

文章编号:1001-1498(2014)06-0758-06

基于中麻黄萌发种子转录组的黄酮类 化合物合成途径基因的挖掘

邓楠, 史胜青, 常二梅, 刘建锋, 兰倩, 江泽平*

(中国林业科学研究院林业研究所 北京 100091)

摘要:应用新一代高通量测序技术(Illumina Solexa),对中麻黄(*Ephedra intermedia*)同萌期的种子进行转录组测序,数据重头(denovo)组装后,共获得52 007条Unigene序列,总长为25 043 596 bp。COG功能注释、GO分类及KEGG代谢通路分析后,获得了64 751个GO功能注释,17 701个COG功能注释以及16 942个KEGG注释,并从KEGG通路中找到参与黄酮类化合物合成途径的关键基因片段283个,其中包含了查尔酮合酶基因、细胞色素相关基因和花青素还原酶等基因。中麻黄转录组测序工作的完成,极大地扩充了麻黄的基因资源,为麻黄属植物分子系统进化分析、抗性和药用功能基因的发掘与利用、遗传改良等研究奠定基础。

关键词:中麻黄;转录组;黄酮类化合物;高通量测序

中图分类号:S567.1

文献标识码:A

Exploring Flavonoid Biosynthetic Pathway Genes Based on Transcriptome of *Ephedra intermedia* Germinating Seeds

DENG Nan, SHI Sheng-qing, CHANG Er-mei, LIU Jian-feng, LAN Qian, JIANG Ze-ping

(Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China)

Abstract: The application of a new generation high-throughput sequencing technology (Illumina solexa) to global transcriptome of different stages of germinated *Ephedra intermedia* Schrenk ex Mey seeds showed that 52 007 unigenes were acquired, and the total length was 25 043 596 bp after de novo assembly; among them, 64 751 annotations were from GO, 17 701 from COG and 16 942 from KEGG after functional annotation against these databases; from KEGG pathway, 283 unigenes were associated with flavonoids biosynthetic pathway which contained chalcone synthase, cytochrome, anthocyanidin reductase, etc. The result could provide some references for the study of molecular systematic evolution, utilization of resistance and medicine related genes, and its genetic improvement in the future.

Key words: *Ephedra intermedia*; transcriptome; flavonoid; high-throughput sequencing

麻黄属(*Ephedra*)植物属于裸子植物门买麻藤目麻黄科,多为灌木、半灌木或草本植物^[1],在我国主要分布在甘肃、新疆、内蒙、宁夏、青海等省区的干旱和半干旱荒漠植物区系^[2],属于超旱生系列植物,是西北荒漠区的主要建群种,在防风固

沙、稳定植被方面起着非常重要的作用^[3]。该属植物与其同类的2种买麻藤类植物(买麻藤(*Gnetum*)和百岁兰(*Welwitschia*)),在系统进化位置上较为特殊^[4],它们虽为裸子植物却具有一系列被子植物的特征,如:双受精、木质部具有导管和昆

收稿日期:2014-04-28

基金项目:中央级公益性科研院所基本科研业务费专项资金(RIF2010-10);林木遗传育种国家重点实验室专项课题(TGB2013012)

作者简介:邓楠(1989—),男,硕士,主要从事系统进化与林木引种研究。idengnan@sina.com

* 通讯作者:博士,研究员,博士生导师,主要从事树木地理方面的研究。Email:jiangzp@caf.ac.cn

虫授粉等^[5-7],具有较高的科学研究意义,因此历来是学者们研究的热点之一,围绕着其系统位置也产生了很多争论和假说^[5-8]。而在实际应用上,麻黄是重要的传统药用植物之一,以富含麻黄碱而闻名。麻黄碱既有利于心血管系统、中枢神经、呼吸系统,又有较好的抗炎作用^[9]。此外,麻黄还富含黄酮类化合物,该成分在抗氧化、抗肿瘤、抗炎、抗菌等方面有着重要作用^[10]。中麻黄(*Ephedra. intermedina* Schrenk ex Mey)是《中国药典》中规定的生药麻黄之一^[11],总生物碱含量较高;并具有耐寒、耐旱、耐贫瘠、耐沙埋等特性,根系发达、分蘖能力强、根幅宽,能够起到防风固沙、改善地区生境和涵养水源的作用^[12];由于生长在西北干旱、半干旱地区,常与蒿子(*Artemisia annua* L.)等组成放牧场,并能够增强牲畜的抗病能力^[13],是西北地区重要的生态、经济药用灌木之一。目前,麻黄中黄酮类化合物的代谢途径研究还较少。并且,在植物的整个生命周期中,种子的阶段最能够适应不良环境,是植物对于逆境适应程度最好的反映^[14],并且荒漠植物种子萌发是一个复杂的过程,荒漠植物都在生长期的进化历程都演化出了与严酷生境相适应的萌发对策^[15]。另外,由于近年全球气候变化及过度采挖,使麻黄资源严重退化而匮乏。通过研究其种子萌发过程,对深入研究其进化与适应及次生代谢等具有重要意义。因此,开展麻黄种子转录组及次生代谢途径研究显得尤为重要。

随着高通量测序技术的不断发展和完善,二代测序技术在转录组水平上的应用越来越普遍^[16],并且非常适合于基因组图谱尚未完成以及遗传数据信息匮乏的物种^[17-19]。因此,本研究以不同萌发时期的中麻黄种子为研究对象,利用高通量转录组测序技术对其转录组进行测序分析,并探讨中麻黄种子转录组基本信息,将有利于发掘麻黄种子萌发过程中抗逆相关基因及其种子中重要生物活性成分代谢关键基因,这对其优良性状的遗传改良及其植被恢复具有重要意义。

1 材料与方法

1.1 实验材料

中麻黄种子采集于甘肃省民勤县沙漠区(东经101°49'41"~104°12'10";北纬38°3'45"~39°27'37"),采种地平均海拔为1400米,大陆性沙漠气候特征明显。选择大小均匀的种子,置于培养皿中湿

润的滤纸上进行萌发,萌发在光照培养箱中进行,温度为25℃,萌发时间为0、1.5、3、6、9、18和24h,每个时间点取100粒种子。

1.2 RNA提取方法

不同萌发时期种子的总RNA提取采用Trizol Reagent方法(Invitrogen),经DNase I消化处理后,1.2%琼脂凝胶电泳检测总RNA的质量。然后,将不同萌发期种子的总RNA进行等量混合后进行分离纯化(Dynabeads mRNA Purification Kit, Invitrogen)。

1.3 转录组测序

使用随机引物以及逆转录酶将纯化后的总RNA反转录为cDNA第一链,并随后合成第二链cDNA。利用试剂盒(QIAquick PCR Purification Kit)对双链cDNA片段进行纯化,对纯化后的双链cDNA进行末端修复及3'末端加'A'。之后,将cDNA片段两端连接上特定的测序接头,电泳回收、PCR扩增测序样本,利用Illumina HiSeq™ 2000进行测序。

1.4 数据的组装及注释

首先通过对原始数据过滤去杂(去除测序接头序列、重复冗余序列、质量差的序列等),随后对总数据进行de novo拼接。使用短reads组装软件Trinity^[20]给转录组从头组装。Trinity首先将具有一定长度overlap的reads连成更长的片段,这些通过reads overlap(组装中不同reads之间产生的重叠)关系得到的不含N的组装片段称之为Contig,最后得到不含N且两端不能再延长的序列,即为Unigene。接着用Blast软件将所获得的Unigene序列分别与Nr(NCBI非冗余蛋白数据库)、COGs(clusters of orthologous groups)和Swiss-Prot比对(E-value < 0.000 01),进行功能注释和分类处理;再对Unigene进行GO(Gene Ontology)功能注释和分类,并用WEGO软件^[21]对GO注释结果分类作图,然后将Unigene与KEGG(the Kyoto Encyclopedia of Genes and Genomes Pathway Database)数据库进行比对,分析相关的代谢通路。

2 结果分析

2.1 中麻黄测序结果

对中麻黄转录组测序共得到13 334 972 reads, Q20值为95.89%, GC碱基含量为45.28%。(表1)。经Trinity软件组装共获得88 672个contigs,52

007 个 Unigenes。

表 1 测序数据统计结果

组装结果	组装参数
Total reads/n	13 334 972
Total nucleotides/bp	1 200 147 480
Contigs/n	88 672
Unigenes/n	52 007

其中, Contig 平均长度为 316 bp, N50 长度为 488 bp。Contig 长度主要分布在 100 ~ 200 bp, 占有所有 Contig 的 52.72%; Contig 长度在 200 ~ 300 bp 的序列数量为 12 419, 占总量的 14.01%; 长度为 300 ~ 400 bp 的 Contig 数量占总量的 9.50%; 长度为 400 ~ 500 bp 的 Contig 数量为 5 141 个, 占总数量的 5.80%; 长度大于 500 bp 的 Contig 占总数的 17.97% (表 2)。

表 2 中麻黄转录组 Contig 的长度分布统计

Contig 长度/bp	总数	百分比/%
100 ~ 200	46 752	52.72
200 ~ 300	12 419	14.01
300 ~ 400	8 426	9.50
400 ~ 500	5 141	5.80
≥500	15 934	17.97

中麻黄转录组测序共获得的 Unigene 序列 52 007 条, 总长为 25 043 596 bp, N50 为 612 bp, 平均长度为 482 bp。Unigene 长度分布在 100 ~ 500 bp 的序列所占总量比例为 66.44%; 长度分布在 500 ~ 1 000 bp 序列有 13 005 条, 所占比例为 25.01%; 长度在 1 000 ~ 1 500 bp 的序列占总量之和的 6.53%; Unigene 长度在 1 500 ~ 2 000 bp 序列有 821 条; 而 Unigene 长度分布大于 2 000 bp 序列所占比例为 0.44% (表 3)。此外, 对 Unigene 位置分布分析, 表明 Unigene 3' 端的 reads 数量比较少; 而与其它部位匹配的 reads 较多 (图 1)。

表 3 中麻黄 Unigene 组装质量统计

Unigene 长度/bp	总数	百分比/%
100 ~ 500	34 556	66.44
500 ~ 1 000	13 005	25.01
1 000 ~ 1 500	3 396	6.53
1 500 ~ 2 000	821	1.58
≥2 000	229	0.44

2.2 中麻黄转录组功能注释

2.2.1 中麻黄转录组 GO 分类分析 为了从宏观上了解中麻黄种子中基因功能的分布特征, 对获得

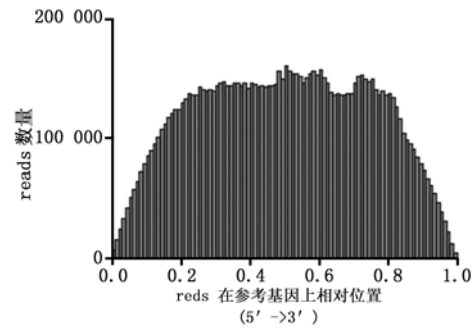
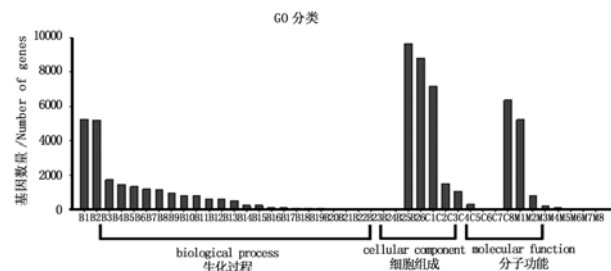


图 1 Reads 在中麻黄 Unigene 的测序随机性结果

的 52 007 个 Unigenes 序列进行 GO 功能分类, 共涉及到 3 个大类 42 种生物功能基因, 分别是生物过程、细胞组分和分子功能。其中, 生物过程占 34.97% (22 644 个), 细胞组分占 44.89% (29 065 个), 分子功能为 20.14% (13 042 个)。但表达基因最多的为细胞组分中的参与细胞部分, 达 9 722 个 (15.01%); 细胞组成表达基因有 8 904 个, 细胞器表达基因占 11.18%; Unigene 参与的基因表达量顺序依次是催化活性 (10.04%)、结合作用 (8.23%)、代谢过程 (8.21%) 和细胞过程 (7.98%) 等所占比例较大 (图 2)。

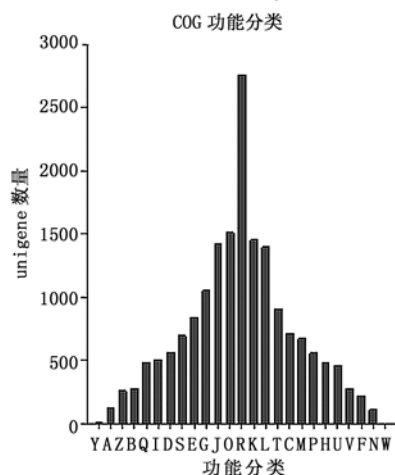


生化过程: B1 代谢过程; B2 细胞过程; B3 刺激反应; B4 定位; B5 胞内定位的建立; B6 发育过程; B7 生物调节; B8 组织或生物起源细胞组件; B9 生物过程调控; B10 涉及多细胞有机体的过程; B11 细胞复制; B12 繁殖过程; B13 信号传递; B14 信号处理; B15 多生物过程; B16 生物过程负调控; B17 生长; B18 生物过程正调控; B19 死亡; B20 免疫系统过程; B21 病毒繁殖; B22 生物附着; B23 染色; B24 节律性过程; B25 移动; B26 细胞杀伤。细胞组成: C1 细胞; C2 细胞组分; C3 细胞器; C4 细胞器组分; C5 高分子配合物; C6 膜附着腔; C7 胞外区; C8 胞外区组分。分子功能: M1 催化活性; M2 结合; M3 载体活性; M4 分子转导活性; M5 酶调节活性; M6 转录调节因子; M7 抗氧化活性; M8 蛋白结合转录因子的活动。

图 2 中麻黄转录组 Unigenes 的 GO 分类图

2.2.2 中麻黄转录组 COG 分类分析 将中麻黄转录组所得的 Unigene 与 COG 数据库进行比对, 根据

比对结果对 Unigenes 进行功能分类和统计,共获得了 17 701 COG 功能注释,分为 26 个功能分类(图 3)。功能分类结果显示 R 类常规功能预测的 Unigene 数量最多(2 753 个),占总注释数的 15.55%;O 类翻译后修饰(1 508,8.52%)、K 类转录(1 451,8.2%)、J 类(1 426,8.26%)、L 类(1 395,7.88%)和 H 类(1 408,5.92%)所占比例比较大;而 W 类注释最少,仅有 2 条 Unigenes。



A RNA 加工和修改;B 染色体结构及动力系统;C 能量生产和转换;D 细胞周期控制、细胞分裂、染色体分区;E 氨基酸运输和代谢;F 核苷酸运输和代谢;G 碳水化合物运输和代谢;H 辅酶运输和代谢;I 脂类运输和代谢;J 翻译、核糖体结构和生物转化;K 转录;L 复制、重组和修复;M 细胞壁/膜/包膜生物合成;N 细胞运动;O 蛋白质翻译后修饰,蛋白转换,分子伴侣;P 无机离子运输和代谢;Q 次生代谢产物生物合成、运输和分解代谢;R 一般功能预测;S 未知功能;T 信号传导机制;U 细胞内运输、分泌和膜泡运输;V 防御机制;W 细胞外结构;Y 核结构;Z 细胞骨架

图3 Unigenes 的 COG 分类结果图

2.2.3 中麻黄的 KEGG 注释 KEGG 作为全球影响力最大的生物代谢数据库之一^[22],涉及系统信息、基因组信息和化学信息^[23]。通过 KEGG 注释发现,中麻黄种子萌发期间共有 16 942 个 Unigenes 参与了 127 个代谢通路,排名前 15 位的代谢通路见表 4。其中代谢途径参与基因最多,达 3 799 个,占有所有 Unigene 的 22.42%;其次是次生代谢产物生物合成途径有 Unigene 1 929 个参与,占 11.39%;再次是参与剪接体的 Unigene 占 4.43% (750 个)。

2.2.4 中麻黄黄酮类化合物合成途径发掘 本研究发现中麻黄种子萌发过程中涉及黄酮类化合物合成途径的 Unigene 有 283 个,几乎涵盖了该合成途径中所有基因(表 5)。其中包含的酶与结构基因有:查尔酮合成酶基因(*CHS*)、羟基肉桂酰 CoA 莽

表 4 KEGG 注释中 Unigene 数量前 15 个通路

编号	通路	Unigene 数量	百分比/%
ko01100	代谢途径	3 799	22.42
ko01110	次生代谢产物生物合成	1 929	11.39
ko03040	剪接体	750	4.43
ko04075	植物激素信号传导	684	4.04
ko04626	植物病原菌互作	642	3.79
ko03013	RNA 运输	617	3.64
ko04141	内质网上的蛋白质加工	532	3.14
ko03010	核糖体	459	2.71
ko00230	嘌呤代谢	451	2.66
ko04120	泛素介导的蛋白降解	413	2.44
ko03008	真核生物中的核糖体合成	386	2.28
ko04144	内吞作用	371	2.19
ko03018	RNA 降解	369	2.18
ko00940	生物合成	353	2.08
ko00500	淀粉和蔗糖代谢	342	2.02

草酸基因、反式肉桂酸 4-单加氧酶基因、二氢黄酮-3-羟化酶基因(*F3H* 或 *FHT*)、二氢黄酮醇-4-还原酶基因(*DFR*)、黄酮类 3',5'-羟化酶基因(*F3'5'H*)、花青素还原酶基因(*ANR*)、黄酮合成酶基因(*FNS1*)等。另外,还发现黄酮类化合物组分中的黄酮与黄酮醇的合成途径的 Unigene 有 106 个。除了有上述的类黄酮 3'-单加氧酶基因、黄酮类 3',5'-羟化酶基因,还包含了黄酮醇 3-O-甲基转移酶基因、黄酮醇 3-O-葡糖基转移酶基因。这说明中麻黄种子萌发期间黄酮与黄酮醇组分可能是其主要产物。

3 讨论

本研究为了更加系统和完整地获取中麻黄种子萌发过程中的转录组数据信息,选用了不同萌发时间的中麻黄种子以获取覆盖度广且较为准确的信息,构建了转录组测序文库并进行了转录组测序工作。与其他转录组研究相比较,如油松(*Pinus tabulaeformis*)转录组测序中 Contig 的平均长度为 382 bp, N50 为 744 bp^[24],根腐线虫(*Pratylenchus coffeae*)转录组中共有 326 971 条 reads,Contig 的平均长度为 458 bp^[25];日本三角涡虫转录组中拼接得到 37 218 条 Unigene 数据,平均长度 468 bp^[26]。相比下说明本研究测序所得到的数据量大,组装效果较好。

黄酮类次生代谢物在植物中广泛存在,不仅在植物与微生物互相识别、防止病原微生物感染以及抵御逆境等过程中均发挥重要作用^[27],还由于具有结构多样、分布广泛,在植物进化过程中的较为保

表5 中麻黄种子萌发期间黄酮类化合物合成途径中的相关基因

编号	名称	一般注释
K01859	<i>CHI1</i>	参与花青素生物合成
K00091	<i>DFRA</i>	催化生成二氢山柰酚
K00660	<i>CHS</i>	催化丙二酸单酰 CoA 与香豆酰 CoA 缩合成查耳酮的反应
K00475	<i>F3H</i>	参与二氢山柰酚 的转换过程
K00487	<i>CYP73A5</i>	控制授粉或紫外防护时必要色素合成的碳通量
K00588	<i>CCOAOMT1</i>	涉及植物细胞壁的增强,以及随着细胞壁形成中伤害的响应和病原体入侵等。
K05278	<i>FLS1</i>	催化二氢黄酮醇转化为黄酮醇
K05277	<i>LDOX</i>	将无色花青素转化为花青素
K05280	<i>CYP75B1</i>	将柚苷配基转化为圣草酚、将二氢山柰酚转化为二氢槲皮素
K13081	<i>LAR</i>	催化儿茶素、阿福豆素和儿茶酸的合成
K13082	<i>DFR</i>	为涉及黄酮代谢的双功能酶,该反应以二氢槲皮素、二氢茨非醇和柑橘素等为基质
K13083	<i>CYP75A1</i>	催化柑橘素和圣草酚 3',5 端的羟基化,催化二氢槲皮素转化为而氢杨梅素
K13065	<i>HST</i>	涉及木质素合成的酰基转移酶
K08695	<i>ANR</i>	涉及凝缩类丹宁的合成,是花青色素转化成为(-)-表儿茶素的主要产物
K09754	<i>CYP98A3</i>	细胞色素 P450 中催化喹啉酸转化为木质素单体,是合成木质素的基本要素

守,常被用于进化研究^[28]。目前对黄酮化合物生物合成各个步骤、催化反应的酶及其基因都有基本研究,对包括拟南芥、金鱼草、松、桃、矮牵牛及葡萄在内的 26 种植物的黄酮合成基因的统计表明,至少已有 103 个黄酮合成相关基因^[29]。本次转录组研究中发现的黄酮类合成基因合成酶的种类与其他物种相比^[30],包含了如 *CHS*、*DFR*、*ANR*、*CHI1* 和 *F3H* 等关键合成基因^[31],并且与吊兰 (*Chlorophytum borivilianum* L.) 中通过转录组发现的黄酮合成大部分先关基因一致,同时与其他物种如淫羊藿属 (*Epimedium* L.)^[30] 植物等对比发现在种类上还是有一定的差异,一方面可能因为物种的差异使部分反应的底物以及条件不同所需的酶不同,另一方面可能是由于种子阶段部分基因没有表达,以及提取、拼接过程中的误差造成的。

虽然非模式生物具有许多模式生物缺少的特征,但是能通过对非模式生物转录组的研究来解答基因进化、遗传育种以及生态等方面的许多问

题^[32];而且与以大量基因序列或分子数据库信息为前提的基因芯片技术相比,该方法更适合基因组图谱尚未完成或数据信息匮乏的物种^[33]。同时由于麻黄为非模式植物,各种类型的生物信息数据库仍在不断更新完善中,可供参考的遗传信息相对较少,还可能存在一些特异性的新基因没有被发现,所以有待进一步的深入研究。本研究所得数据虽然不能覆盖整个中麻黄基因组序列编码区,但仍是目前较为全面的表达信息数据。尤其是以种子萌发时期为研究对象,发现了色氨酸、双萜类、类胡萝卜素和亚麻酸等众多植物激素的代谢途径,对于进一步分析研究麻黄种子萌发中的抗逆的相关基因提供基础。

麻黄作为西北主要的荒漠建群种以及系统位置较为特殊的物种,尤其是还具有一定的药用价值,所以以本次转录组为基础,未来可进一步开发麻黄属的分子标记,从而对西北地区麻黄植物的遗传的结构和多样性开展研究,评估和保护其遗传资源;还能够克隆麻黄中的黄酮合成关键基因,更好利用其药用价值以及开发在系统进化方面的潜力;也为麻黄这种荒漠植物的种子萌发特殊抗逆途径和机制及其优良性状遗传改良^[15] 等提供了大量的遗传数据资源。

参考文献:

- [1] 刘斌心. 中国沙漠植物志[M]. 科学出版社, 1985.
- [2] 雍世鹏, 刘钟龄. 内蒙古植被[Z]. 呼和浩特: 科学出版社, 1985.
- [3] 张晔明, 杨自辉, 王继和, 等. 2 种麻黄光合及其耐逆性分析[J]. 西北植物学报. 2007, 27(7): 1473-1478.
- [4] Huang J, Giannasi D E, Price R A. Phylogenetic relationships in *Ephedra* (Ephedraceae) inferred from chloroplast and nuclear DNA sequences[J]. Molecular Phylogenetics and Evolution. 2005, 35(1): 48-59.
- [5] Price R A. Systematics of the Gnetales: a review of morphological and molecular evidence[J]. International Journal of Plant Sciences. 1996: S40-S49.
- [6] Friedman W E. Double fertilization in *Ephedra*, a nonflowering seed plant: its bearing on the origin of angiosperms[J]. Science. 1990, 247(4945): 951-954.
- [7] Doyle J A. Seed plant phylogeny and the relationships of Gnetales [J]. International Journal of Plant Sciences. 1996: S3-S39.
- [8] Zhong B, Yonezawa T, Zhong Y, et al. The Position of Gnetales among Seed Plants: Overcoming Pitfalls of Chloroplast Phylogenomics [J]. Molecular Biology and Evolution. 2010, 27(12): 2855-2863.
- [9] 查丽杭, 苏志国, 张国政, 等. 麻黄资源的利用与研究开发进展 [J]. 植物学通报. 2002, 19(4): 396-405.

- [10] Chen Z, Hu Y, Wu H, *et al.* Synthesis and biological evaluation of flavonoids as vasorelaxant agents [J]. *Bioorganic medicinal chemistry letters*. 2004, 14(15): 3949–3952.
- [11] 王庆彪. 裸子植物的分子系统发育及麻黄碱类化学成分的进化起源[D]. 复旦大学, 2006.
- [12] 白可喻, 戎郁萍, 徐斌. 甘草和麻黄资源的生物多样性价值和保护[J]. *中国农业资源与区划*. 2009(4): 64–69.
- [13] 赖陈武, 刘颖, 李艾莲. 麻黄离体培养研究进展[J]. *世界科学技术: 中医药现代化*. 2003, 5(2): 37–39.
- [14] Gutterman Y. *Survival strategies of annual desert plants* [M]. Springer, 2002.
- [15] 张勇, 薛林贵, 高天鹏, 等. 荒漠植物种子萌发研究进展[J]. *中国沙漠*. 2005, 25(1): 106–112.
- [16] 李小白, 向林, 罗洁, 等. 转录组测序(RNA-seq)策略及其数据在分子标记开发上的应用[J]. *中国细胞生物学学报*. 2013(05): 720–726.
- [17] Wang E T, Sandberg R, Luo S, *et al.* Alternative isoform regulation in human tissue transcriptomes [J]. *Nature*. 2008, 456(7221): 470–476.
- [18] Sultan M, Schulz M H, Richard H, *et al.* A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome[J]. *Science*. 2008, 321(5891): 956–960.
- [19] 吴琼, 孙超, 陈士林, 等. 转录组学在药用植物研究中的应用[J]. *世界科学技术: 中医药现代化*. 2010(3): 457–462.
- [20] Science For Life Laboratory S, Universitet U, Det M O F V, *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome[J]. *Nature Biotechnology*. 2011, 29(7): 644–652.
- [21] Ye J, Fang L, Zheng H, *et al.* WEGO: a web tool for plotting GO annotations[J]. *Nucleic Acids Res*. 2006, 34(Web Server issue): W293–W297.
- [22] Kanehisa M, Goto S, Sato Y, *et al.* KEGG for integration and interpretation of large-scale molecular data sets[J]. *Nucleic Acids Res*. 2012, 40(Database issue): D109–D114.
- [23] 周国艳, 胡望雄, 徐建红, 等. 整合多个组学(omics)分析植物代谢产物及其功能[J]. *浙江大学学报(农业与生命科学版)*. 2013(03): 237–245.
- [24] Niu S H, Li Z X, Yuan H W, *et al.* Transcriptome characterisation of *Pinus tabulaeformis* and evolution of genes in the *Pinus* phylogeny [J]. *BMC Genomics*. 2013, 14: 263.
- [25] Haegeman A, Joseph S, Gheysen G. Analysis of the transcriptome of the root lesion nematode *Pratylenchus coffeae* generated by 454 sequencing technology[J]. *Molecular and Biochemical Parasitology*. 2011, 178(1): 7–14.
- [26] Qin Y, Fang H, Tian Q, *et al.* Transcriptome profiling and digital gene expression by deep-sequencing in normal/regenerative tissues of planarian *Dugesia japonica* [J]. *Genomics*. 2011, 97(6): 364–371.
- [27] Schijlen E G, Ric De Vos C H, van Tunen A J, *et al.* Modification of flavonoid biosynthesis in crop plants[J]. *Phytochemistry*. 2004, 65(19): 2631–2648.
- [28] Clegg M T, Durbin M L. Flower color variation: a model for the experimental study of evolution [J]. *Proc Natl Acad Sci U S A*. 2000, 97(13): 7016–7023.
- [29] 朱平, 王伟, 程克棣. 药用植物功能基因[J]. *中国生物工程杂志*. 2004, 24(2): 3–8.
- [30] 张华峰, 王瑛, 黄宏文. 黄酮类化合物生物合成途径的进化及其在淫羊藿中的研究展望[J]. *中草药*. 2006, 37(11): 1745–1751.
- [31] Kalra S, Puniya B L, Kulshreshtha D, *et al.* De Novo Transcriptome Sequencing Reveals Important Molecular Networks and Metabolic Pathways of the Plant, *Chlorophytum borivilianum* [J]. *PLoS ONE*. 2013, 8(12): e83336.
- [32] 刘红亮, 郑丽明, 刘青青, 等. 非模式生物转录组研究[J]. *遗传*. 2013(08): 955–970.
- [33] 魏利斌, 苗红梅, 张海洋. 芝麻发育转录组分析[J]. *中国农业科学*. 2012, 45(7): 1246–1256.