

基于 RNA-Seq 技术分析在金属矿上部生长的 芒萁的差异表达基因

邹承武¹, 宋 玮², 宋慈安^{3a,3b,①}, 雷良奇^{3a,3b}

(1. 广西大学农学院, 广西 南宁 530004; 2. 广东工业大学计算机学院, 广东 广州 510006;
3. 桂林理工大学: a. 广西隐伏金属矿产勘查重点实验室, b. 地球科学学院, 广西 桂林 541004)

摘要: 以生长于广西大厂锡多金属矿上部(重金属胁迫区)和未受矿化或污染影响的矿区外围(对照区)的芒萁 [*Dicranopteris pedata* (Houtt.) Nakaike] 为实验材料, 对芒萁叶片进行转录组高通量测序, 并对组装得到的 unigenes 经 NCBI 官方非冗余蛋白质序列数据库(Nr)、NCBI 官方非冗余核苷酸序列数据库(Nt)、KEGG 直系同源数据库(KO)、Swiss-Prot 数据库(Swiss-Prot)、蛋白质家族数据库(Pfam)、基因功能分类体系数据库(GO)和真核生物直系同源序列数据库(KOG)进行注释, 同时分析重金属胁迫区和对照区芒萁叶片间的差异表达 unigenes。结果显示: 测序获得 19.56 Gb clean data, 其中, 重金属胁迫区和对照区芒萁叶片分别含 10.14 和 9.42 Gb clean data。组装得到的 250 582 个 unigenes 中有 120 097 个 unigenes 得到注释, 占 unigenes 总数的 47.93%。与对照区相比较, 重金属胁迫区芒萁叶片中上调和下调差异表达 unigenes 分别有 208 和 620 个, 其中 120 个上调差异表达 unigenes 注释为代谢过程, 占有上调差异表达 unigenes 的 57.69%; 285 个下调差异表达 unigenes 注释为催化活性, 占有下调差异表达 unigenes 的 45.97%。重金属胁迫区芒萁叶片中 15 个 unigenes 与重金属转运和耐受相关, 其中 c44988_g1 和 c84121_g1 的相对表达量分别极显著和显著高于对照区。研究结果显示: 芒萁响应自然金属矿化或矿山重金属污染的基因可以用于生物地球化学找矿和土壤重金属污染检测。

关键词: 芒萁; RNA-Seq; 重金属胁迫; 差异表达基因

中图分类号: Q946-33; Q948.119 文献标志码: A 文章编号: 1674-7895(2017)02-0001-09
DOI: 10.3969/j.issn.1674-7895.2017.02.01

Analysis on differentially expressed genes in *Dicranopteris pedata* grown on metal deposit based on RNA-Seq technology ZOU Chengwu¹, SONG Wei², SONG Cian^{3a,3b,①}, LEI Liangqi^{3a,3b} (1. College of Agriculture, Guangxi University, Nanning 530004, China; 2. School of Computers, Guangdong University of Technology, Guangzhou 510006, China; 3. Guilin University of Technology: a. Guangxi Key Laboratory of Hidden Metallic Ore Deposits Exploration, b. College of Earth Sciences, Guilin 541004, China), *J. Plant Resour. & Environ.*, 2017, 26(2): 1-9

Abstract: Taking *Dicranopteris pedata* (Houtt.) Nakaike grown on tin polymetallic deposits (heavy metal stress area) in Dachang of Guangxi and on the periphery of mining area without mineralization or pollution (the control area) as experimental materials, the transcriptome high-throughput sequencing was conducted on the leaves of *D. pedata*, and assembled unigenes were described by NCBI non-redundant protein sequence database (Nr), NCBI non-redundant nucleotide sequence database (Nt), KEGG Orthology database (KO), Swiss-Prot database (Swiss-Prot), Protein family database (Pfam), Gene Ontology database (GO), and euKaryotic Orthology Groups database (KOG). Meanwhile, differentially expressed unigenes between leaves of *D. pedata* in heavy metal stress area and the control area were analyzed. The results show that 19.56 Gb clean data are acquired by sequencing, in which, leaves of *D.*

收稿日期: 2016-12-06

基金项目: 国家自然科学基金资助项目(41363003; 40972220)

作者简介: 邹承武(1983—), 男, 广西贺州人, 博士, 主要从事植物分子生物学方面的研究。

①通信作者 E-mail: gldysca@126.com

pedata in heavy metal stress area and the control area include 10.14 and 9.42 Gb clean data, respectively. 120 097 unigenes from 250 582 assembled unigenes are described, with 47.93% of total number of unigenes. Compared with the control area, there are 208 and 620 up- and down-regulated differentially expressed unigenes in leaves of *D. pedata* in heavy metal stress area, respectively. In which, 120 up-regulated differentially expressed unigenes are described in metabolism process, accounting for 57.69% of all up-regulated differentially expressed unigenes, and 285 down-regulated differentially expressed unigenes are described in catalytic activity, accounting for 45.97% of all down-regulated differentially expressed unigenes. Fifteen unigenes in leaves of *D. pedata* in heavy metal stress area are associated with heavy metal transport and tolerance, in which relative expression of c44988_g1 and c84121_g1 are extremely significantly and significantly higher than those in the control area, respectively. It is suggested that genes in *D. pedata* in response to natural metal mineralization or heavy metal pollution of mine could be used for biogeochemical prospecting and soil heavy metal pollution detection.

Key words: *Dicranopteris pedata* (Houtt.) Nakaike; RNA-Seq; heavy metal stress; differentially expressed genes

芒萁 [*Dicranopteris pedata* (Houtt.) Nakaike] 隶属于里白科 (Gleicheniaceae) 芒萁属 (*Dicranopteris* Bernh.), 是一种普遍生长于热带、亚热带红壤丘陵、荒坡林缘的蕨类植物, 耐酸、耐旱、耐贫瘠, 属于典型的酸性土壤指示植物^[1]。在金、锡、钨和铈等金属矿床上部生长的芒萁, 其体内与成矿有关的元素明显高于背景区, 产生了生物地球化学异常, 因此可以作为生物地球化学找矿的有效指示植物之一^[2-6]。在受矿区开采污染土壤中生长的芒萁, 其地上部分能吸收大量重金属元素, 属于金属富集型植物, 可用于污染土壤的植物修复^[7-8]。说明芒萁受到高含量金属环境的胁迫后, 体内会聚集大量的金属元素。

大量研究表明: 植物体内聚集的金属元素特别是重金属元素, 可显著影响植物细胞内的基因表达水平^[9-13]。为了揭示芒萁受金属胁迫后的转录调控机制, 明确芒萁响应金属胁迫的相关基因和重要通路, 作者分别采集了生长在广西大厂锡多金属矿上部 and 未受矿化或污染影响的矿区外围的芒萁植株作为实验材料, 进行转录组测序并从头组装 (*de novo* assembly), 分析芒萁体内响应重金属胁迫的相关差异表达基因, 以期进一步研究受重金属胁迫差异表达的基因与自然矿床或开采矿山污染土壤的关系, 并为芒萁应用于生物地球化学找矿和植物修复环境治理提供基础理论依据。

1 材料和方法

1.1 材料

供试芒萁植株分别采自广西大厂锡多金属矿上

部(以下简称为重金属胁迫区)土壤和未受矿化或污染影响的矿区外围(以下简称为对照区)土壤。其中, 重金属胁迫区生长的芒萁植株较对照区矮小, 叶面泛黄伴褐色锈斑。

随机选取重金属胁迫区和对照区同一生长时期的芒萁植物各 3 株, 剪取同一部位、大小一致的新鲜叶片, 同一区域的芒萁叶片各称取 0.1 g, 剪碎并混合均匀, 置于 5 mL 冻存管中, 立即投入液氮中保存, 用于总 RNA 的提取。

在采集芒萁叶片的同时, 采集芒萁植株根际土壤样品并测定土壤中相关矿化元素含量。测定结果表明: 重金属胁迫区土壤中砷、锑、钨、铋、锡、铅、银和镍的含量分别为对照区土壤的 39、38、33、27、26、16、11 和 11 倍。

1.2 方法

1.2.1 总 RNA 提取及检测方法 根据 TRIzol[®] Reagent 试剂(美国 Thermo Fisher Scientific 公司)产品说明书进行实验操作, 提取芒萁叶片总 RNA。每 100 mg 芒萁叶片加入 1 mL TRIzol[®] Reagent, 迅速混匀, 加入 0.2 mL 三氯甲烷, 剧烈震荡 30 s, 室温静置 5 min, 4 °C 条件下 12 000 r · min⁻¹ 离心 15 min。为获得高质量的总 RNA, 避免触碰中间杂蛋白层, 小心地吸取上层水相约 0.6 mL 于 1.5 mL Eppendorf 管中, 加入 10 μL DNase I (RNase free) 消化 DNA, 轻柔混匀 30 s, 室温孵育 10 min, 加入等体积三氯甲烷, 剧烈震荡 30 s, 室温静置 5 min, 4 °C 条件下 12 000 r · min⁻¹ 离心 15 min。小心吸取上层水相 0.4 mL, 加入等体积异丙醇, 室温放置 10 min, 4 °C 条件下 12 000 r · min⁻¹ 离心 10 min, 然后用体积分数 75% 乙

醇清洗沉淀2次,4℃条件下8000 r·min⁻¹离心1 min,弃乙醇,冰上自然干燥沉淀,最后用RNase-free水溶解沉淀。利用Agilent 2100生物分析仪(美国Agilent公司)对总RNA的质量和浓度进行检测和分析。

1.2.2 cDNA文库构建及RNA-Seq测序方法 样品检测合格后,分别取等体积的经DNase I消化的芒萁叶片总RNA,用带有Oligo(dT)的磁珠(美国Thermo Fisher Scientific公司)富集poly(A)mRNA,随后加入5×Fragmentation buffer(美国Illumina公司)在ThermoMixer® F1.5恒温混匀仪(德国Eppendorf公司)中于室温条件下将mRNA打断成短片段,以打断的mRNA为模板,用六碱基随机引物(random hexamers)合成第1链cDNA,然后加入缓冲液、dNTPs、DNA polymerase I和RNase H合成第2链cDNA,再用AMPure XP Beads试剂盒(美国Illumina公司)纯化双链cDNA。纯化后的双链cDNA先进行末端修复、加A尾并连接测序接头,再用AMPure XP Beads试剂盒进行片段大小选择,最后使用ABI 9700型PCR扩增仪(美国Applied Biosystems公司)进行PCR扩增。PCR反应体系总体积为50.0 μL,包含10.0 μL 5×Phusion buffer、1.0 μL PCR Primer PE 1.0、1.0 μL PCR Primer PE 2.0、0.5 μL Phusion DNA Polymerase、0.5 μL 25 mmol·L⁻¹ dNTPs Mix、7.0 μL ddH₂O和30.0 μL已经连接测序接头的连接产物。扩增程序为:98℃预变性30 s;98℃变性10 s,65℃退火30 s,72℃延伸30 s,15个循环;最后在72℃延伸5 min。扩增反应结束后降温至4℃,用AMPure XP Beads试剂盒纯化PCR产物,得到最终的cDNA文库。

cDNA文库构建完成后,先使用Qubit® 2.0荧光定量仪(美国Thermo Fisher Scientific公司)进行初步定量,稀释至1.5 ng·μL⁻¹,随后使用Agilent 2100型生物分析仪(美国Agilent公司)采用DNA 1000分析试剂盒(美国Agilent公司)对cDNA文库的insert size进行检测,insert size符合预期后,使用Illumina qPCR定量操作指南中的qPCR方法对构建的cDNA文库的有效浓度进行准确定量(cDNA文库有效浓度大于2 nmol·L⁻¹),以保证cDNA文库的质量。使用Illumina HiSeq™ 4000平台(美国Illumina公司)对cDNA文库进行测序,测序在北京诺禾致源生物信息科技有限公司完成。

1.2.3 Unigene序列组装 使用组装软件Trinity^[14]将Illumina高通量测序得到的clean reads按顺序进行de novo组装获得unigene。

1.2.4 Unigene序列注释、功能分类及生物学通路分析 使用BLAST软件利用NCBI官方非冗余蛋白质序列数据库(NCBI non-redundant protein sequence database,Nr)、NCBI官方非冗余核苷酸序列数据库(NCBI non-redundant nucleotide sequence database,Nt)、KEGG直系同源数据库(KEGG Orthology database,KO)、Swiss-Prot数据库(Swiss-Prot)、蛋白质家族数据库(Protein family database,Pfam)、基因功能分类体系数据库(Gene Ontology database,GO)和真核生物直系同源序列数据库(euKaryotic Orthology Groups database,KOG)等对unigene进行注释,同时使用Blast2GO^[15]以及Nr注释结果进行GO注释和GO功能分类统计,再联合使用InterProScan 5^[16]进行InterPro注释分析,从而得到unigene的功能注释信息。

1.2.5 差异表达unigene分析 使用RSEM(RNA-Seq by expectation-maximization)^[17]将clean data中的reads映射(mapping)到通过Trinity软件组装获得的转录物上,Bowtie 2^[18]不允许错配,使用DEGseq^[19]进行差异表达unigene检测。所得到的差异表达unigene分别使用goseq软件^[20]和GO富集软件^[21]进行分析。

1.2.6 差异表达unigene的验证 选取转录组测序获得的差异表达unigene,以芒萁的actin基因作为内参基因,根据基因序列使用Primer 6.0软件设计引物并由生工生物工程(上海)股份有限公司合成。使用转录组测序的RNA样品进行反转录获得第1链cDNA,然后通过荧光定量PCR法对选取unigene的表达量进行定量分析。荧光定量PCR实验使用罗氏荧光定量试剂盒(FastStart Universal SYBR Green Master,瑞士Roche公司)并参考说明书进行操作,并在罗氏LightCycler® 480 II实时荧光定量PCR系统(瑞士Roche公司)进行反应。扩增程序为:95℃预变性10 min;95℃变性5 s,72℃退火40 s,45个循环。扩增反应结束后,以芒萁的actin基因进行校正,并将对照区芒萁叶片中待测unigene的转录水平设为“1”,重金属胁迫区芒萁叶片中待测unigene的相对表达量采用2^{-ΔΔC_T}方法^[22]进行计算。

2 结果和分析

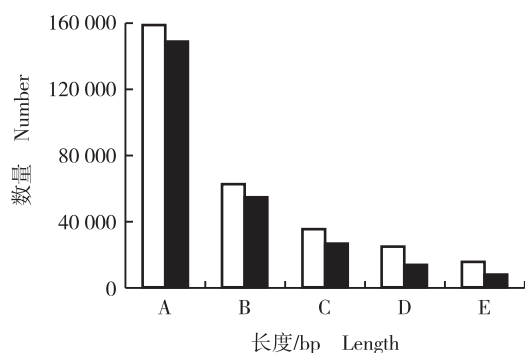
2.1 转录组测序概况及序列组装结果分析

用 Illumina HiSeq™ 4000 平台对构建的芒萁叶片 cDNA 文库进行高通量测序,共获得 135 646 166 条 raw reads,对其进行数据质控,丢弃 5 259 234 条不合格的 reads,获得测序数据 19.59 Gb clean data,其中重金属胁迫区芒萁叶片含 10.14 Gb clean data,对照区芒萁叶片含 9.42 Gb clean data,平均 Q30 分别为 92.94% 和 93.29%。

利用组装软件 Trinity^[14] 对 clean reads 进行混合拼装,获得的转录物平均长度为 567 bp,最大长度为 17 363 bp,选取拼接得到的最长转录物作为 unigenes,获得 250 582 个 unigenes,平均长度为 455 bp,N50 长度为 515 bp,转录物与 unigenes 长度分布见图 1。由图 1 可见:长度为 200~300 bp 的转录物和 unigenes 较多,分别有 158 839 和 149 360 个;长度为 301~500 bp 的转录物和 unigenes 明显减少,分别有 63 027 和 54 594 个;长度为 501~1 000 bp 的转录物和 unigenes 分别有 35 444 和 26 179 个;长度为 1 001~2 000 bp 的转录物和 unigenes 分别有 24 874 和 13 266 个;长度大于 2 000 bp 的转录物和 unigenes 分别有 15 636 和 7 183 个。

2.2 Unigenes 功能注释分析

为了获得芒萁叶片中 unigenes 的功能信息,将组装得到的 250 582 个 unigenes 进行 NCBI 官方非冗余



□: 转录物 Transcript; ■: Unigene.

A: 200-300; B: 301-500; C: 501-1 000; D: 1 001-2 000; E: 大于 2 000 Above 2 000.

图 1 组装获得芒萁叶片中转录物与 unigenes 的长度分布图
Fig. 1 Distribution diagram of length of assembled transcripts and unigenes in leaves of *Dicranopteris pedata* (Houtt.) Nakaike

蛋白质序列数据库(Nr)、NCBI 官方非冗余核苷酸序列数据库(Nt)、KEGG 直系同源数据库(KO)、Swiss-Prot 数据库(Swiss-Prot)、蛋白质家族数据库(Pfam)、基因功能注释数据库(GO)和真核生物直系同源序列数据库(KOG)的 BLAST 注释,各数据库分别注释了 80 536、30 096、37 632、85 035、81 450、83 611 和 52 557 个 unigenes,分别占 unigenes 总数的 32.14%、12.01%、15.02%、33.93%、32.50%、33.37% 和 20.97%。统计结果显示:共有 120 097 个 unigenes 得到注释,占 unigenes 总数的 47.93%;有 130 485 个 unigenes 未被注释,占 unigenes 总数的 52.07%。

GO 数据库对基因及其产物的描述分为生物过程(biological process)、细胞组分(cellular component)和分子功能(molecular function)。对获得 GO 注释的 83 611 个 unigenes 进行分类统计(图 2)。在生物过程中,43 673 个 unigenes 注释为细胞过程(cellular process),44 569 个 unigenes 注释为代谢过程(metabolic process)。在分子功能中,38 014 个 unigenes 注释为催化活性(catalytic activity),38 333 个 unigenes 注释为结合(binding)。

KEGG 是一系列分子作用与互作通路网络,包括新陈代谢(metabolism)、遗传信息处理(genetic information processing)、环境信息处理(environmental information processing)、细胞过程(cellular processes)、生物系统(organismal systems)和人类疾病(human diseases)。为了进一步研究芒萁中活跃的生物学通路,本研究使用 KAAS(KEGG Automatic Annotation Server)^[23] 将 37 632 个 unigenes 定位到 KO 数据库中除人类疾病以外的 5 种分类(图 3)。在二级分类中较为丰富的通路分别为碳水化合物代谢(carbohydrate metabolism)与翻译(translation),分别有 6 007 和 5 282 个 unigenes 参与通路,分别占 KO 数据库注释 unigenes 总数的 15.96% 和 14.04%;氨基酸代谢(amino acid metabolism)通路也较丰富,有 4 095 个 unigenes 参与通路,占 KO 数据库注释 unigenes 总数的 10.88%。这些通路可能在芒萁生化代谢中具有重要作用。

2.3 差异表达 unigenes 检测

以对照区芒萁叶片中 unigenes 表达水平为参照,检测重金属胁迫区芒萁叶片中差异表达 unigenes 的表达情况,选取 $q < 0.005$ 并且 $\log_2(\text{fold change}) > 1$ 的差异表达 unigenes 评估重金属胁迫区和对照区芒萁叶

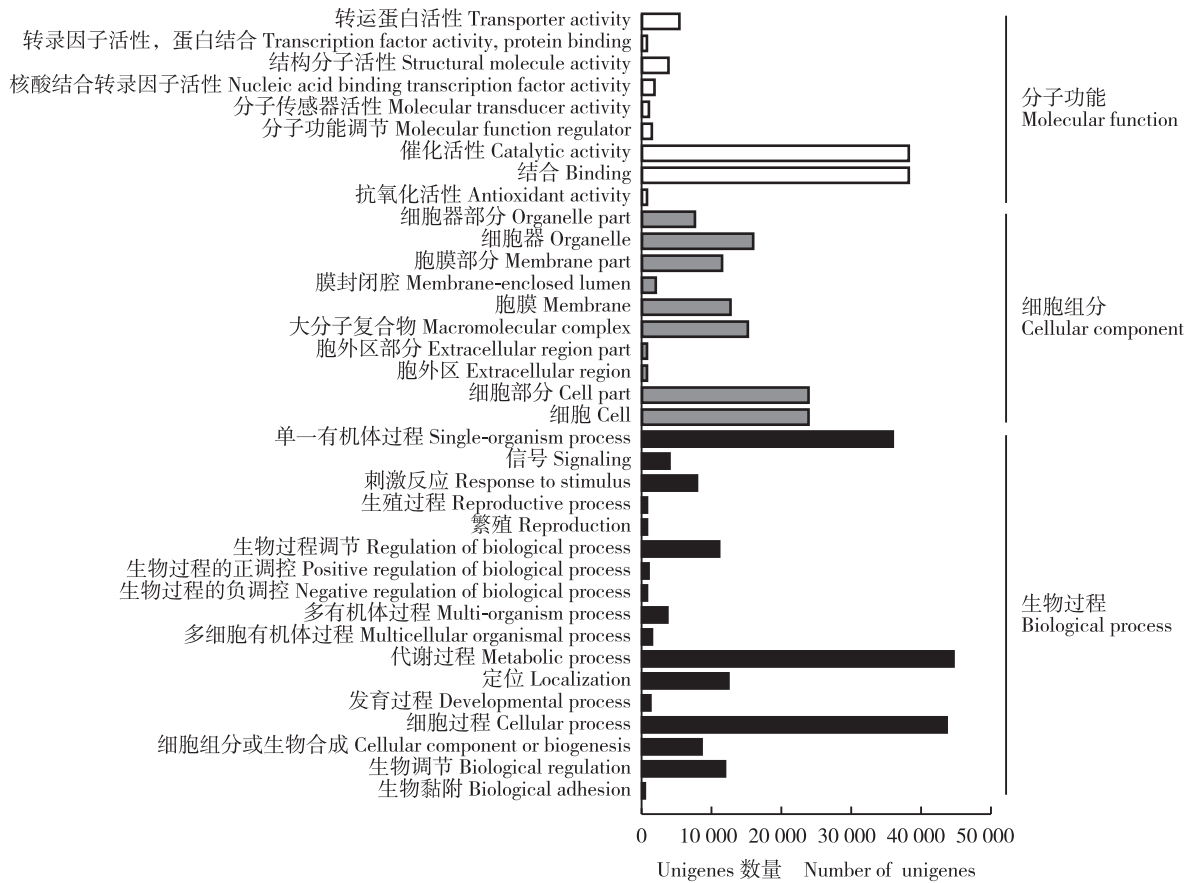
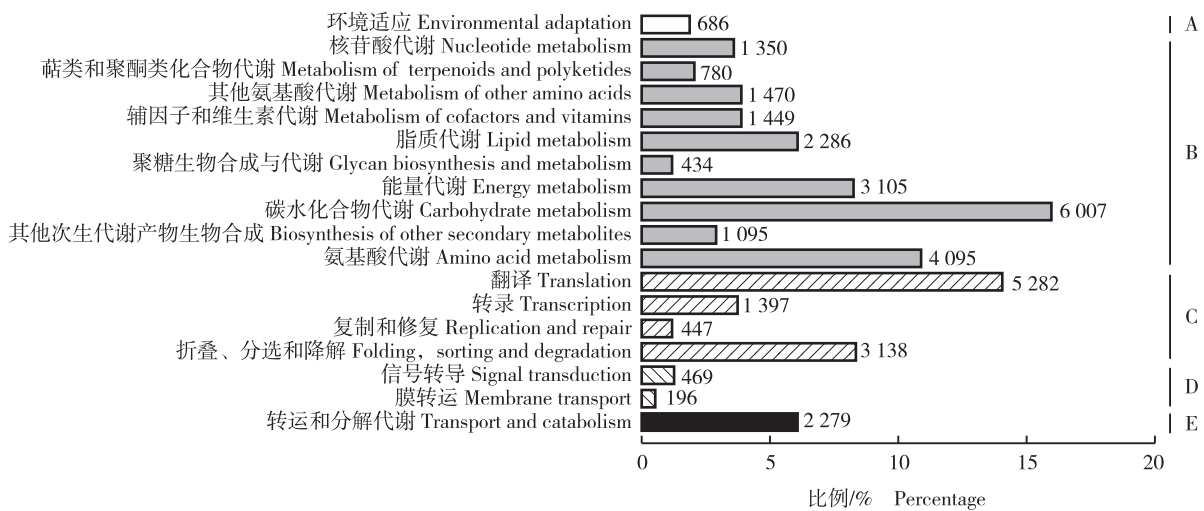


图 2 组装获得的芒萁叶片中 unigenes 的 GO 功能分类结果

Fig. 2 Result of GO function classification of assembled unigenes in leaves of *Dicranopteris pedata* (Houtt.) Nakaike



图中数据表示 unigenes 数量 Datums in the diagram indicate number of unigenes. A: 新陈代谢 Metabolism; B: 生物系统 Organismal systems; C: 遗传信息处理 Genetic information processing; D: 环境信息处理 Environmental information processing; E: 细胞过程 Cellular processes.

图 3 组装获得的芒萁叶片中 unigenes 的 KO 代谢通路分类结果

Fig. 3 Result of KO metabolic pathway classification of assembled unigenes in leaves of *Dicranopteris pedata* (Houtt.) Nakaike

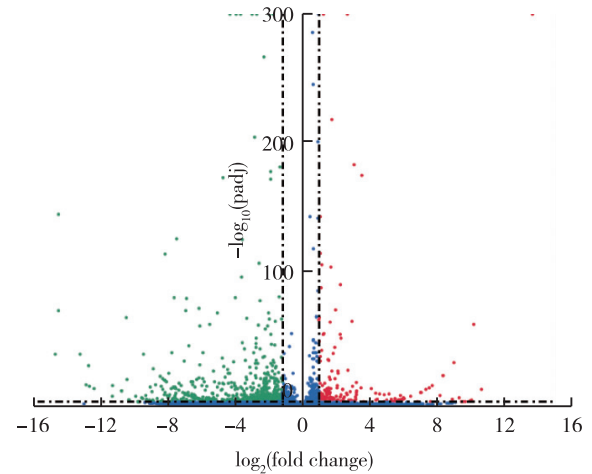
片中的差异表达 unigenes, 结果见图 4。在重金属胁迫区芒萁叶片中上调差异表达 unigenes 有 208 个, 下调差异表达 unigenes 有 620 个, 下调差异表达 unigenes 是上调差异表达 unigenes 的 2.98 倍。

2.4 差异表达 unigenes 的 GO 功能分类分析

对重金属胁迫区芒萁叶片中上调和下调差异表达 unigenes 进行 GO 功能分类富集分析, 结果分别见图 5 和图 6。上调差异表达 unigenes 中有 28 个富集的 GO 条目, 其中有 120 个差异表达 unigenes 注释为代谢过程 (metabolic process), 占有上调差异表达 unigenes 的 57.69%。下调差异表达 unigenes 中有 18 个富集的 GO 条目, 下调差异表达 unigenes 中有 285 个注释为催化活性 (catalytic activity), 占有下调差异表达 unigenes 的 45.97%。

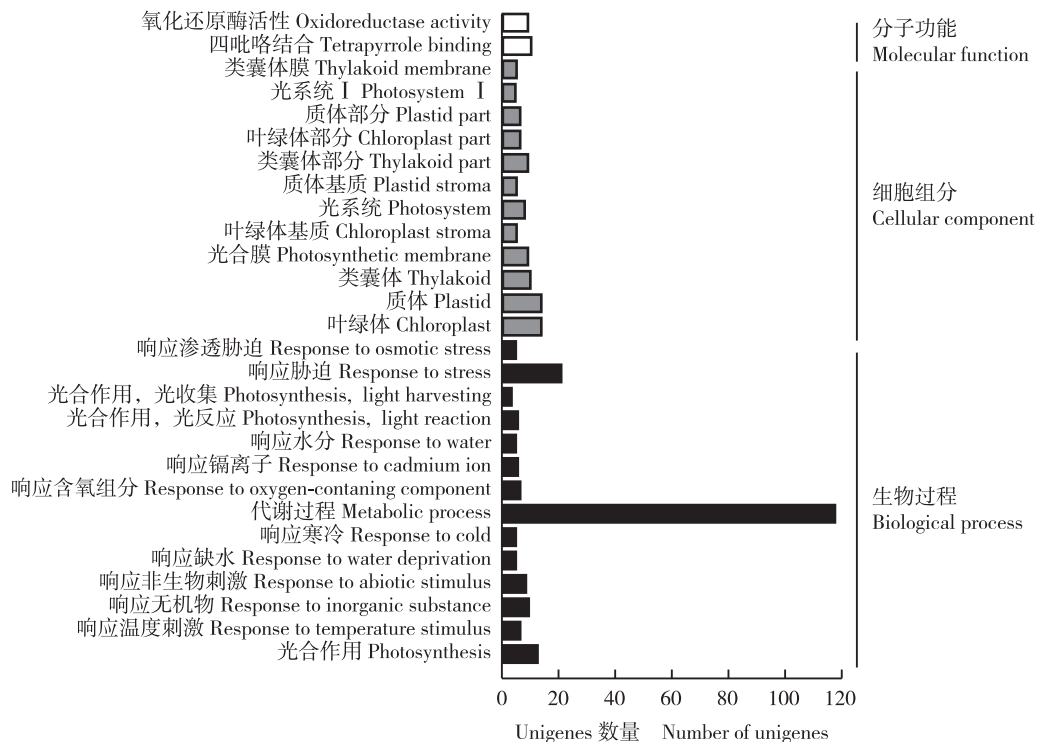
2.5 与重金属转运和耐受相关的 unigenes 及其差异表达情况

对重金属胁迫区芒萁叶片中组装获得的 unigenes 进行 BLASTp 注释, 结果见表 1。结果显示: 15 个 unigenes 与重金属转运和耐受相关。



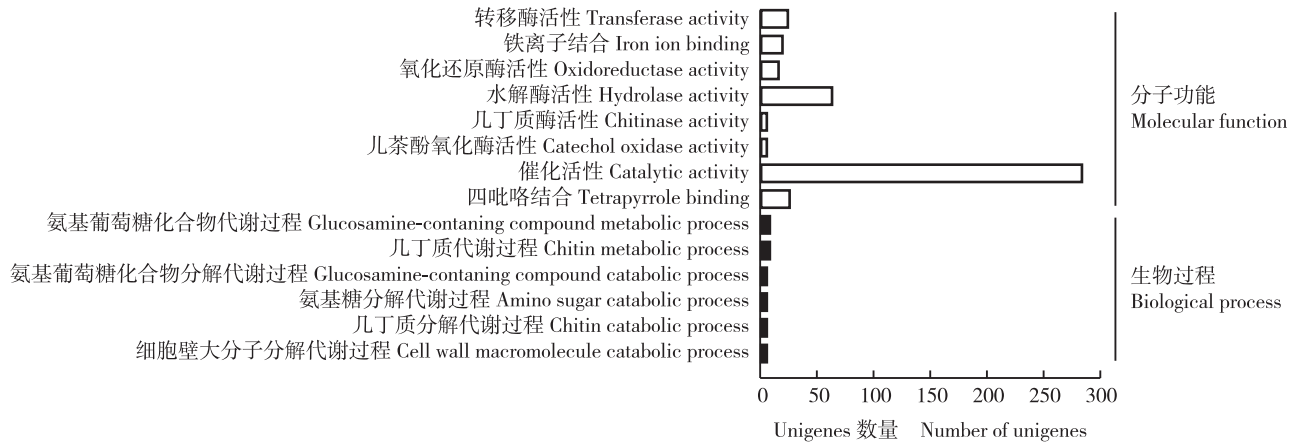
padj: 调整后的 p 值 Adjusted p -value. \cdot : 下调差异表达 unigenes Down-regulated differentially expressed unigenes; \cdot : 除上调和下调以外的差异表达 unigenes Differentially expressed unigenes without up- and down-regulation; \cdot : 上调差异表达 unigenes Up-regulated differentially expressed unigenes. 横向虚线以下及纵向虚线间为无显著性的差异表达 unigenes Differentially expressed unigenes below horizontal dashed line and between vertical dashed lines without significance.

图 4 重金属胁迫区芒萁叶片中差异表达 unigenes 的火山图
Fig. 4 Volcano plot of differentially expressed unigenes in leaves of *Dicranopteris pedata* (Houtt.) Nakaïke in heavy metal stress area



图中不包括被归类为“未知”功能的 unigenes The unigenes classified as function “unknown” are not included in the figure.

图 5 重金属胁迫区芒萁叶片中上调差异表达 unigenes 的 GO 功能分类结果
Fig. 5 Result of GO function classification on up-regulated differentially expressed unigenes in leaves of *Dicranopteris pedata* (Houtt.) Nakaïke in heavy metal stress area



图中不包括被归类为“未知”功能的 unigenes The unigenes classified as function “unknown” were not included in the figure.

图 6 重金属胁迫区芒萁叶片中下调差异表达 unigenes 的 GO 功能分类结果
 Fig. 6 Result of GO function classification on down-regulated differentially expressed unigenes in leaves of *Dicranopteris pedata* (Houtt.) Nakaïke in heavy metal stress area

表 1 重金属胁迫区芒萁叶片中 BLASTp 注释的与重金属转运和耐受相关的 unigenes
 Table 1 Unigenes described by BLASTp and associated with heavy metal transport and tolerance in leaves of *Dicranopteris pedata* (Houtt.) Nakaïke in heavy metal stress area

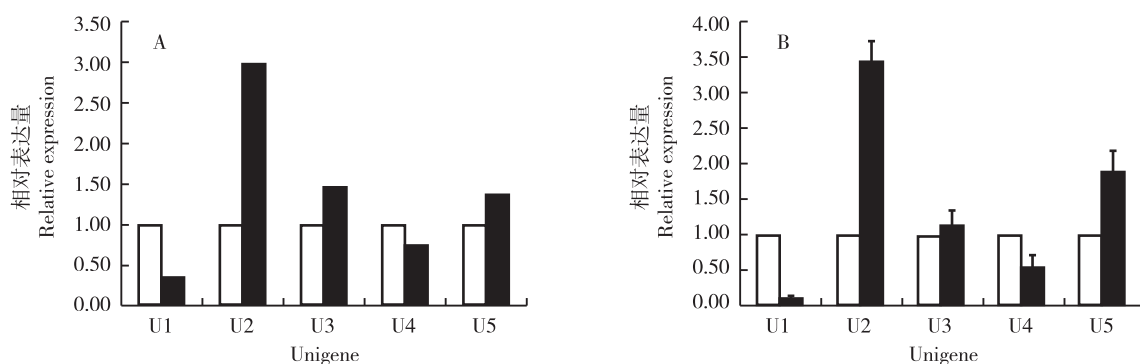
Unigene	序列长度/bp Length of sequence	序列号 Accession number	期望值 Expected value	注释 Description	表达水平 ¹⁾ Regulation ¹⁾
c13489_g1	519	AIG13049.1	2.90E-12	假定重金属转运 P1B-ATPase 2 Putative heavy metal transporting P1B-ATPase 2	UR
c93031_g1	842	KEH18420.1	4.68E-10	重金属相关结构域蛋白 Heavy metal-associated domain protein	UR
c91259_g1	1 516	KEH36986.1	4.17E-24	重金属转运/解毒超家族蛋白 Heavy metal transport/detoxification superfamily protein	UR
c54437_g1	509	KG646770.1	6.93E-57	假定蛋白 Csa_6g133790 Hypothetical protein Csa_6G133790	DR
c107663_g1	237	NP_001052304.1	2.08E-29	重金属相关异戊二烯植物蛋白 26 Heavy metal-associated isoprenylated plant protein 26	S
c11915_g1	722	XP_001697267.1	4.25E-70	一半大小的 ABC 转运蛋白, 膜蛋白 Half-size ABC transporter, membrane protein	UR
c132091_g1	447	XP_001703119.1	7.62E-32	线粒体上的一半大小的 ABC 转运蛋白, 膜蛋白 Mitochondrial half-size ABC transporter, membrane protein	DR
c13999_g1	297	XP_005649995.1	1.19E-11	重金属迁移 Heavy metal translocation	UR
c207040_g1	295	XP_005829762.1	4.15E-11	重金属输出蛋白 hMT1, ABC 超家族 Heavy metal exporter HMT1, ABC superfamily	DR
c44988_g1	1 093	XP_006448292.1	1.44E-16	重金属输出蛋白 hMT1, ABC 超家族 Heavy metal exporter HMT1, ABC superfamily	DR
c92284_g1	1 026	XP_006846435.1	4.41E-35	重金属相关异戊二烯植物蛋白 26 Heavy metal-associated isoprenylated plant protein 26	DR
c84121_g1	1 318	XP_008783549.1	2.91E-34	重金属相关异戊二烯植物蛋白 26 Heavy metal-associated isoprenylated plant protein 26	UR
c177177_g1	241	XP_009348170.1	5.31E-12	重金属耐受样蛋白 Heavy metal tolerance protein-like	UR
c115986_g1	252	XP_010525447.1	1.82E-26	重金属相关异戊二烯植物蛋白 26 Heavy metal-associated isoprenylated plant protein 26	DR
c237853_g1	246	XP_011399202.1	5.66E-18	线粒体 ATP 结合盒亚家族 B 成员 6 ATP-binding cassette sub-family B member 6, mitochondrial	UR

¹⁾ UR: 上调 Up-regulation; S: 相似 Similar; DR: 下调 Down-regulation.

从 15 个与重金属转运和耐受相关的 unigenes 中选择 5 个组装获得的序列长度较长的 unigenes, 分别为 c93031_g1、c91259_g1、c44988_g1、c92284_g1 和 c84121_g1。RNA-Seq 检测结果(图 7-A)显示: c44988_g1 在重金属胁迫区芒萁叶片中的相对表达量只有对照区的 37%, 与对照区差异极显著 ($q < 0.001$); 而 c84121_g1 在重金属胁迫区芒萁叶片中的相对表达量为对照区的 3.2 倍, 与对照区差异显著

($q < 0.005$); 其余 3 个与重金属转运和耐受相关的 unigenes 的相对表达量在重金属胁迫区和对照区间的差异均不显著。

以芒萁的 *actin* 基因为内参基因, 对 5 个 unigenes 的相对表达量进行荧光定量 PCR 检测(图 7-B)。结果显示: 荧光定量 PCR 和 RNA-Seq 检测得到的 5 个 unigenes 的相对表达量存在一定差异, 但是二者反映出的上调和下调的趋势一致。



□: 对照区 The control area; ■: 重金属胁迫区 Heavy metal stress area. U1: c44988-g1; U2: c84121-g1; U3: c91259-g1; U4: c92284-g1; U5: c93031-g1.

A: RNA-Seq 检测 RNA-Seq detection; B: 荧光定量 PCR 检测 Fluorescence quantitative PCR detection.

图 7 重金属胁迫区芒萁叶片中与重金属转运和耐受相关的 unigenes 的差异表达分析
Fig. 7 Differentially expressed analysis on unigenes associated with heavy metal transport and tolerance in leaves of *Dicranopteris pedata* (Houtt.) Nakaike in heavy metal stress area

3 讨 论

不同种类金属胁迫可导致植物某些基因表达水平出现不同的变化^[9-13,24], 这些基因可能潜在指示某种金属的胁迫作用, 或者同种金属不同浓度的胁迫作用, 找到响应金属胁迫的差异表达基因有望为利用植物找矿开辟新途径。用常规的分子生物学手段寻找响应金属胁迫的差异表达基因不仅要求研究对象有基因组数据支持, 而且工作量大, 耗资巨大。RNA-Seq 高通量测序技术, 不但可以快速全面地了解样本间的差异表达基因, 而且测序价格越来越低, 为将来利用该技术寻找指示矿藏的植物基因提供了可能。

芒萁的遗传背景研究较少, 在 NCBI 上尚未找到已经公开报道的芒萁参考基因组或转录组信息。本研究应用 RNA-Seq 高通量测序技术对重金属胁迫区和对照区的芒萁进行转录组测序, 获得 19.56 Gb

clean data, 其中重金属胁迫区和对照区芒萁叶片中分别含 10.14 和 9.42 Gb clean data, 获得的芒萁转录组信息不但能用于寻找响应金属胁迫相关基因, 还为芒萁遗传信息的相关研究提供了数据参考。

将转录组测序分析结果与数字基因表达谱相结合是筛选、分析无参考基因组物种差异表达基因最有效的方法之一。本研究设定对照区芒萁叶片中的基因表达水平为参照来检测重金属胁迫区芒萁叶片中差异表达 unigenes 的表达情况, 重金属胁迫区芒萁样品中上调差异表达 unigenes 有 208 个, 下调差异表达 unigenes 有 620 个, 说明重金属胁迫影响了基因的正常表达水平。推测芒萁能富集和耐受高浓度的重金属元素可能与某些重金属转运和耐受相关的基因调控有关。本研究对重金属胁迫区芒萁叶片中的 unigenes 进行 BLASTp 注释, 明确了这些 unigenes 在细胞代谢过程的具体分类和生物通路, 共有 15 个 unigenes 被注释为与重金属转运和耐受相关, 其中

c44988_g1 和 c84121_g1 的相对表达量分别极显著和显著高于对照区,这2个 unigenes 可以作为潜在的用于植物地球化学找矿的候选基因。

本研究中,芒萁的转录物中仅 47.93% 的 unigenes 被注释,可能由于芒萁的基因组和转录组还没有相关研究,在数据库中没有相应的参考序列,而芒萁作为比较古老的植物,与很多已进行基因组测序的物种亲缘关系较远,所以根据其他物种的基因信息进行注释获得的注释基因数量较少。此外,由于芒萁转录物可能比较复杂,常规组装手段获得的 unigenes 平均长度太短,数量太大,影响后续的数据分析,今后需要进一步优化组装的参数,或采用第三代测序技术,提高测序读长,从而使更多的芒萁基因得到注释。对未能获得注释的响应重金属胁迫的 unigenes 也可以开展基因功能研究,以期获得能够稳定响应某种重金属胁迫的基因用于植物地球化学找矿和土壤重金属污染检测。

综上所述,生长于重金属胁迫区的芒萁与对照区的 unigenes 的表达差异较大,这些差异表达 unigenes 可以作为植物地球化学找矿和土壤重金属污染检测的候选基因,但还需要后续的研究进一步验证。

参考文献:

- [1] 陆树刚. 蕨类植物学[M]. 北京: 高等教育出版社, 2007: 71-74.
- [2] 季峻峰, 崔卫东, 孙承轅. 湖南黄金洞金矿床植物地球化学勘查的初步研究[J]. 物探与化探, 1992, 16(6): 470-473.
- [3] 马跃良. 广东省河台金矿生物地球化学特征及遥感找矿意义[J]. 矿物学报, 2000, 20(1): 80-86.
- [4] 陈代演, 邹振西, 任大银. 植物找矿法在寻找铀矿床中的初步应用[J]. 矿物岩石地球化学通报, 2000, 19(4): 397-400.
- [5] 徐金鸿, 徐瑞松, 夏 斌. 广东鼎湖山斑岩钨矿区生物地球化学特征[J]. 地球与环境, 2006, 34(1): 23-28.
- [6] 陈 杨, 蒙梦平, 宋慈安. 植物对土壤中微量元素的吸收与转移及对生物地球化学异常形成的影响: 以广西盐田岭锡石硫化物矿床为例[J]. 地球与环境, 2012, 40(2): 208-218.
- [7] 刘足根, 杨国华, 杨 帆, 等. 赣南钨矿区土壤重金属含量与植物富集特征[J]. 生态学杂志, 2008, 27(8): 1345-1350.
- [8] 杨胜香, 田启建, 梁士楚, 等. 湘西花垣矿区主要植物种类及优势植物重金属蓄积特征[J]. 环境科学, 2012, 33(6): 2038-2045.
- [9] LOUIE M, KONDOR N, DEWITT J G. Gene expression in cadmium-tolerant *Datura innoxia*: detection and characterization of cDNAs induced in response to Cd²⁺ [J]. Plant Molecular Biology, 2003, 52: 81-89.
- [10] ROUT J R, SAHOO S L. Antioxidant enzyme gene expression in response to copper stress in *Withania somnifera* L. [J]. Plant Growth Regulation, 2013, 71: 95-99.
- [11] HUANG B, XIN J, YANG Z, et al. Suppression subtractive hybridization (SSH)-based method for estimating Cd-induced differences in gene expression at cultivar level and identification of genes induced by Cd in two water spinach cultivars[J]. Journal of Agricultural and Food Chemistry, 2009, 57: 8950-8962.
- [12] TADDEI S, BERNARDI R, SALVINI M, et al. Effect of copper on callus growth and gene expression of *in vitro*-cultured pith explants of *Nicotiana glauca* [J]. Plant Biosystems, 2007, 141: 194-203.
- [13] NEVRTALOVA E, BALOUN J, HUDZIECZEK V, et al. Expression response of duplicated *metallothionein 3* gene to copper stress in *Silene vulgaris* ecotypes [J]. Protoplasma, 2014, 251: 1427-1439.
- [14] GRABHERR M G, HAAS B J, YASSOUR M, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data [J]. Nature Biotechnology, 2013, 29: 644-652.
- [15] CONESA A, GÖTZ S. Blast2GO: a comprehensive suite for functional analysis in plant genomics [J]. International Journal of Plant Genomics, 2008(2008): 619832.
- [16] JONES P, BINNS D, CHANG H Y, et al. InterProScan 5: genome-scale protein function classification [J]. Bioinformatics, 2014, 30: 1236-1240.
- [17] LI B, DEWEY C N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome [J]. BMC Bioinformatics, 2011, 12: 323.
- [18] LANGMEAD B, SALZBERG S L. Fast gapped-read alignment with Bowtie 2 [J]. Nature Methods, 2012, 9: 357-359.
- [19] WANG L, FENG Z, WANG X, et al. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data [J]. Bioinformatics, 2010, 26: 136-138.
- [20] YOUNG M D, WAKEFIELD M J, SMYTH G K, et al. goseq: Gene Ontology testing for RNA-seq datasets [J]. Nature Methods, 2012, 9: 357-359.
- [21] BOYLE E I, WENG S, GOLLUB J, et al. GO::TermFinder: open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes [J]. Bioinformatics, 2004, 20: 3710-3715.
- [22] LIVAK K J, SCHMITTGEN T D. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔC_T} method [J]. Methods, 2001, 25: 402-408.
- [23] MORIYA Y, ITOH M, OKUDA S, et al. KAAS: an automatic genome annotation and pathway reconstruction server [J]. Nucleic Acids Research, 2007, 35: W182-W185.
- [24] GHELICH S, ZARINKAMAR F, SOLTANI B M, et al. Effect of lead treatment on medicarpin accumulation and on the gene expression of key enzymes involved in medicarpin biosynthesis in *Medicago sativa* L. [J]. Environmental Science and Pollution Research, 2014, 21: 14091-14098.

(责任编辑: 张明霞)