



# 铁皮石斛组培苗和栽培苗的 circRNA 检测及分析

黄路瑶, 郁吉锋, 钱央央, 徐 涛

(浙江理工大学生命科学与医药学院, 杭州 310018)

**摘要:** 为了分析铁皮石斛组培苗与有菌根的栽培苗的 circular RNA (circRNA) 变化, 采用高通量测序方法对铁皮石斛组培苗和栽培苗进行 circRNA 检测, 并分析它们之间的差异。结果显示: 两组 4 个样品共测出 4115 组不同的 circRNA, 其中有显著性差异的有 27 种, 在实验组中 17 种为显著上调, 10 种为显著下调; GO 富集分析一共检测出 99 组数据, 其中显著性基因最多是核基因; KEGG 富集分析表明, circRNA 共有富集在 7 条 KEGG 通路上, 其中丁酸甲酯代谢途径(Butanoate metabolism)和错配修复途径(Mismatch repair)具有显著性差异。以上结果可为进一步研究铁皮石斛形成菌根前后 circRNA 的变化提供参考。

**关键词:** 铁皮石斛; 组培苗; 栽培苗; circRNA; 差异性表达

中图分类号: Q94-3

文献标志码: A

文章编号: 1673-3851 (2020) 11-0840-06

## Detection and analysis of circRNA in tissue cultured seedling and cultivated seedling of *Dendrobium officinale*

HUANG Luyao, YU Jifeng, QIAN Yangyang, XU Tao

(College of Life Sciences and Medicine, Zhejiang Sci-tech University, Hangzhou 310018, China)

**Abstract:** In order to analyze the changes of circular RNA (circRNA) of *Dendrobium officinale* between the tissue cultured seedlings and the cultivated seedlings with mycorrhiza, the experiment was conducted to detect the circRNA of the tissue cultured seedlings and cultivated seedlings of *D. officinale* by high-throughput sequencing method, and the differences between them were analyzed. The results showed that 4115 groups of different circRNAs were detected in the four samples, of which 27 were significantly different. In the experimental group, 17 were significantly up-regulated and 10 were significantly down-regulated. A total of 99 sets of data were detected by GO enrichment analysis, of which the most significant genes were nucleus. KEGG enrichment analysis found that differential circRNAs were commonly enriched in 7 KEGG pathways and Butanoate metabolism and Mismatch repair had significant differences. The above results can provide a reference for further research on the changes of circRNA before and after the formation of mycorrhiza in *D. officinale*.

**Key words:** *Dendrobium officinale*; tissue culture seedling; cultivation seedlings; circRNA; differential expression

## 0 引 言

Circle RNA (circRNA) 是一类在 5' 和 3' 末端通过共价键连接的单链闭合 RNA 分子, 目前从哺

乳动物<sup>[1-2]</sup>、蠕虫<sup>[3]</sup>、昆虫<sup>[4-5]</sup>、鱼类<sup>[6]</sup>、植物<sup>[7]</sup> 及真菌<sup>[8]</sup> 中都检测到了 circRNA, 在真核生物中具有普遍表达性。作为内源性 ncRNA 家族的重要成员, circRNA 与多种生物学过程有关, 可以调节转录, 调

收稿日期: 2020-06-28 网络出版日期: 2020-09-03

基金项目: 浙江省重点研发计划项目(2015C02030)

作者简介: 黄路瑶(1995—), 女, 江西宜春人, 硕士研究生, 主要从事天然药用植物资源开发方面的研究。

通信作者: 徐 涛, E-mail: xutao1998cn@163.com

节可变剪接并与 miRNA 或蛋白质相互作用<sup>[9-10]</sup>。circRNA 在植物中表现为特定的细胞类型、组织或发育阶段的表达模式,并且在非生物胁迫和生物胁迫下,circRNA 都会产生差异性表达<sup>[11]</sup>。在非生物胁迫条件下,Ye 等<sup>[12]</sup>在缺磷培养的水稻根部样品中检测到 27 个 circRNA 差异表达,Zuo 等<sup>[13-14]</sup>在冷藏甜椒和冷藏番茄果实中分别鉴定出 36 个和 163 个,Gao 等<sup>[15]</sup>在低温胁迫下的葡萄叶片中鉴定出 475 个,Wang 等<sup>[16]</sup>在脱水胁迫下的小麦中检测到了 62 个。在病原体侵染植物期间,circRNA 也有显著的差异性表达,目前已从被病原体侵染的拟南芥叶片<sup>[17]</sup>、番茄<sup>[18]</sup>、棉花<sup>[19]</sup>和玉米<sup>[20]</sup>检测出显著差异表达的 circRNA。然而,在能与真菌结合形成菌根的兰科石斛属植物中有关 circRNA 差异表达未见报道。

铁皮石斛 (*Dendrobium officinale* Kimura et Migo) 是一种药用价值很高的兰科植物,一般利用人工种植栽培;作为一种附生兰,在自然条件下,铁皮石斛一生都会与真菌共生,并形成独特的菌根结构<sup>[21]</sup>。铁皮石斛需要与真菌共生形成菌根才能获取营养,而 circRNA 在植物根系生长发育中起着重要作用<sup>[22]</sup>。本文通过高通量测序与生物信息学分析手段,对铁皮石斛栽培苗(有菌根)及无菌组培苗进行 circRNA 检测,以期对铁皮石斛与真菌共生的机制提供依据,也为铁皮石斛促生长营养菌剂的开发提供参考。

## 1 材料与方法

### 1.1 实验材料

选取同批次组织培养的一年生铁皮石斛苗在种植基地进行栽培,待其形成菌根并可稳定存活后,将形成菌根的栽培苗及同批组培苗取茎部作为实验样品,以无菌组培苗为对照组,以有菌根的栽培苗为实验组进行测序。

### 2.2 参考基因的线性分析

#### 2.2.1 将读数匹配到基因组上

用 FASTQC 去除低质量的读数后,4 个样品的总有效读数在 6960 万~8200 万之间,结果如表 2 所示。由表 2 可知,用 tophat 和 tophat-fusion 将其匹配到铁皮石斛基因组上,其中 4 个样品的基因匹

## 1.2 实验方法

### 1.2.1 样品中 rRNA 去除

使用 Trizol 试剂(Invitrogen,CA,美国)提取总 RNA。使用 RIN 数大于 7.0 的 Bioanalyzer 2100 和 RNA 6000 Nano LabChip 试剂盒(安吉伦,美国)对总 RNA 数量和纯度进行分析。使用约 10  $\mu$ g 的总 RNA 消除核糖体 RNA,纯化后,用 RNase R 消化 1 h,在高温下使用二价阳离子将 RNA 片段破碎成小片,并按照 mRNA-Seq 样品制备试剂盒(Illumina,圣地亚哥,美国)的操作规程,将裂解的 RNA 片段反转录以创建最终的 cDNA 文库,并插入片段平均大小为 300 bp( $\pm$ 50 bp)配对末端文库,在 Illumina Hiseq 4000 上进行了配对末端测序。

### 1.2.2 生物信息学分析

在 Illumina Hiseq 4000 测得原始数据后,用软件 Cutadapt 和 FASTQC 去除低质量的读数(Reads)和接头,通过软件 Tophat 和 Tophat-fusion 将数据定位到基因组上,并用软件 CIRCexplorer 识别每一个样本中的 circRNA,合并成为一个转录表达谱,用 Perl statistics psckage 鉴定 circRNA 的差异表达。

## 2 结果与分析

### 2.1 FASTQC 质量检测

本次实验以组培苗为对照组,以栽培苗为实验组,共对 4 个样品进行测序。测序的原始数据用 FASTQC 质检(见表 1)。表 1 表明,在本次测序中,样品原始数据中总读数在 9440 万~9980 万之间,有效读数率均在 98.01%以上,Q20 均在 99.84%以上,Q30 均在 96.80%以上,且样品的 GC 含量在 41.0%~42.0%之间,GC 含量正常,以上数据表明本次测序结果质量很好,可用于后续分析。

表 1 FASTQC 质检结果

样品	总读数/万	碱基数/G	有效读数/万	碱基数/G	有效读数率/%	Q20/%	Q30/%	GC 含量/%
对照组 1	9981	14.97	9824	14.74	98.43	99.84	96.92	41.0
对照组 2	9753	14.63	9558	14.34	98.01	99.85	97.22	41.5
实验组 1	9941	14.91	9773	14.66	98.31	99.85	96.84	42.0
实验组 2	9447	14.17	9281	13.92	98.25	99.88	96.83	42.0

配率为 70.43%~73.82%,匹配到单基因读数的比率为 46.21%~48.97%,匹配到融合基因的读数比率为 23.53%~25.49%,匹配到单基因的读数约为融合基因的 2 倍;另外,4 个样品匹配到 DNA 正链与负链的比率均为 33.84%~35.74%,无显著差异。

表2 匹配基因结果

样品	总有效读数/万	匹配到基因的 读数比率/%	匹配到单基因 读数比率/%	匹配到融合基 因读数比率/%	匹配到 DNA 正 链比率/%	匹配到 DNA 负链 比率/%
对照组 1	6960	73.82	48.97	24.85	35.48	35.74
对照组 2	7595	72.34	48.81	23.53	34.73	35.07
实验组 1	8200	70.43	46.21	24.22	33.84	34.05
实验组 2	8102	72.71	47.21	25.49	35.10	35.12

## 2.2.2 对匹配基因读数的来源分析

基因来源一般分为外显子、内含子和基因间区,对匹配到单基因读数(Unique mapped reads, UMR)和匹配到融合基因的读数(Multi mapped reads, MMR)进一步比对分析,发现两组样品的 UMR 和 MMR 在基因来源方面具有差异性,结果如表 3 和表 4 所示。表 3 和表 4 显示,其中 UMR 来源外显子区域的比例为 58.08%~63.68%,MMR 的比例为 56.43%~64.37%,在外显子来源中,UMR 与 MMR 无明显差距,但对照组都略高于实验组;UMR 来源内含子比例为 27.21%~31.20%,MMR 的比例为 20.83%~25.22%,在内含子来源方面,UMR 高于 MMR,实验组略高于对照组;UMR 来源基因间隔区的比例为 8.40%~12.91%,MMR 的比例为 14.80%~18.35%,在基因间区来源方面,MMR 高于 UMR,对照组与实验组无明显差距。

表3 匹配到单基因读数的来源统计

样品	外显子/%	内含子/%	基因间区/%
对照组 1	63.68	27.93	8.40
对照组 2	63.31	27.21	9.48
实验组 1	58.35	31.20	10.45
实验组 2	58.08	29.00	12.91

表4 匹配到融合基因读数的来源统计

样品	外显子/%	内含子/%	基因间区/%
对照组 1	61.78	21.05	17.17
对照组 2	64.37	20.83	14.80
实验组 1	56.43	25.22	18.35
实验组 2	59.65	24.05	16.29

## 2.3 circRNA 的显著性差异分析

本次高通量测序共测得 4115 组不同的环状 RNA,而在这 4115 组环状 RNA 中,有显著性差异的有 27 种 ( $|\log_2FC| > 1$ , 且  $p \leq 0.05$ ),且皆是外显子,其中:17 种为显著上调(见表 5),10 种为显著下调(见表 6)。由表 5 和表 6 可知,在显著上调的 17 种 circRNA,有 6 种转录自 DNA 正链,其余都转录自 DNA 负链,而其中外显子个数最多的是 circRNA3008,拥有 7 个外显子,其次是 circRNA2827,拥有 5 个外显子;而 10 种显著下调的 circRNA 除 circRNA1270 是转录自 DNA 正链外,其他 9 个都是转录自 DNA 负链,其中拥有外显子个数最多的是 circRNA776,拥有 7 个外显子,其次是 circRNA1379,拥有 4 个外显子。

表5 显著上调的 circRNA

circRNA 编码	染色体编号	起始位点	终止位点	DNA 链	外显子个数	基因名称	log2FC	上调/下调
circRNA2567	NW_018262509.1	221292	221740	—	2	LOC110093560	inf	上调
circRNA2827	NW_018300769.1	96604	97410	+	5	LOC110109905	inf	上调
circRNA2716	NW_018321052.1	381197	381461	—	2	LOC110093180	inf	上调
circRNA2836	NW_018258747.1	154057	162739	—	2	LOC110107129	inf	上调
circRNA2940	NW_018290372.1	101028	101490	+	2	LOC110105758	inf	上调
circRNA945	NW_018274601.1	220584	220923	—	2	LOC110099297	5.17	上调
circRNA2521	NW_018270372.1	49730	50285	—	3	LOC110098101	inf	上调
circRNA2709	NW_018256747.1	477408	489271	—	2	LOC110102630	inf	上调
circRNA2912	NW_018298246.1	28552	28936	—	2	LOC110109647	inf	上调
circRNA2756	NW_018318886.1	127529	128276	+	4	LOC110092112	inf	上调
circRNA1301	NW_018282725.1	146594	160369	—	3	LOC110102763	4.25	上调
circRNA120	NW_018320961.1	91686	91977	+	2	LOC110092904	2.67	上调
circRNA3195	NW_018315962.1	2329	3256	+	2	LOC110116255	inf	上调
circRNA2646	NW_018278168.1	243952	244172	+	1	LOC110100946	inf	上调
circRNA676	NW_018277039.1	470362	470672	—	2	LOC110100400	3.74	上调
circRNA3008	NW_018274601.1	442087	444118	—	7	LOC110099299	inf	上调
circRNA3083	NW_018306256.1	590973	591346	+	2	LOC110112620	inf	上调

注:log2FC 即为折叠倍数(fold change)的对数,该值为每组样品所测的 circRNA 的量均值比(实验组比对照组);inf 代表无穷大。下同。

表 6 显著下调的 circRNA								
circRNA 编码	染色体编号	起始位点	终止位点	DNA 链	外显子个数	基因名称	log2FC	上调/下调
circRNA1379	NW_018260064.1	546169	553467	—	4	LOC110110106	-inf	下调
circRNA1178	NW_018254202.1	766428	766912	—	3	LOC110096707	-inf	下调
circRNA719	NW_018280475.1	413976	414442	—	2	LOC110101691	-inf	下调
circRNA1175	NW_018292777.1	105123	106217	—	1	LOC110106909	-inf	下调
circRNA298	NW_018273632.1	54585	55193	—	2	LOC110099184	-inf	下调
circRNA460	NW_018260106.1	62568	63466	—	3	LOC110111746	-inf	下调
circRNA351	NW_018279587.1	428084	428336	—	2	LOC110101406	-inf	下调
circRNA1060	NW_018256669.1	3206	3665	—	2	LOC110099217	-inf	下调
circRNA776	NW_018274755.1	433365	434992	—	7	LOC110099829	-inf	下调
circRNA1270	NW_018295038.1	378543	378812	+	2	LOC110108036	-inf	下调

2.4 差异 circRNA 聚类热图分析

各组样品的差异 circRNA 聚类热图如图 1 所示，图中颜色代表 circRNA 在该样品中检测的数目。图 1 表明，在 27 组具有显著差异的 circRNA 中，上调的 circRNA2709 和 circRNA2046 在两个实验组中表达量差距较大，下调的 circRNA1178 在两个对照组中的表达量差距较大。

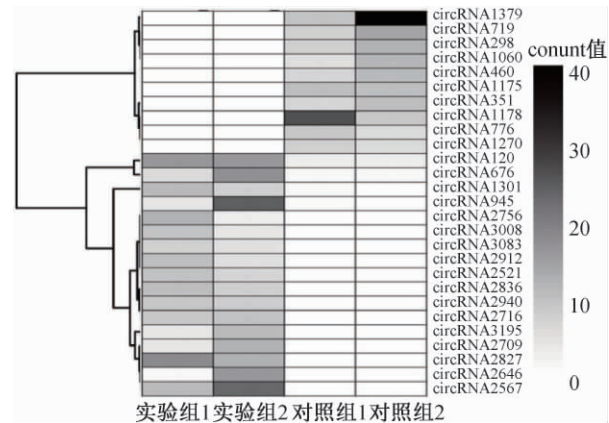


图 1 差异 circRNA 聚类热图

2.5 差异 circRNA 火山图分析

以 log2FC 为横坐标，以  $t$  检验显著性检验  $-\log_{10}(p)$  为纵坐标，以所测得的 circRNA 为原始数据，得到的差异 circRNA 火山图(图 2)。在图 2 中， $\log_2FC > 1.0$  部分代表在实验组中上调， $\log_2FC < -1.0$  为下调， $-\log_{10}(p) > 1.0$  可认为具有一定的差异性， $-\log_{10}(p) > 1.3$  为具有显著的差异性。其中有部分只在实验组检测出的 circRNA 或只在对照组检测出的 circRNA，因为换算成的 log2FC 值趋近正负无穷大而未能在图中显示出来(这部分可参照表 5 和表 6)。从图 2 中可以看出，在此次所测得的两组样品中，circRNA 的显著性差绝大部分在 0.5 以下，少部分在 0.5~1.3 之间，极少部分在 1.3 以上，而在后两个区间中，在实验组中上调的 circRNA 数量都比下调的要高。

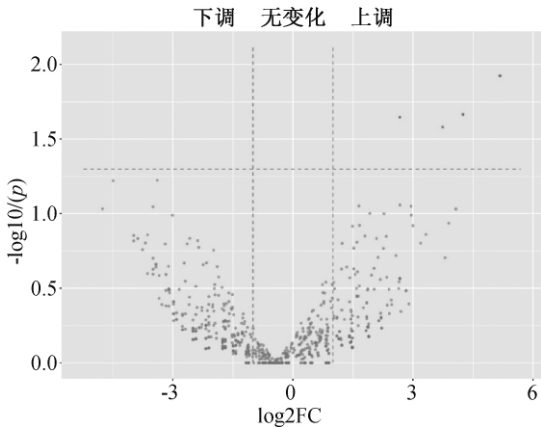


图 2 差异 circRNA 火山图

2.6 差异 circRNA 的 GO 富集分析

GO 富集分析条目由三个部分组成，分别是生物学过程 (Biological process, BP)，细胞组成 (Cellular component, CC) 和分子功能 (Molecular function, MF)<sup>[23]</sup>，在本次 GO 富集分析中，一共检测出 99 组数据，其中 BP 部分 36 组，CC 部分 27 组，MF 部分 36 组。在以上数据中，注释为特定 GO 并具有显著性差异的基因数(S gene number)在 BP 组最大值为 2，分别是蛋白磷酸化基因 (Protein phosphorylation) 和生物过程基因 (Biological\_process)；CC 最大值为 7，即核基因 (Nucleus)；MF 最大值为 4，即 ATP 结合基因 (ATP binding)，具体如图 3 所示。

2.7 差异 circRNA 的 KEGG 富集分析

通过对 KEGG 富集分析(表 7 和图 4)，发现显著差异的 circRNA 共富集在 7 条 KEGG 通路，7 条 KEGG 通路中，有 6 条为上调 circRNA 的 KEGG 富集通路，1 条为下调 circRNA 的 KEGG 富集通路。其中丁酸甲酯代谢途径 (Butanoate metabolism) 和错配修复途径 (Mismatch repair) 具有显著性差异，而磷酸肌醇代谢途径 (Inositol phosphate metabolism) 和磷脂酰肌醇信号系统 (Phosphatidylinositol signaling system) 差异性较小。

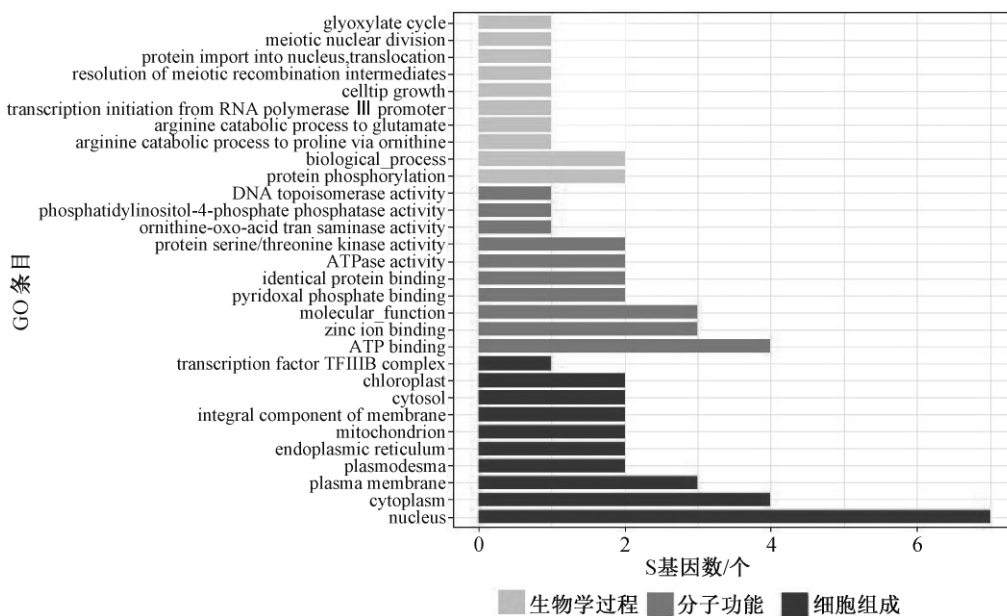


图3 差异 circRNA 的 GO 富集图

表7 差异 circRNA 的 KEGG 富集分析表

通路 ID	通路名称	circRNA 编码	上调/下调	p 值
ko00650	丁酸甲酯代谢途径	circRNA2827	上调	0.04
ko03430	错配修复途径	circRNA1301	上调	0.05
ko00250	丙氨酸,天门冬氨酸和谷氨酸代谢途径	circRNA2827	上调	0.06
ko03440	同源重组途径	circRNA120	上调	0.09
ko00330	精氨酸和脯氨酸代谢途径	circRNA1379	下调	0.09
ko00562	磷酸肌醇代谢途径	circRNA676	上调	0.12
ko04070	磷脂酰肌醇信号系统	circRNA676	上调	0.13

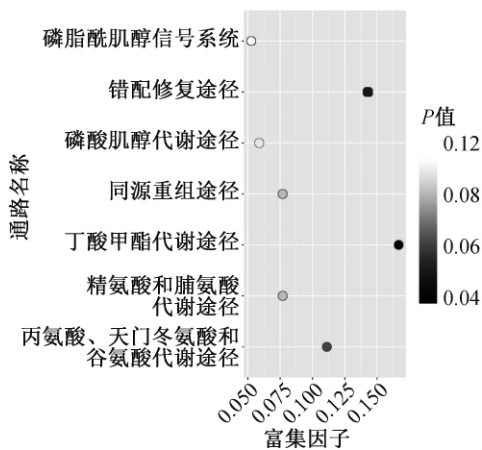


图4 KEGG 通路的富集统计

### 3 结论

本文通过高通量测序,对铁皮石斛有菌栽培苗和无菌组培苗的测序结果进行分析,得出以下主要结论:

a) 在本次测序中,样品原始数据总读数在 9440 万~9980 万之间,去除低质量读数后,总有效读数

在 6960 万~8200 万之间,栽培苗的总有效读数大于组培苗。

b) 在铁皮石斛中匹配到单基因(UMR)的比率约为匹配到融合基因(MMR)的两倍,基因主要来源外显子,其次为内含子,基因间区最少;对于 UMR 和 MMR 的外显子来源,组培苗都略高于栽培苗,对于 UMR 和 MMR 内含子来源,组培苗都略低于栽培苗。

c) 4 个样品中共检测出 4115 组不同的 circRNA,有显著性差异的 circRNA 有 27 种,其中在栽培苗中 17 种为显著上调,10 种为显著下调;对差异 circRNA 进行 GO 富集分析,发现核基因的显著性差异基因数最多;对差异 circRNA 进行 KEGG 富集分析,发现丁酸甲酯代谢途径和错配修复途径具有显著性差异。

### 参考文献:

- [1] Dong R, Ma X K, Chen L L, et al. Increased complexity of circRNA expression during species evolution[J]. RNA Biology, 2017, 14(8):1064-1074.

- [2] Liu Y Q, Zhang X, Liu M, et al. Direct detection of circRNA in real samples using reverse transcription-rolling circle amplification[J]. *Analytica Chimica Acta*, 2020, 1101: 169-175.
- [3] George A K, Master K, Majumder A, et al. Circular RNAs constitute an inherent gene regulatory axis in the mammalian eye and brain [J]. *Canadian Journal of Physiology And Pharmacology*, 2019, 97(6):463-472.
- [4] Hu X L, Zhu M, Zhang X, et al. Identification and characterization of circular RNAs in the silkworm midgut following *Bombyx mori* cytoplasmic polyhedrosis virus infection[J]. *RNA Biology*, 2018, 15(2):292-301.
- [5] Guo R, Chen D F, Chen H Z, et al. Systematic investigation of circular RNAs in *Ascosphaera apis*, a fungal pathogen of honeybee larvae[J]. *Gene*, 2018, 678: 17-22.
- [6] He L B, Zhang A D, Xiong L, et al. Deep circular RNA sequencing Provides Insights into the Mechanism Underlying Grass Carp Reovirus Infection [J]. *International Journal of Molecular Sciences*, 2017, 18(9):1977-1993.
- [7] Zhang P J, Meng X W, Chen H J, et al. PlantCircNet: A database for plant circRNA-miRNA-mRNA regulatory networks [J]. *Database-The Journal of Biological Databases and Curation*, 2017, 11: 1-8.
- [8] Huo L, Zhang P, Li C X, et al. Genome-wide identification of circRNAs in pathogenic basidiomycetous yeast *Cryptococcus neoformans* suggests conserved circRNA host genes over kingdoms[J]. *Genes*, 2018, 9(3):118-131.
- [9] 刘致新, 孙文洲, 李佳欣, 等. circRNA 的生物学功能及其在卵巢癌中的研究进展[J]. *癌症进展*, 2020, 18(4):347-350.
- [10] 张晓云, 孔鹏, 窦永青, 等. circRNA 肿瘤相关生物学功能的研究进展[J]. *生命的化学*, 2020, 40(4): 607-613.
- [11] 高振. 葡萄 circRNA 鉴定分析及 Vv-circATS1 的抗冷功能研究[D]. 上海:上海交通大学, 2019, 8-11.
- [12] Ye C Y, Chen L, Liu C, et al. Widespread noncoding circular RNAs in plants[J]. *New Phytologist*, 2015, 208(1):88-95.
- [13] Zuo J H, Wang Y X, Zhu B Z, et al. Analysis of the coding and non-coding RNA transcriptomes in response to bell pepper chilling [J]. *International Journal of Molecular Sciences*, 2018, 19(7):125-135.
- [14] Zuo J H, Wang Y X, Zhu B Z, et al. Deciphering the roles of circRNAs on chilling injury in tomato [J]. *Biochemical And Biophysical Research Communications*, 2016, 479(2):132-138.
- [15] Gao Z, Li J, Luo M, et al. Characterization and cloning of grape circular RNAs identified the cold resistance-related vv-circATS1 [J]. *Plant Physiology*, 2019, 180(2):966-985.
- [16] Wang Y X, Yang M, Wei S M, et al. Identification of circular RNAs and their targets in leaves of *Triticum aestivum* L. under dehydration stress[J]. *Front Plant Science*, 2017, 1(5):2024-2032.
- [17] 王彦如. 在拟南芥中检测和验证环形 RNA 及其在植物与病原物互作中的作用研究[D]. 南京:南京农业大学, 2016, 13-26.
- [18] 张敏. 番茄晚疫病密切相关 circRNA 及其功能初探 [D]. 大连:大连理工大学, 2018, 17-21.
- [19] Xiang L X, Cai C W, Cheng J R, et al. Identification of circularRNAs and their targets in *Gossypium* under *Verticillium* wilt stress based on RNA-seq[J]. *Peerj*, 2018, 6(3): e4500.
- [20] 王猛. 玉米小斑病菌 O 小种分化鉴定与非编码 RNA 效应物研究[D]. 上海:上海交通大学, 2018, 75-100.
- [21] 杨建文, 凌鸿, 张盈. 内生真菌对兰科药用植物种子萌发作用研究进展[J]. *菌物学报*, 2018. 37(1):22-34.
- [22] Liu S, Wu L, Qi H, et al. LncRNA/circRNA-miRNA-mRNA networks regulate the development of root and shoot meristems of *Populus* [J]. *Industrial Crops & Products*, 2019, 133:333-347.
- [23] 张艳俊, 杨毅清, 袁心悦, 等. 基于转录组测序对小麦抗叶锈性相关基因的筛选与分析[J]. *河北农业大学学报*, 2020, 43(1):11-16.

(责任编辑:唐志荣)