

# 基于马尾松干旱转录组的抗旱功能 SSR 位点分析

杜明凤<sup>1,2</sup>, 丁贵杰<sup>1\*</sup>

(1. 贵州大学林学院, 贵州省森林资源与环境研究中心, 贵州 贵阳 550025; 2. 贵州师范大学喀斯特研究院, 贵州 贵阳 550001)

**摘要:** [目的] 了解马尾松干旱胁迫转录组序列的功能分布和 SSR 位点分布特征, 并探索与干旱关联的 SSR 位点。 [方法] 对马尾松幼苗进行持续干旱胁迫, 选取干旱 10、15、25 d 及正常供水对照的马尾松针叶样品, 通过总 RNA 提取、Illumina 测序、raw reads 去冗、Trinity 拼接获得 unigenes。利用 Blast 比对, 对 Unigene 序列进行 GO、KOG、KEGG 功能注释及分类; 利用 Misa 软件进行 SSR 位点批量搜寻, primer 3.0 软件进行规模化 SSR 引物设计; 利用 GOSec (1.10.0)、KOBAS (v2.0.12) 软件对差异表达的含 SSR 位点 Unigene 进行 GO、KEGG 显著性富集分析。 [结果] 马尾松转录组 194 821 个 Unigene 中, 101 806 个 Unigene 获得注释, 包含 64 943 个 GO 功能注释、35 880 个 KOG 功能注释及 30 882 个 KEGG 注释。搜寻到 6 728 个 SSR 位点, 分布于 6 367 个 Unigene 中, SSR 出现频率为 3.45%; 重复类型以单、三、二核苷酸为主, 分别占总 SSR 的 35.82%、33.03% 和 25.22%; 重复基序以 A/T、AT/AT、AG/CT、AGC/CTG、AAG/CTT 为主; 基序长度以 10~20 bp 的短序列 SSR 为主; 基序重复次数以 5~10 次重复占优势; 批量设计 13 338 对 SSR 引物。422 个含 SSR 位点 Unigene 具有差异表达; KEGG 富集分析发现, 有 11 个含 SSR 位点的差异 Unigene 参与了光合作用、植物激素信号传导及类胡萝卜素合成等 3 个与干旱响应相关的代谢途径。 [结论] 从较高质量的马尾松转录组中获得 101 806 个具有注释的 Unigene; 从 6 367 个 Unigene 中挖掘出 6 728 个 SSR 位点; 422 个含 SSR 位点的差异表达 Unigene 中筛选出 11 个与干旱关联的 SSR 功能位点, 为抗旱功能基因定位及马尾松抗旱分子机制研究奠定基础。

**关键词:** 马尾松; 转录组; 基因注释; SSR; 基因差异表达

中图分类号: S718.46

文献标识码: A

文章编号: 1001-1498(2018)05-0009-11

## Analysis of SSR Loci of Functional Gene Linked to Drought Resistance Based on Transcriptome Sequences in *Pinus massoniana* under Drought Stress

DU Ming-feng<sup>1,2</sup>, DING Gui-jie<sup>1</sup>

(1. College of Forestry/Institute for Forest Resources & Environment of Guizhou, Guizhou University, Guiyang 550025, Guizhou, China;

2. School of Karst Science, Guizhou Normal University, Guiyang 550001, Guizhou, China)

**Abstract:** [Objective] The transcriptome data of *Pinus massoniana* under drought stress were used to clarify the function distribution of sequences, as well as the characteristics and distribution patterns of SSR loci, and to explore the key SSR loci linked to drought-resistant gene. [Method] The *P. massoniana* needle samples under lingering drought stress for 10, 15, and 25 days and the corresponding samples with sufficient water as the control (CK) were selected to extract the total RNA. Illumina sequencing were performed to generate raw reads. After removal of low-quality data, the transcriptome assembly was conducted using Trinity software. The unigenes of transcriptome were annotated by aligning with several public databases via BLAST program, including GO (Gene Ontology), KOG (eukaryotic orthologous groups), and KEGG (Kyoto Encyclopedia of Genes and Genomes). The SSRs loci were ex-

收稿日期: 2017-05-22

基金项目: 国家自然科学基金课题(31260183); 国家科技支撑课题(2015BAD09B0102); 贵州省重大专项课题(黔科合重大专项字[2012]6001); 贵州大学研究生创新基金(研农 2015006)

作者简介: 杜明凤(1979—), 女, 博士研究生, 副教授, 主要从事森林培育、植物分子遗传与育种研究。E-mail: dmf1979@126.com.

\* 责任作者: 丁贵杰(1960—), 男, 博士生导师, 教授, 主要从事森林培育、林木抗逆机制研究。E-mail: gjding@gzu.edu.cn

amined using Misa software, and the PCR amplification SSR primers were designed using Primer 3.0 software. GO and KEGG enrichment analysis were implemented using GOSep (1.10.0) and KOBAS software, respectively, to determine the major process of biological process and metabolic pathways of differentially expressed unigenes contained SSR loci. [ **Result** ] A total of 101 806 unigenes were annotated from 194 821 unigenes of transcriptome. Among them, 64 973 functional annotations were from GO database, 35 880 from KOG database and 30 882 from KEGG database. Moreover, 6 728 SSR loci were identified and distributed in 6 367 unigenes, and their average frequency of SSRs was 3.45%. Among all the SSR motifs, mononucleotide, trinucleotide and dinucleotide were the major repeated types, with occurrence frequency of 35.82%, 33.03% and 25.22%, respectively; the form of A/T, AT/AT, AG/CT, AGC/CTG, and AAG/CTT were the most frequent motifs, the length from 10 to 20 bp were the most repeat motifs, and the SSR repeat numbers from 5 to 10 were the most repeat numbers of motifs. A total of 13 338 pairs of SSR primers were designed for marker development of *P. massoniana*. Furthermore, among the 6 367 unigenes containing SSR loci, 422 unigenes were differentially expressed on drought stress versus the control. Enriched analysis of KEGG pathway showed that 11 unigenes containing SSR loci were significantly enriched into three KEGG pathways, including photosynthesis, plant hormone signal transduction and carotenoid biosynthesis, which were linked to the plant response to drought stress. [ **Conclusion** ] A total of 101 806 unigenes were annotated from a higher quality of transcriptome database in *P. massoniana*, 6 728 SSR loci were identified and distributed from 6 367 unigenes, 11 SSR loci from 422 differentially expressed genes containing SSR loci were identified linking to the plant response to drought stress. These results can be used for the subsequent study on molecular mechanism for drought resistance and functional gene localization in *P. massoniana*.

**Keywords:** *Pinus massoniana*; transcriptome; gene annotation; SSR; differentially expressed unigene

转录组学借助高通量测序技术快速获取海量转录本,从 RNA 水平解析物种在特定环境下的基因功能表达、生物学过程及分子运行机制,是一种快捷高效的分子生物学研究手段,更适用于基因组较大、且基因信息相对匮乏的非模式生物<sup>[1-2]</sup>。基于转录组的 SSR 标记,不仅具有 SSR 多态性高、重复性好等特性,更具有转录组经济、高效、信息量大等优势<sup>[3]</sup>;众多林木已相继开展转录组 SSR 标记开发研究,如地中海松(*Pinus halepensis* Mill.)<sup>[4]</sup>、红松(*Pinus koraiensis* Sieb. et Zucc.)<sup>[5]</sup>等。同时,转录组 SSR 标记源自转录区编码序列,直接反映基因表达的变异<sup>[6]</sup>,成为与基因表达直接关联的功能标记<sup>[7]</sup>,涉及基础代谢、信号转导及转录等基因调控的各个方面<sup>[8]</sup>;因此,通过转录组学分析手段,可快速精准地锁定目的基因的 SSR 功能标记,加快目的基因定位研究进程。但借助转录组数据挖掘 SSR 功能标记的研究非常少,仅见杨洋等<sup>[9]</sup>通过茄子转录数据找到 1 个与热胁迫相关的 SSR 候选位点。目前,SSR 标记结合 BSA 法(Bulked Segregant Analys),在水稻(*Oryza sativa* L.)<sup>[10]</sup>、大豆(*Glycine max* (Linn.) Merr.)<sup>[11]</sup>、向日葵(*Helianthus annuus* L.)<sup>[12]</sup>等作物中找到耐旱基因的连锁标记,为耐旱功能基因定位及耐旱性育

种奠定了基础。

马尾松(*Pinus massoniana* Lamb.)是中国南方主要造林树种,具有速生、丰产、综合利用程度高等特点,在我国森林资源发展和生态建设中具有重要地位<sup>[13]</sup>。马尾松分子遗传学研究发展迅速,SSR<sup>[14]</sup>、ISSR<sup>[15]</sup>、IRAP<sup>[16]</sup>等标记在马尾松遗传图谱构建、遗传结构变异、遗传多样性研究中被广泛运用;有学者曾利用马尾松近缘种 EST 序列<sup>[17]</sup>、基因组<sup>[2]</sup>或转录组<sup>[18]</sup>开发 SSR 标记,但开发的 SSR 标记仍较少,无法满足分子标记辅助育种需求。近年来,南方频发的季节性干旱严重威胁马尾松生长,抗旱种质选育已成为应对干旱逆境的关键。课题组前期选择 3 个速生性和适生性较好的马尾松优良家系进行干旱胁迫,通过对其形态、生长及生理等指标进行对比分析,筛选出马尾松抗旱种质<sup>[19]</sup>,并对其进行高通量测序获得干旱胁迫转录本数据。本研究将对该转录本进行 Unigene 功能注释及分类,分析 SSR 位点分布特征;在此基础上,对含 SSR 位点的 Unigene 序列进行差异表达以及 GO、KEGG 显著性富集分析,深度挖掘与功能基因直接关联的 SSR 标记,为后续马尾松 SSR 规模性标记开发、抗旱分子机制、功能基因定位等研究奠定基础。

## 1 材料与方法

### 1.1 试验材料及干旱处理

材料为长势一致的马尾松 2 年生幼苗(苗高 65 cm 左右),是课题组前期筛选出的抗旱家系<sup>[19]</sup>。试验设对照组和干旱组,对照组每 3~5 d 浇 1 次水,维持正常水分,干旱组在浇水后自然持续干旱至 30 d。对照组和干旱组在 10、15、25 d 时,对应的土壤相对含水量分别为 81.9%、80.3%、81.2% 和 57.8%、46.6%、30.1%,分属于湿润(75%~80%)、轻微干旱(55%~60%)、中度干旱(45%~50%)、重度干旱(30%~35%)<sup>[20]</sup>。选取供水 10 d(CK)及干旱 10、15、25 d 的同位针叶,重复 2 次共 8 个样品,交诺禾致源生物有限公司完成 RNA 提取、质检及转录组测序。

### 1.2 转录组数据及 unigene 功能注释

通过 Hiseq2000 高通量测序平台获得 8 个针叶转录组数据(NCBI 数据库 Accession: SRX 2327081~2327084; 2310441~2310444)。用 NGSQCToolkit v 2.5 软件去除接头序列、poly-A 及低质量序列,用 Trinity 软件完成序列拼接,筛选大于 200 bp 的 unigene 序列。通过 Blast 序列比对,对 unigene 进行 Nr、Nt、Swiss-Prot、PFAM、GO、KOG、KEGG 7 大数据库信息注释,注释成功的 unigene 根据功能的不同,进一步划分不同的 GO 基因功能、KOG 类别及 KEGG 代谢途径。

### 1.3 SSR 位点的挖掘及引物设计

用 Misa(MIcorSAtelescope Identification Tool) 软件对 unigene 进行 SSR 位点检测。参数设置为单核苷酸重复数 10 次或 10 次以上,二~六核苷酸的最小重复数分别为 6、5、5、5、5,SSR 位点侧翼序列长度  $\geq 50$  bp。用 primer 3(2.3.5 版,默认参数)对含 SSR 位点的 unigene 批量设计引物,引物设计参照通用标准,并按下列标准进一步筛选引物:引物错配 5' 端低于 3 个碱基,3' 端低于 1 个碱基;引物中无 SSR;去除可匹配其他 unigene 的引物,筛选唯一匹配引物;用 SSRFinder 校验 SSR,将产物序列搜寻的 SSR 与 MISA 结果比较,保留具有相同 SSR 产物的引物。为验证引物有效性,对随机筛选的 11 对引物用 1 个 DNA 材料进行 PCR 扩增,扩增条带用 1.5% 的琼脂糖凝胶电泳检测。

### 1.4 含 SSR 位点的 unigene 差异表达及富集分析

对含 SSR 位点的 unigene 进行差异表达分析,以

及 GO、pathway 富集分析。RPKM 法(*Reads Per Million Kilobases*)计算 unigene 表达量。采用 DESeq 法(1.10.1)比较干旱胁迫与正常供水(CK)之间的 unigene 表达量,设定 Fold change  $\geq 2$  和 FDR  $< 0.05$  来筛选样品间具有显著性差异的 unigene。针对差异表达 unigene,用 GOSec(1.10.0)、topGO(2.10.0) 软件进行 GO 富集分析,用 KOBAS(v2.0.12) 软件进行 KEGG 富集分析,均取校正 *P*-Value  $< 0.05$ 。

### 1.5 qRT-PCR 验证

从转录组中随机筛选 4 个 unigenes 进行 qRT-PCR 检验,验证转录组数据可靠性。利用宝生物 RNA LA PCR Kit 试剂盒对转录组测序 RNA 进行反转录合成 cDNA,Primer Premier 5.0 软件设计 unigene 特异性引物,UBC 选作内参基因,ABI 7500 Real Time System 进行 PCR 扩增。反应体系 20.0  $\mu$ L: SYBR mix 10.0  $\mu$ L,正反向引物各 0.5  $\mu$ L,cDNA 模板 1.0  $\mu$ L,ddH<sub>2</sub>O 8.0  $\mu$ L;扩增程序:95°C 60 s,95°C 10 s(40 个循环),61°C 30 s,72°C 30 s;加溶解曲线程序:95°C 15 s,61°C 60 s,95°C 15 s。差异表达分析采用  $2^{-\Delta\Delta CT}$  定量分析法。

## 2 结果与分析

### 2.1 马尾松转录组的功能注释

2.1.1 转录组序列注释 基于马尾松干旱胁迫转录组,通过低质量 Raw reads 的筛除,高质量 clean reads 的 Trinity 拼接,共获得大于 200 bp 的 Unigene 194 821 个。通过 7 大数据库的 Blast 比对,有 101 806 个 Unigene 获得注释,注释率 52.26%;其余的 93 015 个 Unigene 未获注释,可能为新基因。其中,NR 数据库比对所获得的 66 825 个注释中,16 323 个 Unigene 与云杉属(*Picea*)、3 040 个 Unigene 与松属(*Pinus*)的序列同源,其序列匹配数远高于其它物种。

### 2.1.2 转录组的 GO、KOG、KEGG 功能注释及分类

对注释成功的 Unigene 进行 GO、KOG、KEGG 功能注释并分别归类(图 1~3),该结果将反映马尾松干旱胁迫过程中表达基因的功能分布整体情况。GO 功能分类显示,64 943 个 Unigene 被成功注释,占总数 33.33%。获得的 128 326 个功能注释,被划分为 3 大类:46 779 个生物过程(36.45%)、52 470 个分子功能(40.89%)和 29 077 个细胞组分(22.66%)(图 1)。生物过程包含 25 个功能亚类,主要涉及代谢过程(34 306, 73.34%)、细胞过程

(34 293, 73.31%)、单细胞有机体过程(27 445, 58.67%);与胁迫响应相关的有生物调控(9 993, 21.36%)、胁迫响应(6 741, 14.41%)、信号(3 679, 7.86%)等。分子功能包含 10 个功能亚类,代表性功能包括蛋白结合(32 711, 62.34%)、催化活性(29 984, 57.15%)、转运活性(4 442, 8.47%)等;与胁迫响应相关的有核酸结合转录因子(1 708, 3.26%)、转录因子活性(602, 1.15%)、抗氧化活性(467, 0.89%)。细胞组分包含 21 个功能亚类,主要类型有细胞(18 463, 63.50%)、细胞组分(18 441, 63.42%)、细胞器(12 184, 41.90%)、高分子复合物(11 475, 39.46%)及细胞膜(9 248, 31.81%)等。该结果表明马尾松参与了广泛的细胞过程和代谢活动,几乎涵盖所有干旱胁迫过程的生命活动。

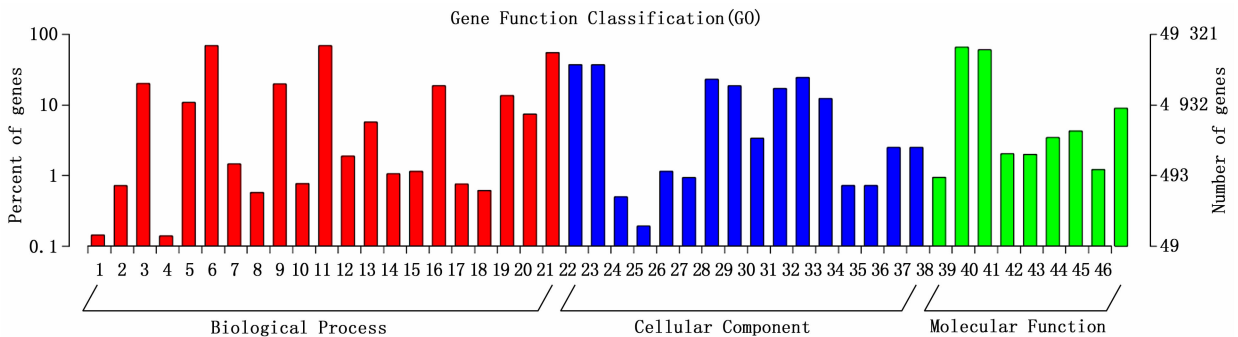
KOG 功能分类显示,35 880 个 Unigene 比对到同源序列,占总数 18.42%;共获得 39 989 个注释,涉及全部的 26 个 KOG 功能类别(图 2)。其中,一般功能预测比例最大(6 179, 17.22%);随后依次为翻译后修饰、蛋白翻转、分子伴侣(4 486, 12.50%),

翻译、核糖体结构和生物合成(3 197, 8.91%),能量产生和转化(2 962, 8.26%),信号转导机制(2 659, 7.41%)、脂类转运及代谢(2 129, 5.93%)等;而胞外结构和细胞迁移的比例最小,分别仅涉及 72 和 23 个 Unigenes。

KEGG 注释结果显示:30 882 个 Unigene 获得 KO 注释,占总数 15.85%,涉及 284 条代谢途径(图 3)。其中,Unigene 注释最多的代谢途径主要涉及糖代谢(4 040, 13.08%)、氨基酸代谢(3 082, 9.98%)、翻译(2 970, 9.62%)、信号转导(2 817, 9.12%)、能量代谢(2 310, 7.48%)、脂类代谢(2 118, 6.86%)等各类代谢及环境适应,表明干旱胁迫下马尾松的各类代谢活动、信号转导过程非常活跃。

## 2.2 马尾松转录组 SSR 特性

2.2.1 SSR 分布特征及与转录组 Unigene 分布的关系 通过 Misa 搜索,从 194 821 个 Unigenes 中获得 6 728 个 SSR 位点,分布于 6 367 个 Unigenes 中,其中,6 031 个 Unigenes 只含 1 个 SSR 位点,336 个 Unigenes 含多个( $\geq 2$ )SSR 位点,出现频率 3.45%,

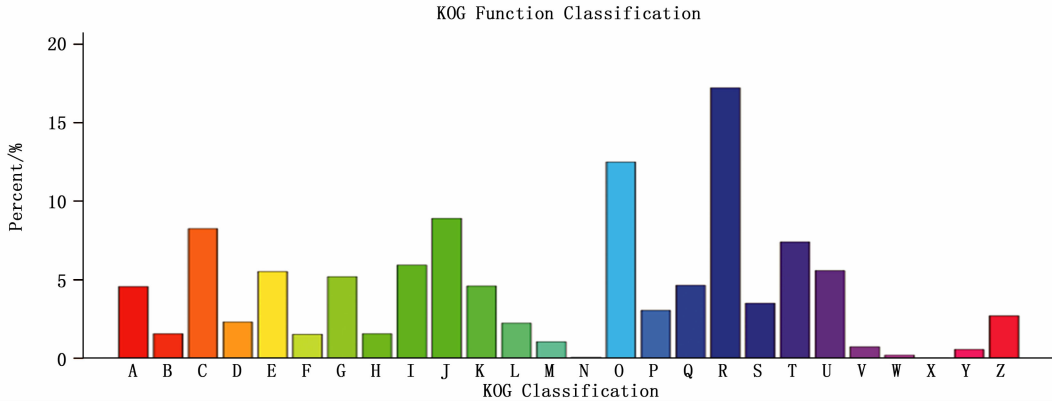


注:生化过程: 1. 生物附着; 2. 生物调节; 3. 细胞集合; 4. 细胞杀伤; 5. 细胞组成或生物合成; 6. 细胞过程; 7. 发育过程; 8. 免疫系统; 9. 定位; 10. 运动; 11. 代谢过程; 12. 多细胞进程; 13. 多个有机体过程; 14. 生物过程负调控; 15. 生物过程正调控; 16. 生物过程调控; 17. 繁殖; 18. 繁殖过程; 19. 刺激反应; 20. 信号传导; 21. 单一有机体过程。细胞组成: 22. 细胞; 23. 细胞组分; 24. 细胞外基质; 25. 细胞外基质组分; 26. 胞外区; 27. 胞外区组分; 28. 大分子复合物; 29. 膜; 30. 膜封闭腔; 31. 膜组分; 32. 细胞器; 33. 细胞器组分; 34. 其他有机体; 35. 其他有机体组分; 36. 病毒体; 37. 病毒体组分。分子功能: 38. 抗氧化活性; 39. 结合; 40. 催化活性; 41. 分子功能调节; 42. 分子转导活性; 43. 核酸结合转录因子活性; 44. 结构分子活性; 45. 蛋白结合转录因子活性; 46. 转运活性。

Note: Biological Process: 1. biological adhesion; 2. biological regulation; 3. cell aggregation; 4. cell killing; 5. cellular component organization or biogenesis; 6. cellular process; 7. developmental process; 8. immune system process; 9. Localization; 10. Locomotion; 11. metabolic process; 12. multicellular organismal process; 13. multi-organism process; 14. negative regulation of biological process; 15. positive regulation of biological process; 16. regulation of biological process; 17. Reproduction; 18. reproductive process; 19. response to stimulus; 20. Signaling; 21. single-organism process. Cellular Component: 22. Cell; 23. cell part; 24. extracellular matrix; 25. extracellular matrix component; 26. extracellular region; 27. extracellular region part; 28. macromolecular complex; 29. Membrane; 30. membrane-enclosed lumen; 31. membrane part; 32. Organelle; 33. organelle part; 34. other organism; 35. other organism part; 36. Virion; 37. virion part. Molecular Function: 38. antioxidant activity; 39. Binding; 40. catalytic activity; 41. molecular function regulator; 42. molecular transducer activity; 43. nucleic acid binding transcription factor activity; 44. structural molecule activity; 45. transcription factor activity, protein binding; 46. transporter activity.

图1 马尾松转录组 Unigene 的 GO 分类图

Fig. 1 GO categorization of non-redundant unigenes in *P. massoniana* transcriptome.



注: A:RNA 加工和修饰;B:染色体的结构和动态;C:能量产生和转换;D:细胞周期控制,细胞分裂,染色体分离;E:氨基酸的运输和代谢;F:核苷酸的运输和代谢;G:碳水化物的运输和代谢;H:辅酶的运输和代谢;I:脂质的运输和代谢;J:翻译,核糖体结构和生物合成;K:转录;L:复制,重组和修复;M:细胞壁/膜/胞外被膜;N:细胞运动;O:翻译后修饰,蛋白质折叠,分子伴侣;P:无机盐的运输和代谢;Q:次生代谢物的生物合成,运输及分解代谢;R:一般功能预测;S:功能未知;T:信号转导机制;U:细胞内运输,分泌和囊泡运输;V:防御机制;W:细胞外结构;X:未知蛋白;Y:核结构;Z:细胞骨架。

Note: A:RNA processing and modification;B:Chromatin structure and dynamics;C:Energy production and conversion; D:Cell cycle control, cell division, chromosome partitioning;E:Amino acid transport and metabolism; F:Nucleotide transport and metabolism;G:Carbohydrate transport and metabolism;H:Coenzyme transport and metabolism;I:Lipid transport and metabolism;J:Translation, ribosomal structure and biogenesis;K: Transcription;L:Replication, recombination and repair;M:Cell wall/membrane/envelope biogenesis;N:Cell motility;O:Posttranslational modification, protein turnover, chaperones;P:Inorganic ion transport and metabolism;Q:Secondary metabolites biosynthesis, transport and catabolism; R:General function prediction only;S:Function unknown;T:Signal transduction mechanisms;U:Intracellular trafficking, secretion, and vesicular transport;V:Defense mechanisms;W:Extracellular structures;X:Unnamed protein;Y:Nuclear structure;Z: Cytoskeleton.

图 2 马尾松转录组 Unigene 的 KOG 分类图

Fig. 2 KOG annotation of putative proteins in *P. massoniana* transcriptome.

平均距离 15.97 kb(表 1)。SSR 出现频率随 Unigenes 长度的增加而增加,在各 Unigenes 长度分组中

依次为 1.64%、2.61%、4.77%、8.51% 及 15.71%。

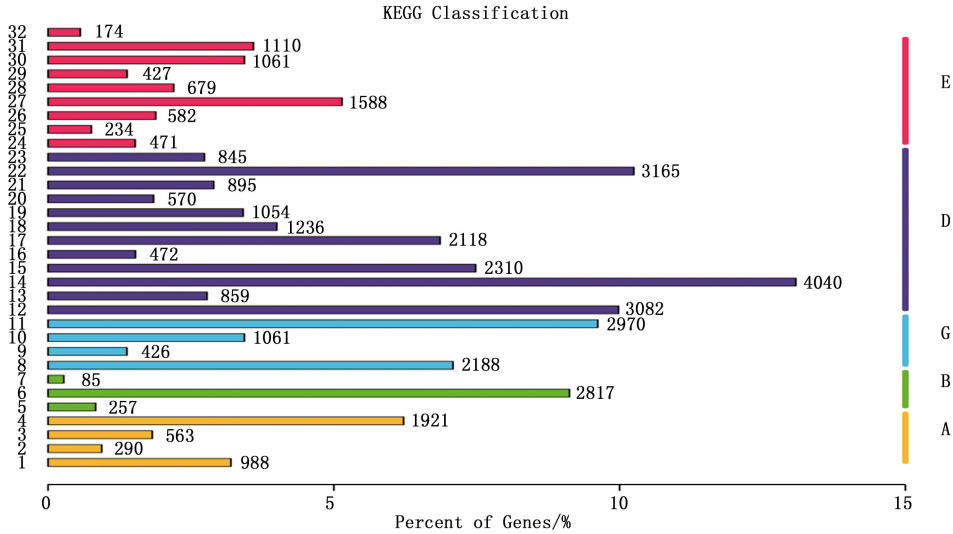
表 1 不同 Unigenes 长度分组中 SSR 位点数、出现频率及平均距离

Table 1 Number, frequency and mean distance of SSR in different unigene length distribution of *P. massoniana*

长度 Length/bp	基因数 Gene number	含 SSR 的基因数 Number of unigene contained SSR	SSR 位点数 Number of SSR loci	出现频率 Frequency/%	平均距离 Mean distance/kb
≤300	97 486	1 569	1 597	1.64	67.30
301 ~ 500	45 841	1 156	1 197	2.61	89.79
501 ~ 1 000	28 051	1 266	1 337	4.77	80.38
1 001 ~ 2 000	15 081	1 188	1 283	8.51	83.77
≥2 001	8 362	1 188	1 314	15.71	81.79
合计 Total	194 821	6 367	6 728	3.45	15.97

2.2.2 SSR 的重复类型及频率 马尾松转录组中单核~六核苷酸的 SSR 重复类型均有分布(表 2)。单、二、三核苷酸重复类型的出现频率占优势,共 6 329 个,占总 SSR 位点的 94.07%,其中,单核苷酸最多,为 2 410 个,占 35.82%,其次三核苷酸为 2 222 个,占 33.03%;二核苷酸为 1 697 个,占 25.22%;其余四、五、六核苷酸重复类型的数量较少,分布相对分散。

2.2.3 SSR 的基序类型及频率 马尾松转录组 SSR 包含 70 种重复基序,单核至六核苷酸的重复基序分别为 2、4、10、22、12、20 种。出现频率以单核苷酸 A/T (2 332 个,占 34.66%),二核苷酸 AT/AT (791 个,占 11.76%)、AG/CT(579 个,占 8.61%)、AC/GT(392 个,占 5.83%),三核苷酸 AGC/CTG (443 个,占 6.58%)、AAG/CTT(317 个,占 4.71%) 较多,其余基序频率均相对较低(表 3)。



注:(A)细胞过程: 1. 细胞生长和死亡; 2. 细胞运动; 3. 细胞通讯; 4. 运输和代谢。(B)环境信息处理: 5. 膜转运; 6. 信号转导; 7. 信号分子的相互作用。(C)遗传信息处理: 8. 折叠,分类和降解 9. 复制和修复; 10. 转录; 11. 翻译。(D)代谢: 12. 氨基酸代谢; 13. 其他次生代谢产物的生物合成; 14. 碳水化合物代谢; 15. 能量代谢; 16. 多聚糖生物合成与代谢; 17. 脂质代谢; 18. 代谢辅助因子和维生素; 19. 其他氨基酸代谢; 20. 萜类和酮类化合物代谢; 21. 核苷酸代谢; 22. 总代谢; 23. 外来物质的降解和代谢。(E)有机系统: 24. 循环系统; 25. 发展; 26. 消化系统; 27. 内分泌系统; 28. 环境适应; 29. 排泄系统; 30. 免疫系统; 31. 神经系统; 32. 感觉系统。

Note: (A) Cellular Processes: 1. Cell growth and death; 2. Cell motility; 3. Cellular commiunity; 4. Transport and catabolism. (B) Environmental Information Processing: 5. Membrane transport; 6. Signal transduction; 7. Signaling molecules and interaction. (C) Genetic Information Processing: 8. Folding, sorting and degradation; 9. Replication and repair; 10. Transcription; 11. Translation. (D) Metabolism: 12. Amino acid metabolism; 13. Biosynthesis of other secondary metabolites; 14. Carbohydrate metabolism; 15. Energy metabolism; 16. Glycan biosynthesis and metabolism; 17. Lipid metabolism; 18. Metabolism of cofactors and vitamins; 19. Metabolism of other amino acids; 20. Metabolism of terpenoids and polyketides; 21. Nucleotide metabolism; 22. Overview; 23. Xenobiotics biodegradation and metabolism. (E) Organismal Systems: 24. Circulatory system; 25. Development; 26. Digestive system; 27. Endocrine system; 28. Environmental adaptation; 29. Excretory system; 30. Immune system; 31. Nervous system; 32. Sensory system.

图3 马尾松转录组 Unigene 的KEGG 分类图

Fig.3 KEGG annotation of putative proteins in *P. massoniana* transcriptome.

表2 不同 SSR 重复类型在马尾松转录组中的出现频率  
Table 2 Occurrence SSR in *P. massoniana* transcriptome

重复类型 Repeat type	数目 Number	比例 Percentage/%	平均长度 Mean length/bp	平均距离 Mean distance/kb	出现频率 Mean frequency/%
单核苷酸 Mononucleotide	2 410	35.82	11.31	44.59	1.24
二核苷酸 Dinucleotide	1 697	25.22	14.17	63.33	0.87
三核苷酸 Trinucleotide	2 222	33.03	16.21	48.37	1.14
四核苷酸 Quadnucleotide	108	1.61	21.36	995.12	0.05
五核苷酸 Pentanucleotide	17	0.25	25.00	6 321.94	0.01
六核苷酸 Hexanucleotide	21	0.31	39.16	5 117.76	0.01
复合型 Composition	253	3.76	70.49	424.79	0.13
合计 Total	6 728	100.00	16.09	15.97	3.45

表3 马尾松转录组 SSR 重复基序的数量及频率  
Table 3 The number and frequency of the motifs in SSR of *P. massoniana* transcriptome

重复基序 Motif	数目 Number	频率 Frequency/%	重复基序 Motif	数目 Number	频率 Frequency/%
A/T	2 332	34.66	AAT/ATT	239	3.55
C/G	90	1.34	ACC/GGT	186	2.76
AC/GT	392	5.83	ACG/CGT	110	1.63
AG/CT	579	8.61	ACT/AGT	32	0.48
AT/AT	791	11.76	AGC/CTG	443	6.58
CG/CG	22	0.33	AGG/CCT	244	3.63
AAC/GTT	260	3.86	ATC/ATG	205	3.05
AAG/CTT	317	4.71	CCG/CGG	231	3.43
			其余 Other	255	3.79

2.2.4 SSR 的基序重复次数及长度 马尾松转录组 SSR 基序的重复次数介于 5~23 次之间,随重复次数增加,SSR 数量呈递减趋势(表 4)。5~10 次重复的 SSR 位点数 5 447 个,占总数 80.97%;11 次重复及以上的 SSR 位点数 1 180 个,占 17.54%。其中,单核苷酸以 10 次重复的基序(1 284 个,占

19.09%)最多,二核苷酸以 6 次重复的基序(820 个,占 12.19%)最多,三~六核苷酸中均以 5 次重复的基序最多。SSR 的长度从 10~212 bp 不等,长度为 10~20 bp 的 SSR 位点最多,共 6 149 个,占 SSR 位点总数的 91.39%;长度大于 20 bp 的共 579 个,占 8.61%(表 5)。

表 4 马尾松转录组 SSR 基序的重复次数及分布

The number of repeats and distribution of the motifs in SSR of *P. massoniana* transcriptome

重复类型 Repeat type	重复次数 repeat number												合计 Total	比例 Percentage/%
	5	6	7	8	9	10	11	12	13	14	15	>15		
单核苷酸 Mononucleotide						1 284	491	248	119	79	48	153	2 422	36.00
二核苷酸 Dinucleotide		820	392	298	161	75	36	2					1 784	26.52
三核苷酸 Trinucleotide	1 563	503	178	21	1		1						2 267	33.70
四核苷酸 Quadnucleotide	89	26	2	1			1						119	1.77
五核苷酸 Pentanucleotide	15												15	0.22
六核苷酸 Hexanucleotide	10	3	2	3			1			1			20	0.30
合计 Total	1 677	1 352	574	323	162	1 359	530	250	119	80	48	153	6 627	98.50
比例 Percentage/%	24.93	20.10	8.53	4.80	2.41	20.20	7.88	3.72	1.77	1.19	0.71	2.27	98.50	

表 5 马尾松转录组 SSR 基序长度及频率

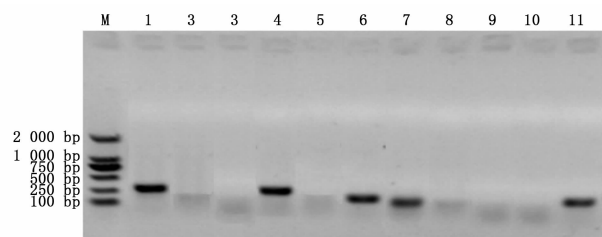
Table 5 The length and frequency of the motifs in SSR of *P. massoniana* transcriptome

项目 Item	基序长度 motif length/bp												合计 Total
	10	11	12	13	14	15	16	17	18	19	20	>20	
SSR 数量 Number of SSR	1 280	492	1 043	119	452	1 583	314	23	650	18	175	579	6 728
频率 Frequency/%	19.02	7.31	15.5	1.77	6.72	23.53	4.67	0.34	9.66	0.27	2.6	8.61	100

2.2.5 SSR 引物设计 从含 SSR 位点的 6 367 个 Unigenes 中成功筛选 4 446 个 Unigenes 并设计出 13 338 对 SSR 引物。其中,引物长度 18~27 bp,GC 含量 40%~55%,退火温度( $T_m$ )57~63℃,正、反向引物退火温度差低于 5℃,PCR 产物大小 100~280 bp。随机筛选 11 对 SSR 引物进行 PCR 扩增检测,5 对引物能有效扩增出 PCR 产物(图 4),引物转化率 45.5%,表明引物具有一定可行性。

### 2.3 含 SSR 位点 Unigene 的功能分析

2.3.1 含 SSR 位点 Unigene 的差异表达 基于转录组功能注释和 SSR 位点挖掘,对含 SSR 位点的 6 367 个 Unigenes 进行差异表达分析,共获得 422 个差异表达 Unigene(图 5)。干旱胁迫 10、15、25 d 与正常供水相比,分别有 325、147、183 个差异表达 Unigene,其中,上调表达分别为 196、66、87 个,下调表达分别为 129、81、96 个;特异性差异表达 Unigene 分别为 181、21、60 个,不同程度胁迫的共差异表达



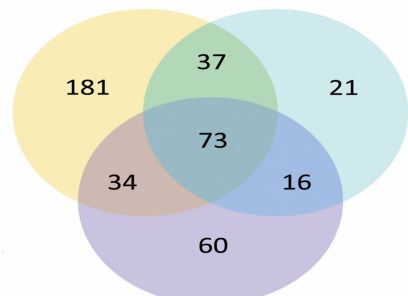
注:M:分子量标记,1~11 为随机引物 1~11。

Note: M mean marker, the number 1~11 mean random primer 1~11.

图 4 马尾松 SSR 引物 PCR 扩增检测

Fig.4 The amplification of SSR primers from transcriptome of *P. massoniana*

Drought 10 days versus control Drought 15 days versus control



Drought 25 days versus control

图 5 不同程度胁迫下的 Unigene 表达总体差异

Fig.5 Venn diagram of the differentially expressed genes identified in three comparisons

Unigene 为 73 个。其中胁迫 10 d 的差异表达 Unigene 数量最多,表明干旱 10 d 时马尾松响应胁迫的各途径中 Unigene 表达量丰富。

**2.3.2 差异表达 Unigene 的 GO 和 KEGG 显著性富集分析** 通过 GO 和 KEGG 显著性富集分析,进一步确定差异表达 Unigene 行使的主要生物学功能以及参与的主要代谢途径及信号转导通路。GO 显著性富集分析发现,422 个含 SSR 位点的差异 Unigenes 中有 261 个参与了 3 大类生物学功能。生物学过程中,有机环化合物合成(GO:1901362)富集的 Unigenes 最多(51),其次为氧化还原过程(GO:0055114)(43)、新陈代谢调控(GO:0019222)(37);分子功能中,氧化还原酶活性(GO:0016491)富集的 Unigenes 最多(42),其次为转运活性(GO:0005215)(29)、核酸结合转录因子(GO:0001071)(16);细胞组分中,胞外区(GO:0005576)富集的 Unigenes 最多(18),其次为线粒体内膜蛋白复合物(GO:0098800)等,表明上述显著富集的生物学功能可能涉及马尾松干旱胁迫响应过程。

KEGG 显著性富集分析发现(图 6),422 个含 SSR 位点的差异 Unigenes 中有 97 个被富集到 53 个代谢途径中,其中,光合作用(ko00195)、类胡萝卜素

合成(ko00906)、植物激素信号传导(ko040753)等 3 个代谢途径被显著富集( $P < 0.05$ ),表明这 3 个代谢途径与马尾松干旱逆境应答相关。光合作用途径富集了 4 个含 SSR 位点 Unigenes,包括 2 个 ATP 合成酶(ATP synthase; c94519\_g2, c88154\_g1),1 个光系统 II(photosystem II; c77320\_g1),1 个氧化还原酶(oxidoreductase; c85918\_g1),均下调表达。类胡萝卜素合成途径富集了 3 个含 SSR 位点 Unigenes,1 个铁离子(iron ion binding; c89714\_g1)呈上调表达,其余 1 个氧化还原酶(oxidoreductase; c78714\_g)和 1 个黄素腺嘌呤二核苷酸(FAD; c69125\_g1)均下调表达。植物激素信号传导途径富集了 4 个含 SSR 位点 Unigenes,1 个茉莉酸(JA; c88597\_g2)呈上调表达,其余 1 个蛋白磷酸酶(PP2C; c68631\_g1)和 2 个生长素(IAA; c92989\_g1, c77087\_g3)均下调表达。上述结果表明,11 个含 SSR 位点 Unigenes 可能参与了马尾松干旱响应过程。一方面马尾松光合作用明显减弱,生理、生长变缓;另一方面,马尾松启动干旱防御保护机制,通过上调 JA 表达、下调 PP2C 表达,延迟干旱损伤。结合转录组 SSR 位点数据,筛选出上述 11 个重要干旱响应基因的 SSR 位点信息。

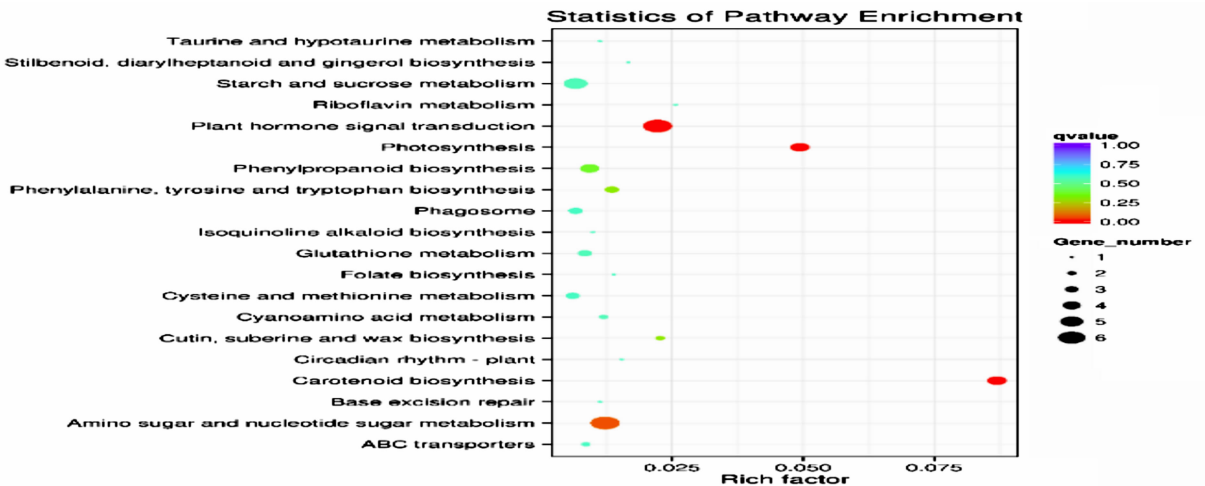


图 6 含 SSR 位点的差异 Unigene 的 KEGG 显著性富集分析

Fig. 6 Significant enrichment KEGG pathways of differentially expressed genes contained SSR

## 2.4 qRT-PCR 验证

qRT-PCR 结果(图 7)显示,随干旱持续,2 个 Unigenes (c71819\_g3, c85755\_g1) 的基因表达量呈递减变化,2 个 Unigenes (c95186\_g2, c93699\_g2) 呈先升后降变化。3 个 Unigenes (c71819\_g3, c85755\_g1, c95186\_g2) 的 qRT-PCR 变化与转录水平 DEG 的

变化基本一致;1 个 Unigene (c93699\_g2) 在第 10 天和第 15 天的 qRT-PCR 扩增倍数高于转录水平 DEG 的变化倍数,但二者变化趋势一致;说明转录组结果有效可靠。



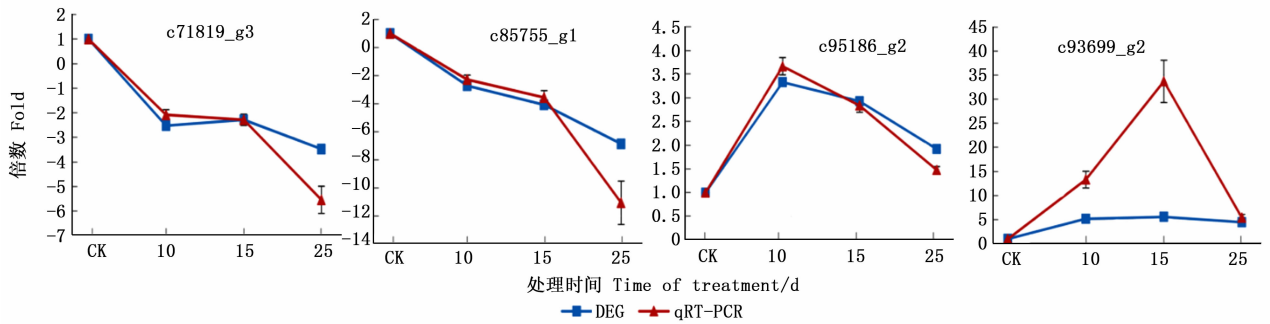


图7 荧光定量 PCR 验证

Fig.7 Real-time PCR Validations of *P. massoniana*

## 3 讨论

### 3.1 转录组序列功能注释

基于高通量测序技术,本研究获得马尾松干旱胁迫转录组 194 821 个 Unigene,通过 Blast 比对,获得 101 806 个 Unigene 序列注释,远高于马尾松均一化测序获得的 33 772 个 Unigene 注释量<sup>[21]</sup>,表明组装效果好,注释信息丰富。NR 数据库 24.42% 的 Unigene (16 323 个)被注释到系统进化关系紧密的云杉属和松属,且在所有注释物种中其匹配数量最多,表明序列注释结果较好,注释成功率较高。

Unigene 功能注释及分类,可初步确定其编码的蛋白质功能,是深入解析转录组信息的前提和基础。本研究中,COG 分类涉及全部 26 个功能类别,表明注释信息全面,几乎涵盖马尾松整个生命过程。GO 分类中,大部分 Unigene 参与初生代谢、细胞结构、生物调控、胁迫刺激响应、信号等生物学过程,表明多数 Unigene 生理活动与干旱响应有关。KEGG 分类中,多数 Unigene 参与次生代谢、植物激素合成、信号转导通路,表明被注释 Unigene 可能参与各类干旱胁迫响应过程。这些 Unigene 的发掘为后续基因功能验证、抗逆机制研究奠定了基础。

### 3.2 马尾松转录组 SSR 特性

基于转录组 194 821 个 Unigene,本研究检测出 SSR 位点平均距离 15.97 kb,出现频率 3.45%;高于马尾松基因组的 3.2%<sup>[2]</sup>,低于马尾松近缘种 EST-SSR 的 4.08%<sup>[17]</sup>。与其他松树相比,高于海岸松 (*Pinus pinaster* Ait.) 的 2.1%<sup>[22]</sup>,低于红松的 4.24%<sup>[5]</sup>。总体上,针叶树 SSR 出现频率及变化幅度较小,说明同类植物变化趋势接近;其差异主要与物种基因组大小、含 SSR 的基因比例以及转录时含 SSR 基因的表达丰度有关。与阔叶树相比,远低于

桉树 (*Eucalyptus robusta* Smith) 的 14.99%<sup>[23]</sup>,说明遗传距离较远的物种,SSR 出现频率差异较大,其差异可能与其进化地位有关。松属起源于 2.5 亿年前的中生代三叠纪<sup>[24]</sup>,远早于起源距今 3 650 ~ 6 500 万年的桉属<sup>[23]</sup>,松属在长期进化积累及自然选择压力下趋于稳定,其基因组进化速度及变异程度相应小于起源较晚的桉属。由此可知,马尾松 SSR 发生频率较低,印证了松属 SSR 分布较低的观点<sup>[25]</sup>。

物种 SSR 重复类型多数以两、三核苷酸为主<sup>[26]</sup>,由于三核苷酸突变不易引起物种突变,面对重大突变压力时,物种更倾向选择三核苷酸<sup>[27]</sup>,且起源越早、压力选择累积越多的物种,其三核苷酸重复类型的富集越明显<sup>[28]</sup>,如起源较早的火炬松 (*Pinus taeda* L.)<sup>[29]</sup>、地中海松<sup>[4]</sup>等。本研究发现,马尾松三核苷酸重复类型频率较高,占 SSR 总数 33.03%,与上述研究结果相似,表明松科植物的自然选择机制具有明显的趋同倾向。此外,马尾松的单核苷酸重复类型频率也较高,占 SSR 总数 35.82%,但与刘公秉<sup>[17]</sup>的 24.49% 差异较大,可能与 SSR 位点重复次数的阈值有关,刘公秉的单核苷酸阈值为 15 次而本研究为 10 次,阈值不同将导致检测到的单核苷酸比例不同。综上所述,马尾松 SSR 重复类型以单、三核苷酸为主。

在数量足够大、无偏倚性的理想情况下,4 种碱基随机组合产生的二至五核苷酸的重复基序分别为 4、10、33、102 种<sup>[30]</sup>。本研究马尾松二~六核苷酸的重复基序分别为 4、10、22、12、20 种,且以 AT、AG、AC、AGC、AAG 为主,存在明显的偏倚性,可能与 SSR 高级基元自身长度的限制有关<sup>[31]</sup>;同时,许多松树如火炬松<sup>[29]</sup>、地中海松<sup>[4]</sup>也存在类似的偏倚性,故这种偏倚性还可能与松属固有的遗传特性有关。

SSR 产生于 DNA 复制过程中的碱基错配,短序列 SSR ( $12 \leq L < 20$  bp) 的碱基错配率小,其突变率远低于长序列 SSR ( $L \geq 20$  bp)<sup>[32]</sup>。本研究中,20 bp 以下基序占 SSR 总数 91.39%,20 bp 以上基序仅占 8.61%,以短序列基序为主,表明马尾松 SSR 具有较强的碱基错配修复能力,为马尾松精准化、规模化 SSR 标记开发提供了重要保障。基于转录组数据大规模设计 SSR 备选引物 13 338 对,随机引物有效扩增率 45.5%,这一结果处于 Bai 等的 50%<sup>[2]</sup> 和刘公秉等的 37.78%<sup>[17]</sup> 之间,表明本研究开发的 SSR 引物具有一定的有效性和通用性。下一步将针对目的基因进行引物筛选,可为马尾松分子辅助育种和遗传多样性研究奠定基础。

### 3.3 马尾松 SSR 位点基因的功能挖掘

转录组 SSR 锚定基因编码序列,表征具体功能,深入挖掘转录组 SSR 标记功能信息,可实现直接与目的性状靶向标记<sup>[6-7]</sup>。为快速搜寻马尾松具有抗旱功能的 SSR 标记,本研究筛选出具有差异表达的 422 个含 SSR 位点 Unigene,并进行 GO 和 KEGG 富集分析,深度挖掘其干旱胁迫下主要的生物过程及生化代谢和信号途径。KEGG 发现,光合作用、类胡萝卜素合成以及植物激素信号传导等 3 个代谢通路被显著富集,涉及 11 个差异表达的 SSR 位点 Unigene,表明其可能与干旱逆境应答直接关联。锁定这 11 个功能表达 Unigene,从转录组中获得对应的 SSR 位点信息,可为后续马尾松抗旱 SSR 标记指纹筛选及抗旱优良种质的选育等研究奠定基础。

编码 JA(c88597\_g2) 和 PP2C(c68631\_g1) 的 2 个 Unigene 在植物激素信号传导通路中被显著富集。JA 作为启动植物防御机制的重要信号,通过上调表达激活防御反应途径,诱导下游防御基因转录或表达,最终产生防御物质以延缓或抵制逆境伤害<sup>[33]</sup>。本研究编码 JA 的 Unigene 亦呈上调表达,并参与了植物激素信号传导的代谢反应。此外,编码 PP2C 的 Unigene 被诱导下调表达,参与 ABA 信号传导,与苜蓿 (*Medicago sativa* Linn)<sup>[34]</sup>、玉米 (*Zea mays* L.)<sup>[35]</sup> 等在逆境胁迫下 PP2C 的变化趋势相同,表明 PP2C 作为重要的功能基因,通过负调节方式直接或间接参与马尾松在干旱胁迫下的信号传导。因此,JA 和 PP2C 可能参与了马尾松的干旱逆境应答。针对这 2 个含 SSR 位点的 Unigene,未来可进一步构建抗旱、不抗旱品系的 F2 群体,将 BSA 法与 SSR 标记分析相结合,深入开展抗旱基因功能

定位研究。

## 4 结 论

本研究从马尾松干旱胁迫转录组中获得 101 806 条具有注释的 Unigene,丰富了马尾松基因信息资源;从 6 367 个 Unigene 中搜寻到 6 728 个 SSR 位点,并初步设计 13 338 对 SSR 标记引物,为马尾松 SSR 分子标记规模化开发及遗传多样性研究奠定了基础;422 个含 SSR 位点的差异表达基因参与了 3 个与干旱响应关联的代谢途径,包括植物激素信号传导、光合作用、类胡萝卜素合成,从中筛选出 11 个重要的 SSR 功能位点,为马尾松抗旱分子机制研究,特别是抗旱功能基因的定位研究奠定了基础。

### 参考文献:

- [1] Pandey G, Misra G, Kumari K, et al. Genome-wide development and use of microsatellite markers for large-scale genotyping applications in foxtail millet [*Setaria italica* (L.)][J]. DNA Research, 2013, 20(2): 197-207.
- [2] Bai T D, Xu L, Xu M, et al. Characterization of masson pine (*Pinus massoniana* Lamb.) microsatellite DNA by 454 genome shotgun sequencing [J]. Tree Genetics & Genomes, 2014, 10(2): 429-437.
- [3] Kalia R K, Rai M K, Kalia S, et al. Microsatellite markers: an overview of the recent progress in plants [J]. Euphytica, 2011, 177(3): 309-334.
- [4] Pinosio S, González-Martínez S C, Bagnoli F, et al. First insights into the transcriptome and development of new genomic tools of a widespread circum-Mediterranean tree species, *Pinus halepensis* Mill [J]. Molecular Ecology Resources, 2014, 14(4): 846-856.
- [5] 张 振, 张含国, 莫 迟, 等. 红松转录组 SSR 分析及 EST-SSR 标记开发 [J]. 林业科学, 2015, 51(8): 114-120.
- [6] Singh R K, Jena S N, Khan S, et al. Development, cross-species / genera transferability of novel EST-SSR markers and their utility in revealing population structure and genetic diversity in sugarcane [J]. Gene, 2013, 524(2): 309-329.
- [7] Choudhary S, Sethy N K, Shokeen B, et al. Development of chickpea EST-SSR markers and analysis of allelic variation across related species [J]. Theoretical and Applied Genetics, 2009, 118(3): 591-608.
- [8] Kantety R V, La Rota M, Matthews D E, et al. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat [J]. Plant Molecular Biology, 2002, 48: 501-510.
- [9] 杨 洋, 田时炳, 王永清, 等. 茄子耐热相关 EST-SSR 分子标记的研究 [J]. 西南农业学报, 2012, 25(5): 1798-1804.
- [10] 曲延英, 穆 平, 李雪琴, 等. 水、旱栽培条件下水稻叶片水势与抗旱性的相关分析及其 QTL 定位 [J]. 作物学报, 2008, 34(2):

- 198-206.
- [11] 刘莹,盖钧镒,吕慧能,等.大豆耐旱种质鉴定和相关根系性状的遗传与QTL定位[J].遗传学报,2005,32(8):856-863.
- [12] 张永虎,陈雪,于海峰,等.向日葵芽期抗旱相关性状丙二醛含量的SSR分子标记分析[J].内蒙古农业大学学报,2012,33(3):25-29.
- [13] 丁贵杰,周志春,王章荣,等.马尾松纸浆用材树种培育与利用[M].北京:中国林业出版,2006:1-34.
- [14] 何卫龙,张逢凯,潘婷,等.利用大配子体构建马尾松遗传图谱[J].分子植物育种,2014,12(3):421-431.
- [15] 杜明凤,丁贵杰.不同种源马尾松ISSR遗传结构及影响因素分析[J].广西植物,2016,36(9):1068-1075.
- [16] Fan F H, Cui B W, Zhang T, et al. LTR-retrotransposon activation, IRAP marker development and its potential in genetic diversity assessment of masson pine (*Pinus massoniana*) [J]. Tree Genetics & Genomes, 2014, 10(1):213-222.
- [17] 刘公秉,季孔庶.基于松树EST序列的马尾松SSR引物开发[J].分子植物育种,2009,7(4):833-838.
- [18] 梅利那,范付华,崔博文,等.基于马尾松转录组的SSR分子标记开发及种质鉴定[J].农业生物技术学报,2017,25(6):991-1002.
- [19] 杜明凤,丁贵杰,赵熙州.不同家系马尾松对持续干旱的响应及抗旱性评价[J].林业科学,2017,53(6):21-29.
- [20] 王艺,丁贵杰.马尾松菌根化苗木对干旱的生理响应及抗旱性评价[J].应用生态学报,2013,24(3):639-645.
- [21] 王晓锋,何卫龙,蔡卫佳,等.马尾松转录组测序和分析[J].分子植物育种,2013,11(3):385-392.
- [22] Chagné D, Chaumeil P, Ramboer A, et al. Cross-species transferability and mapping of genomic and cDNA SSRs in pines [J]. Theoretical and Applied Genetics, 2004, 109(6):1204-1214.
- [23] 阎毛毛,戴晓港,李淑娴,等.松树、杨树及桉树表达基因序列微卫星比对分析[J].基因组学与应用生物学,2011,30(1):103-109.
- [24] 陆素娟,李乡旺.松属的起源、演化及扩散[J].西北林学院学报,1999,14(3):1-5.
- [25] Salzer K, Sebastiani F, Guerli F, et al. Isolation and characterization of polymorphic nuclear microsatellite loci in *Pinus cembra* L. [J]. Molecular Ecology Resources, 2009, 9(3):858-861.
- [26] Liang X, Chen X, Hong Y, et al. Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species [J]. BMC Plant Biology, 2009, 9(1):35.
- [27] Garg R, Patel R K, Tyagi A K, et al. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification [J]. DNA Research, 2011, 18(1):53-63.
- [28] Bérubé Y, Zhuang J, Rungis D, et al. Characterization of EST-SSRs in loblolly pine and spruce [J]. Tree Genetics & Genomes, 2007, 3(3):251-259.
- [29] Echt C S, Saha S, Deemer D L, et al. Microsatellite DNA in genomic survey sequences and unigenes of loblolly pine [J]. Tree Genetics & Genomes, 2011, 7(4):773-780.
- [30] Rota L R, Kantety R V, Yu J K. Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat and barley [J]. BMC Genomics, 2005, 6(2):23.
- [31] 李小白,张明龙,崔海瑞.油菜EST资源的SSR信息分析[J].中国油料作物学报,2007,29(1):20-25.
- [32] Temnykh S, DeClerck G, Lukashova A, et al. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential [J]. Genome Research, 2001, 11(8):1441-1452.
- [33] 张勇,范佳,赵兴延,等.植物防御信号物质JA/SA对桃蚜解毒酶谷胱甘肽-S-转移酶及唾液腺基因C002表达诱导反应[J].中国科学:生命科学,2016,46(5):665-672.
- [34] Millward T A, Zolnierowicz S, Hemmings B A. Regulation of protein kinase cascades by protein phosphatase 2A [J]. Trends in Biochemical Sciences, 1999, 24(5):186-191.
- [35] Liu L, Hu X, Song J, et al. Over-expression of a *Zea mays* L. protein phosphatase 2C gene (*ZmPP2C*) in *Arabidopsis thaliana* decreases tolerance to salt and drought [J]. Journal of Plant Physiology, 2009, 166(5):531-542.

(责任编辑:张研)