



小花草玉梅高通量转录组测序 与花发育基因的挖掘

戎利勤,李晓冬,刘虎岐*

(西北农林科技大学 生命科学学院,陕西杨陵 712100)

摘要:该研究采用高通量测序技术 Illumina HiSeq 2000,对小花草玉梅花器官进行转录组测序,挖掘参与其花发育相关的基因。测序结果得到 54 513 822 个序列读取片段(reads),包含 7 826 726 115 bp 的碱基序列信息。将测序数据进行序列组装后,获得 43 767 个单基因簇(unigenes),平均长度为 926 bp。转录物注释结果显示,28 130 条 unigenes 有同源比对信息。在 43 767 条 unigenes 中检测到 5 015 个 SSR 位点,且 SSR 不同重复基序类型中,出现频率最高的是 AG/CT,其次是 AAG/CTT 和 ACC/GGT。通过转录因子分析,进一步筛选得到了 12 个与花发育紧密相关的 MADS 基因,分别为 *FUL1*,*FUL2*,*AP3-1*,*AP3-2*,*AP3-3*,*PI1*,*PI2*,*AG1*,*AG2*,*SEP1*,*SEP3* 和 *AGL6*。实时荧光定量 PCR 分析表明,与小花草玉梅正常花相比,全白、绿白相间、五瓣变异、全绿变异花中的 *FUL1*、*SEP1*,*SEP3* 和 *AGL6* 基因均显著上调表达,而 12 个基因在极端变异花中的表达水平与正常花的差异均不明显。定量结果经主成分分析显示,*AGL6*,*SEP3*,*FUL1*,*PI2* 及 *SEP1* 的表达量均为小花草玉梅花形态建成的主要指标。研究结果在一定程度上丰富了小花草玉梅的基因信息,为后续研究其花器官变异的分子机制提供了基础数据。

关键词:小花草玉梅;转录组;高通量测序;花发育基因

中图分类号:Q785; Q789 文献标志码:A

Transcriptome Analysis of *Anemone rivularis* var. *flore-minore* Based on High-throughput Sequencing Technology and Flower Development Gene Screening

RONG Liqin, LI Xiaodong, LIU Huqi*

(College of Life Science, Northwest A&F University, Yangling, Shaanxi 712100, China)

Abstract: The transcriptome of *Anemone rivularis* var. *flore-minore* was sequenced by Illumina HiSeq 2000 platform to find flower development gene. The result indicated that 54 513 822 sequence reads containing 7 826 726 115 bp nucleotide sequence information were obtained. A total of 43 767 unigenes were formed by initial sequence splicing, with an average read length of 926 bp. Annotation analysis of unigene indicated that 28 130 unigenes had homolog in public protein database. Total 5 015 unigenes included SSRs and the highest frequency was AG/CT, followed by AAG/CTT and ACC/GGT. Through the analysis of transcription factors, 12 MIKC-type MADS genes closely related to flower development were selected, and they were *FUL1*,*FUL2*,*AP3-1*,*AP3-2*,*AP3-3*,*PI1*,*PI2*,*AG1*,*AG2*,*SEP1*,*SEP3* and *AGL6*. Quantitative PCR analysis showed that the expression of *FUL1*,*SEP1*,*SEP3* and *AGL6* were significant up regulate in white, green-white, five-flap and green plants than that in normal flower of *A. rivularis* var. *flore-*

收稿日期:2018-03-13;修改稿收到日期:2018-07-17

基金项目:国家中医药管理局 2017 年中医药公共卫生服务补助专项“全国中药资源普查项目”(财社〔2017〕66 号)

作者简介:戎利勤(1992—),女,在读硕士研究生,主要从事花发育基因研究。E-mail:1004415776@qq.com

* 通信作者:刘虎岐,博士,副教授,硕士生导师,主要从事植物发育生物学及分子生物学的研究。E-mail:liuhuqi@nwafu.edu.cn

minore, and the expression of 12 genes had no significant difference in normal and extreme variation plants. The expression of *AGL6*, *SEP3*, *FUL1*, *PI2* and *SEP1* were the main indexes of flower morphogenesis based on Principal Component Analysis (PCA). These results greatly enriched genetic information of *A. rivularis* var. *flore-minore* and provided basic data for the study of mechanism on diversity of floral organ variants in *A. rivularis* var. *flore-minore* in the future.

Key words: *Anemone rivularis* var. *flore-minore*; transcriptome; high-throughput sequencing technology; flower development gene

花是被子植物独有的观赏及生殖器官,花发育的研究对植株发育及演化过程有重要影响^[1]。在拟南芥、金鱼草等模式植物中已克隆到许多与花分生组织形成和花器官形成等相关的基因^[2]。Coen 等开创性地提出了花器官发育的 ABC 模型^[3]。但此模型是以核心真双子叶植物为基础的。毛茛科属于基部真双子叶植物^[4],花器官发育并不严格遵守基于核心真双子叶植物的 ABC 模型,该类群 MADS-box 基因调控范围更广,延展到相邻花器官,使花器官形态具有高度多样性^[5-6]。

小花草玉梅(*Anemone rivularis* var. *flore-minore*)是毛茛科银莲花属草玉梅的变种^[7]。小花草玉梅的花器官在野生条件下发生了变异,变异在花被片、雄蕊和雌蕊中均有体现^[8]。根据其花器官变异部位及程度将变异分为 5 类:全白、绿白相间、五瓣、全绿和极端变异^[9]。

Illumina HiSeq 2000 第二代高通量测序平台使用 PE90 技术,在测序速度以及通量方面都得到了优化^[10]。本研究通过高通量测序技术 Illumina HiSeq 2000,在小花草玉梅中进一步挖掘参与其花形态建成的重要基因。小花草玉梅作为基部真双子叶植物,转录组数据对其变异机制的研究奠定了基础,为人们更全面研究被子植物的系统发育与遗传演化提供基础数据。

1 材料和方法

1.1 试验材料

2016 年 7 月于陕西省陇县地区采集小花草玉梅植株的正常花器官和变异花器官,经液氮速冻后于 -80 °C 贮存备用。对正常花器官进行转录组测序分析。

1.2 方法

1.2.1 RNA 提取与测序文库的构建 按照 Trizol 法提取样品的总 RNA,用带有 Oligo(dT)的磁珠富集 mRNA。将 mRNA 打断成片段后逆转录成 dsDNA,在经过纯化洗脱之后做末端修复、加 polyA 并连接测序接头,制备测序文库,富集测序样本。

1.2.2 转录组测序 通过 Illumina HiSeq 2000 测序平台进行高通量测序。为得到较高质量的结果,原始数据中要去除带接头的,重复的以及测序质量很低的读序,再用 Trinity 软件进行转录组重头组装,先将具有一定长度重叠的 reads 连成更长的不含 N 的片段 Contig,之后将来自同一转录本的不同 Contig 连接,得到两端不能再延长的非冗余序列。

1.2.3 转录组数据分析 使用 BLAST 程序,将小花草玉梅 unigenes 与 NCBI 的非冗余蛋白数据库(Non-redundant protein database, NR),去冗余的蛋白序列数据库(SwissProt protein database, Swiss-Prot),蛋白质直系同源数据库(Cluster of Orthologous Groups, COG)和基因功能和代谢途径数据库(Kyoto encyclopedia of genes and genomes, KEGG)等数据库进行 Blastx 比对,得到与给定 unigene 具有最高序列相似性的蛋白,得到该 unigene 的注释信息。根据 NR 注释信息,使用 Blast2GO 软件得到基因的基因本体论数据库(Gene Ontology, GO)注释信息。

1.2.4 花发育基因的实时定量 PCR 试验 将转录组得到的序列与北京大学的转录因子数据库(Plant TFDB)进行比对分析,筛选出了 12 个与花发育相关的 MADS 基因。对 12 个基因在小花草玉梅正常花和 5 种变异花(图 1, B~F)中做定量反应。5 种

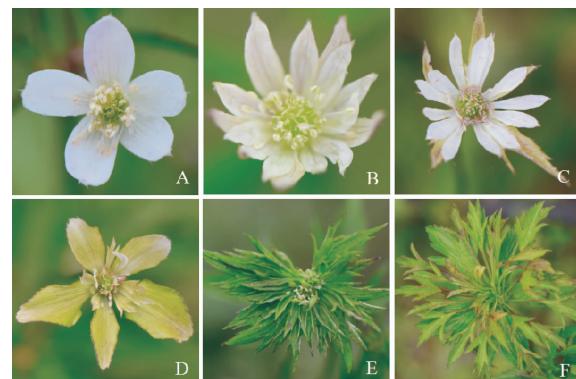


图 1 小花草玉梅正常花(A)和变异花(B~F)的形态

Fig. 1 Morphology of normal (A) and variant flowers

(B~F) in *Anemone rivularis* var. *flore-minore*

表 1 引物序列和产物长度

Table 1 Primer sequences and the length of the products

基因 Gene	正向引物 Forward primer(5'→3')	反向引物 Reverse primer(5'→3')	产物长度 The length of the product/bp
FUL1	AAGCATCCAGGGTGGCATGAG	CTTCGTCCCGCTGCAGG	121
FUL2	TGATTAAGCATCCAGGGAGGAATCAGAG	GTCACCAACTGTGAGGAACATGTTCAAATGG	180
AP3-1	CGATCCGCCAGCACCATC	GCCATTAACTACTCACACTGAAACAAACCAG	169
AP3-2	CCCGACATGCCACTATGGATTCA	AGCAAGGGTTAAACCATATGAGCT	151
AP3-3	CATGTTCTCTTGACCCGGCA	ATCCATCCCAGTCCTTCGCC	199
PI1	ATCGCATAAACAGGAACAGCAGATGG	CAAATTGGCTGGATTGGCTGGACT	136
PI2	GCACCTCCGAAGGATAGTCTGTAC	TGGACAATGAACCAAGCAGCAGGAAATGAA	97
AG1	CCAAATCTAGCAGCAACGGGAGAT	ACGGCCACGGCTAGAG	201
AG2	GACTTGTCTGACCCACAAACTGCT	GGTTTGAGTTCTGCAGAATTCAATCTGCTG	113
SEP1	CCTTGGAACATTCATAGGGTATGCATCTG	CTCCAAGCACCAAAGAACAGAAAAGATGT	205
SEP3	ACATTGGCTCTGGCCTCC	TCGCCAACCGTAGAAATGGC	184
AGL6	ATCTCCGCCGAAGGAGCGTC	CGATTGCAAGGTGGGCTC	189
β-actin	GAGCCCAGAGGTGCTCTT	AGCAATGCCAGGGAACATGG	165

变异花分别为全白变异(图 1,B),绿白相间变异(图 1,C),五瓣变异(图 1,D),全绿变异(图 1,E)和极端变异(图 1,F)。利用 Primer5 设计定量 PCR 引物(表 1),选择 β-actin 作为内参基因,使用相对定量 $2^{-\Delta\Delta Ct}$ 法分析。运用 SPSS 软件对基因表达量进行主成分分析,筛选出参与花形态建成的主要基因。

2 结果与分析

2.1 小花草玉梅转录组数据的组装

对小花草玉梅花器官转录组进行测序,共得到 54 513 822 个读序(reads),Q20 的百分率为 95.49%、Q30 的百分率为 88.90%、GC 含量百分比为 46.33%。以上结果表明,转录组测序数据质量可靠,可用于后续分析。序列组装后,最终得到了 43 767 条 unigenes,总长度为 40 565 399 bp,平均长度与 N50 分别为 926 和 1 431 bp。其中,大于 2 000 bp 的序列共有 4 280 条,占 unigenes 总数的 9.78%,说明测序质量较好。

2.2 小花草玉梅 unigene 的功能注释、分类和代谢途径分析

使用 BLAST 程序将测序得到的 unigenes 分别与 NR、Swissprot、COG、KEGG 数据库比对,结果(表 2)显示,在 NR 注释成功的 unigenes 总 unigenes 数的百分率最高(64.09%),在 KEGG 注释成功的 unigenes 占比则最低,为 26.44%。

对 4 个数据库的注释信息分析,共有 9 379 条

表 2 小花草玉梅 unigenes 的功能注释结果

Table 2 Summary of the annotation percentage of *A. rivularis* var. *flore-minore* as compared to public database

类别 Category	数量 Quantity	比例 Proportion/%
NR	28 051	64.09
SwissProt	21 101	48.21
COG	17 440	39.85
KEGG	11 574	26.44
得到注释的基因 Annotation genes	28 130	64.27
未得到注释的基因 Genes without annotation	15 637	35.73

unigenes 在所有数据库中同时标注成功,占 unigenes 总数的 21.43%。在所有数据中,只在某一个数据库中注释成功的 unigenes 为 5 105 条,比例为 11.66%。在以上 4 个数据库中至少 1 个数据库注释成功的 unigenes 有 28 130 条。用相似序列匹配得到的近缘物种中,莲花(*Nelumbo nucifera*)所占比例最高(28.34%),其次是葡萄(*Vitis vinifera*,8.53%),可可(*Theobroma cacao*,5.41%)。

2.3 小花草玉梅的 SSR 分析

利用 MISA 软件在小花草玉梅花器官的 unigenes 中检测到 5 015 个 SSR 位点,占 unigenes 总序列的 11.46%。SSR 的类型丰富,其中,三核苷酸重复所占比例最高,达到了 49.77%;五核苷酸重复

所占比例最低,为3.05%;二核苷酸重复、四核苷酸重复和六核苷酸重复所占比例分别为27.32%、6.66%和13.20%。在搜索到的SSR中,出现频率最高的5类基序为:AG/CT(20.7%)、AAG/CTT(15.8%)、ACC/GGT(8.4%)、ATC/ATG(7%)、AGC/CTG(5.1%)。上述SSR特征的分析,有助于开展小花草玉梅花器官的通用性标记开发等研究。

2.4 小花草玉梅花发育基因的定量实验

采用qPCR方法,以小花草玉梅正常花为对照,设定基因的表达量为1,得出各基因的相对表达量,分别对12个MIKC型MADS-box基因在小花草玉梅正常花和变异花中的表达量水平进行了研究,结果(图2)表明,与小花草玉梅正常花相比,全白、绿白相间、五瓣变异、全绿变异花中的 $FUL1$ 、 $SEP1$ 、 $SEP3$ 和 $AGL6$ 基因均显著上调表达,而12个基因在极端变异花中的表达水平与正常花的差异均不明显。

对12个花发育基因在正常花和5种变异花中的表达量进行主成分分析。按特征值 >1 的原则,提取了3个主成分,贡献率分别为41.54%、33.62%、13.45%,累计方差贡献率达88.61%,涵盖了大部分信息。由主成分的初始因子载荷矩阵(表3)可以看出, $AGL6$ 、 $SEP3$ 、 $FUL1$ 、 $PI2$ 及 $SEP1$ 的表达量与第1主成分呈显著正相关, $AG2$ 、 $AP3-1$ 、 $AG1$ 、 $AP3-3$ 、 $PI1$ 的表达量与第2主成分呈显著正相关。主成分分析表明,与第1主成分呈显

著正相关的指标,即 $AGL6$ 、 $SEP3$ 、 $FUL1$ 、 $PI2$ 及 $SEP1$ 的表达量均可作为小花草玉梅花形态发育的主要指标。

图3是小花草玉梅的12个花发育MADS基因的表达量经主成分分析后的前3个主成分构建的三维空间。分析的样本有6个,分别为小花草玉梅正常花和5种变异花。具有不同花形态的小花草玉梅分布在三维空间中不同的位置。6个样品表型不同,各自起主要调控作用的基因也不相同,因此它们位于空间中的不同位置。

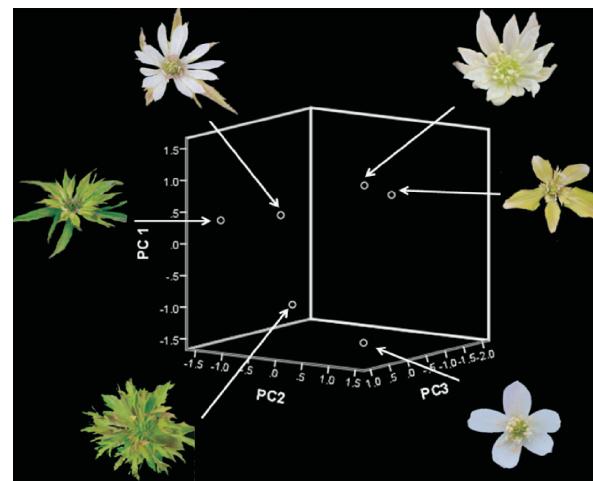


图3 不同形态的小花草玉梅在MADS基因三维空间中的分布位置

Fig. 3 *A. rivularis* var. *flore-minore* species placed in 3D space of MADS gene

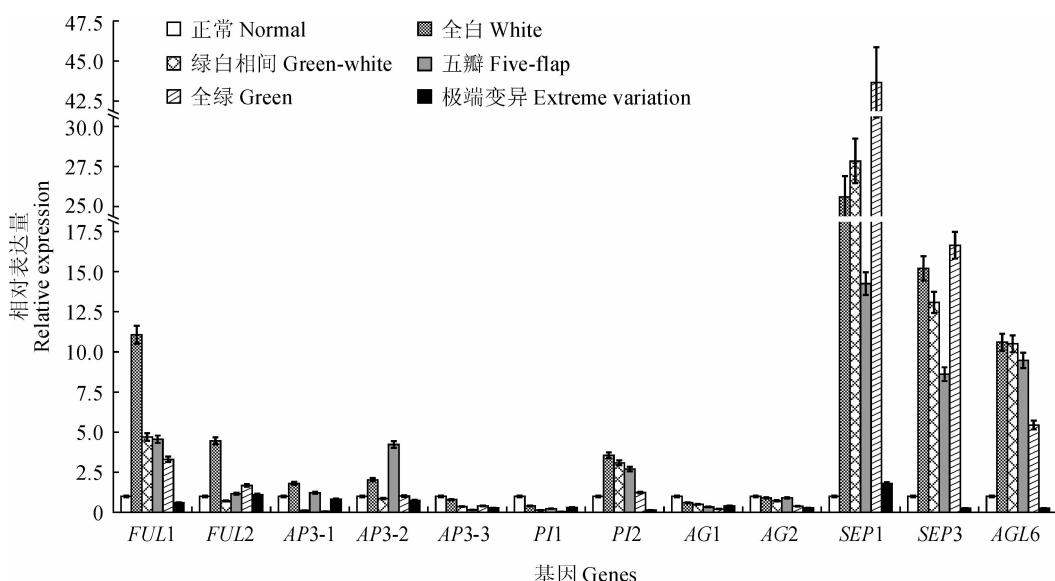


图2 12个MIKC型MADS-box基因在小花草玉梅正常花和变异花中的相对表达量

Fig. 2 Relative expression level of 12 MIKC-type MADS genes in normal and variant flowers in *A. rivularis* var. *flore-minore*

表3 花发育基因表达量的初始因子载荷矩阵及主成分贡献率

Table 3 Factor matrix and principal component contribution rate of the expression of flower developmental genes

指标 Index	主成分 1 Principal component1	主成分 2 Principal component2	主成分 3 Principal component3
FUL1	0.86	0.45	0.13
FUL2	0.57	0.51	0.30
AP3-1	0.14	0.80	-0.32
AP3-2	0.45	0.26	-0.80
AP3-3	-0.24	0.77	0.57
PI1	-0.63	0.76	0.10
PI2	0.85	0.39	-0.08
AG1	-0.50	0.79	0.23
AG2	0.18	0.85	-0.19
SEP1	0.75	-0.40	0.45
SEP3	0.90	-0.17	0.35
AGL6	0.93	0.13	-0.12
特征值 Eigenvalue	4.985	4.034	1.614
贡献率 Contribution rate/%	41.538	33.620	13.446
累计贡献率 Cumulative contribution rate/%	41.538	75.158	88.604

3 讨 论

Illumina 高通量测序技术的测序数据量大、效率高且成本低^[11]。本研究共得到 43 767 条 unigenes, 平均长度为 926 bp, 高于薏苡幼苗叶片 737.85 bp^[12]、半夏珠芽 751 bp^[13], 梁山慈竹 857.89 bp^[14], 说明小花草玉梅序列组装效果较好。N50 值为 1 431 bp, N50 值越大说明组装得到的长片段就越多, 组装效果就越好^[15]; Q20 的百分率为 95.49%、Q30 的百分率为 88.90%、高于它们的规定限值(>80%), 以上研究结果表明, 本研究采用的双端测序的方法, 增加了测序深度, 且提高了拼接的效率和准确性^[16]。此次测序质量可靠, 可以满足转录组分析的基本要求。

将测序结果分别与 NR、Swissprot、KOG、KEGG 数据库比对, 得到每个数据库注释的 unigenes 占总 unigenes 分别为 64.09%、48.21%、39.85% 和 26.44%。在以上 4 个数据库中至少在一个数据库注释成功的 unigenes 占比为 64.27%。有 35.73% 的 unigenes 未得到注释, 一方面可能是因为 unigenes 序列片段长度过短, 或是非编码序列, 难以进行同源性比对; 另一方面, 小花草玉梅基因组和转录组遗传信息匮乏, 某些未被注释的基因可能

是其特有的基因。将数据在 NR 数据库中进行分析, 发现大多数 unigenes(28.34%)能够成功匹配到莲花的蛋白质序列, 其次是葡萄(8.53%)和可可(5.41%), 这可能与目前 NCBI 数据库中莲花、葡萄和可可的基因组数据比较丰富有关。

在黑种草中, 有研究者通过基因敲除技术来研究花器官形成机制, 研究表明决定花器官特征的基因有 AP3-1、AP3-2、AP3-3、PI1、PI2、AG1、SEP1、SEP2、SEP3 和 AGL6, 它们表达量的不同会影响花形态的建成^[17]。本研究中的小花草玉梅花器官在野生条件下发生了变异, 王超等的研究将变异分为 5 类, 分别为全白变异、绿白相间变异、五瓣变异、全绿变异和极端变异。张婷等研究发现, 正常植株和绿白相间变异植株的 AP3-3 基因序列是不同的, 在变异植株的上游调控区有一段 49 bp 的插入; 突变植株的 AP3-3 序列与正常相比有 4 个碱基突变点^[18]。参照前人对小花草玉梅的研究, 本研究选取了 12 个花发育相关的 MADS 基因分别在小花草玉梅正常花和 5 种变异花中做实时定量 PCR 实验, 通过主成分分析表明, AGL6、SEP3、FUL1、PI2 及 SEP1 的表达量为小花草玉梅花形态建成的主要影响指标。

本研究首次建立了小花草玉梅转录组数据库,

有助于其分子遗传信息的大量扩充,初步挖掘了12条花发育相关的MADS基因,为进一步研究小花草

玉梅花形态发育及变异的分子机制,开展花发育的基因克隆及功能验证等研究提供了宝贵资源。

参考文献:

- [1] 常鸿莉. 毛茛科花形态发生和花变态及其系统学意义[D]. 西安:西北大学,2005.
- [2] XU Z H, CHONG K. Plant developmental biology: past, present and future[J]. *Acta Botanica Sinica*, 2002, **44**(9): 1 085-1 095.
- [3] COEN E S, MEYEROWITZ E M. The war of the whorls: genetic interactions controlling flower development[J]. *Nature*, 1991, **353**(6 339): 31-37.
- [4] SOLTIS D E, SOLTIS P S, ALBERT V A, et al. Missing links: the genetic architecture of flowers and floral diversification[J]. *Trends in Plant Science*, 2002, **7**(1): 22-31.
- [5] KIM S, KOH J M, KONG H, et al. Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators[J]. *Plant Journal*, 2005, **43**(5): 724-744.
- [6] ENDRESS P K. Angiosperm floral evolution: morphological developmental framework [J]. *Advances in Botanical Research*, 2006, **44**(6): 1-61.
- [7] 常鸿莉,任毅,冯鲁田. 小花草玉梅变态花萼片的形态学研究[J]. 植物分类学报,2005, **43**(3): 225-232.
CHANG H L, REN Y, FENG L T. Morphological observations on metamorphosed sepals in *Anemone rivularis* var. *flore-minore* (Ranunculaceae) [J]. *Acta Phytotaxonomica Sinica*, 2005, **43**(3): 225-232.
- [8] 刘宏顾,狄维忠. 小花草玉梅变态花的形态学研究 I. 变态花的形态学观察[J]. 西北植物学报,1991, **11**(4): 292-298.
LIU H Q, DI W Z. Morphological studies on the metamorphosed flowers of *Anemone rivularis* var. *flore-minore* Maxim. I. morphological observation on the metamorphosed flowers[J]. *Acta Botanica Boreali-Occidentalia Sinica*, 1991, **11**(4): 292-298.
- [9] 王超,戎利勤,冯璐杰,等. 小花草玉梅花器官变异多样性机制初步研究[J]. 园艺学报,2017, **44**(1): 89-96.
WANG C, RONG L Q, FENG L J, et al. Preliminary study of mechanism on diversity of floral organs variation in *Anemone rivularis* var. *flore-minore* [J]. *Acta Horticulturae Sinica*, 2017, **44**(1): 89-96.
- [10] HAO D C, GE G B, XIAO P G, et al. The first insight into the tissue specific *Taxus* transcriptome via Illumina second generation sequencing[J]. *Plos One*, 2011, **6**(6): e21220.
- [11] 郭云霞,刘永斌,荣威恒. 转录组研究新技术: RNA-seq 及其应用[J]. 遗传, 2011, **33**(11): 1 191-1 202.
QI Y X, LIU Y B, RONG W H. RNA-Seq and its applications: a new technology for transcriptomics[J]. *Hereditas*, 2011, **33**(11): 1 191-1 202.
- [12] 黄玉兰,殷奎德,向君亮. 蒜苔幼苗叶片转录组分析[J]. 农业生物技术学报,2017, **25**(3): 386-396.
HUANG Y L, YIN K D, XIANG J L. Analysis transcriptome of Coix (*Coix lachryma-jobi*) leaf at seedling stage[J]. *Journal of Agricultural Biotechnology*, 2017, **25**(3): 386-396.
- [13] 叶德,杨支力,李东海,等. 基于高通量测序半夏珠芽转录组研究[J]. 浙江理工大学学报,2017, **37**(2): 282-288.
YE D, YANG Z L, LI D H, et al. Transcriptome analysis of *Pinellia ternata* (Thunb) Breit based on high-throughput sequencing technology[J]. *Journal of Zhejiang Sci-Tech University*, 2017, **37**(2): 282-288.
- [14] 王身昌,胡尚连,曹颖,等. 梁山慈竹高通量转录组测序及差异表达基因分析[J]. 华北农学报,2016, **31**(3): 65-71.
WANG S C, HU S L, CAO Y, et al. High-throughput RNA-seq and analysis on differential expressed gene from *Dendrocalamus farinosus* [J]. *Acta Agriculturae Boreali-Sinica*, 2016, **31**(3): 65-71.
- [15] XU Y, ZHOU W W, ZHOU Y J, et al. Transcriptome and comparative gene expression analysis of *Sogatella furcifera* (Horváth) in response to southern rice black-streaked dwarf virus[J]. *PLoS One*, 2012, **7**(4): e36238.
- [16] FULLWOOD M J, WEI C L, LIU E T, et al. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses[J]. *Genome Research*, 2009, **19**(4): 521-532.
- [17] WANG P P, LIAO H, ZHANG W G, et al. Flexibility in the structure of spiral flowers and its underlying mechanisms[J]. *Nature Plants*, 2015, (2): 15 188.
- [18] 张婷,邢妮,王超,等. 小花草玉梅正常和自然变异植株的AP3-3基因研究[J]. 西北植物学报,2016, **36**(2): 231-240.
ZHANG T, XING N, WANG C, et al. Cloning and sequence analysis of AP3-3 gene in normal plant and natural variant from *Anemone rivularis* var. *flore-minore* [J]. *Acta Botanica Boreali-Occidentalia Sinica*, 2016, **36**(2): 231-240.

(编辑:宋亚珍)