信息学预测及验证 [J]. 中草药, 2014, 45(8): 1129-1135.
- [22] 李崇奇,沈文涛,言普,等.巨桉 miRNA 及其靶基因生物信息学预测 [J]. 南方农业学报, 2014, 45(9): 1532-1538.
- [23] 夏薇,李静,周丽霞.椰子保守 microRNA 预测和特征分析 [J]. 广东农业科学, 2014, 41(14): 130-135.