

## RNA-Seq 在药用植物研究中的应用

刘厚伯, 上官艳妮, 潘胤池, 赵梓岐, 李 林, 徐德林\*  
遵义医学院 细胞生物学教研室, 贵州 遵义 563099

**摘 要:** 转录组测序技术 (RNA-Seq) 是近 10 年新崛起的一种在组学水平剖析基因功能和互作关系的研究方法, 广泛应用于分子生物学等前沿领域。近年来, 该技术在药用植物研究中得到广泛应用, 相关报道也日渐增多。对相关文献进行系统整理, 归纳 RNA-Seq 技术在药用植物功能基因发掘、基因网络分析、遗传机制揭示、分子标记开发等领域中的应用。同时, 依据 RNA-Seq 的技术特点和药用植物研究的发展需求, 对该技术在中药材中的进一步应用进行展望, 以期为 RNA-Seq 技术应用于中药系统基因组学的研究提供参考。

**关键词:** 转录组测序技术; 药用植物; 系统基因组学; 功能基因挖掘; 基因网络分析; 分子标记开发

**中图分类号:** R282.1      **文献标志码:** A      **文章编号:** 0253 - 2670(2019)21 - 5346 - 09

**DOI:** 10.7501/j.issn.0253-2670.2019.21.031

## Applications of RNA-Seq technology on medicinal plants

LIU Hou-bo, SHANGGUAN Yan-ni, PAN Yin-chi, ZHAO Zi-qi, LI Lin, XU De-lin  
Department of Cell Biology, Zunyi Medical University, Zunyi 563099, China

**Abstract:** RNA-Sequencing (RNA-Seq) is a newly developed method to analyze gene function and interaction at the omics level, which is widely used in frontier field in molecular biology and other fields. In recent years, this technology has been intensively used in the research of medicinal plants, and growing number of reports are published. This paper systematically sorted out relevant literatures and summarized applications of RNA-Seq technology in functional gene discovery, gene network analysis, genetic mechanism revelation and development of molecular marker of medicinal plants. Meanwhile, according to the technical characteristics of RNA-Seq and the development needs of medicinal plant research, this article brings the future prospects regarding RNA-seq technology in Chinese medicinal materials, and intending to provide inspirations for researches on Chinese materia medica based on RNA-Seq.

**Key words:** RNA-Seq; medicinal plant; phylogenomics; functional gene discovery; gene network analysis; development of molecular marker

我国药用植物有 1 万多种, 大多数药用植物遗传背景不清楚, 基因组信息缺乏, 遗传信息和功能基因的研究亦极为薄弱, 而遗传信息的匮乏制约了中药研究和开发。因此, 阐明和完善重要药用植物遗传信息的研究对于医药卫生的发展、药用资源开发等方面具有重要的作用。随着分子生物学技术的高速发展, 将分子生物学技术运用到不同学科门类中, 对于解析不同物种的重要性状特征等具有重要

的作用。分子生药学是在分子水平上研究生药的鉴定、生产和成分的科学<sup>[1]</sup>。运用分子生药学可以最大限度地获得中药材的药效成分, 从而弥补中药材来源不足, 在中药材研究中有十分重要的作用。

转录组测序技术 (RNA-Seq) 是分子生药学研究中的重要技术, 是利用高通量测序技术对组织或细胞中所有 RNA 反转录而成的 cDNA 文库进行测序, 反映其表达水平。目前该技术广泛应用于水稻

收稿日期: 2019-02-01

**基金项目:** 国家自然科学基金项目 (31560079); 国家自然科学基金项目 (31560087); 中医药管理局中医药、民族医药科学技术研究课题 (QZYY-2019-060); 贵州省中药现代化科技产业研究开发专项项目 (黔科合 ZY 字[2013]3002 号); 遵义市 2014 年度“15851”人才工程项目 (201424); 贵州省科技厅人才成长项目 (KY[2017]194); 遵义医学院博士启动基金资助项目 (F-809); 贵州省优秀青年科技人才项目 (黔科合平台人才[2019]5657)

**作者简介:** 刘厚伯, 女, 在读硕士, 研究方向为中药材。E-mail: 1207906601@qq.com

**\*通信作者** 徐德林, 男, 博士, 副教授, 研究方向为中药材遗传育种。E-mail: xudelint2000@163.com

*Oryza sativa* L.<sup>[2]</sup>、玉米 *Zea mays* L.<sup>[3]</sup>、拟南芥 *Arabidopsis thaliana* (L.) Heynh.<sup>[4]</sup>等模式植物中，无需预先设计探针就可以得到较为全面的转录组信息，并进行定位、注释分类以及功能分析。

因此，在中药研究中将分子生物学与 RNA-seq 技术相结合具有重要意义。本文将目前 RNA-Seq 技术在药用植物基因组研究中的应用情况进行介绍，并初步探讨 RNA-Seq 技术在药用植物研究中存在的问题并对今后的发展趋势进行展望。

### 1 RNA-Seq 概念及优缺点

RNA-Seq 是利用高通量测序技术检测信使 RNA (messenger RNA, mRNA)、小 RNA (small RNA) 及非编码 RNA (non-coding RNA) 的序列，并反映出其表达水平。与基因芯片相比，RNA-Seq 无需预先设计探针即可对特定条件下任意物种生长发育阶段整体转录活动进行检测，提供更精确的数字化信号、更高的检测通量以及更广泛的检测范围，同时可以发现新的转录本，因而其成为目前深入研究转录组复杂变化活动的强大且颇具优越性的技术手段<sup>[5]</sup>。

RNA-Seq 具有以下优势：(1) 通量高；(2) 灵敏度高；(3) 分辨率高；(4) 不受限制性<sup>[6]</sup>。RNA-Seq 的发展为相关研究开辟了新的道路，不可否认其也存在一些缺陷，如测序成本过高、如何对获得的大量数据进行分析等亟需解决的问题。但随着技术的逐步完善和广泛应用，这些不足将被有效避免。

### 2 RNA-Seq 技术步骤

RNA-Seq 测序技术步骤可归纳为 3 步：构建测序文库、测序、生物信息分析 (图 1)。

#### 2.1 构建测序文库

首先，用合适的试剂收集要分析的组织样本。

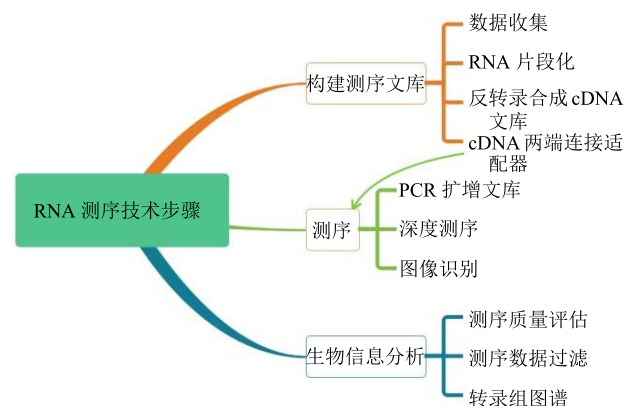


图 1 RNA-Seq 技术基本步骤

Fig. 1 Basic steps of RNA-Seq technology

在采集样本后，通过有机提取或吸附到纺丝柱的硅膜上制备总 RNA。接着，总 RNA 被转换成 cDNA 片段的文库。大多数情况下总 RNA 需要被分离，从中纯化出 mRNA。然后随机打断 mRNA，通过 RNA 片段 (RNA 水解) 制备双链 cDNA 文库，并逆转录成 cDNA。生成片段 cDNA 后，将测序适配器连接到片段的两端。在这个过程中，关于转录本方向的信息完全丢失，但是通过使用硫酸氢钠将胞嘧啶转化为尿苷，可以保持横向信息；然后，生成的 C-T 转换位置标记每个转录本的编码链。

#### 2.2 测序

目前被广泛使用 RNA-seq 技术的平台包括：454 焦磷酸测序系统、ABSOLID 系统和 Illumina 基因组分析仪。扩增测序文库中的 cDNA 并在测序平台上测序，获得大量读段。对有效读段进行序列拼接，获得大量 unigene。

#### 2.3 生物信息分析

对 unigene 进行功能注释、分类、代谢途径分析、SSR 位点搜索等生物信息学分析。对测序结果进行个性化分析，解析某一生命现象或过程的本质<sup>[7]</sup>。在信号处理之后，新一代测序 (next-generation sequencing, NGS) 平台会生成数以百万计的短序列，这些序列与它们的基本调用质量分数相关联，这些质量分数表示每个基本调用的可靠性。这些短序列的长度在 2~450 bp 内，这取决于 NGS 平台的类型。一旦所有的短序列都经过了适当的筛选、映射或装配，就可以对它们进行计数，从而可以从属于某一特定基因外显子的短序列总数中推断出基因的表达水平。通过这种方式，每个碱基都可以得到一个表达分数，从而得到一个基因组规模的转录组图谱<sup>[8]</sup>。

### 3 RNA-Seq 在药用植物研究中的概况

目前已完成转录组测序的药用植物有灵芝 *Ganoderma lucidum* Karst、茯苓 *Poria cocos* (Schw.) Wolf、丹参 *Salvia miltiorrhiza* Bunge、长春花 *Catharanthus roseus* (L.) G. Don、杜仲 *Eucommia ulmoides* Oliver 等，这些转录组序列数据均已公开，这有助于在相关物种上开展基因和启动子挖掘、分子遗传标记开发、基因网络分析、遗传机制分析等研究。然而，药用植物遗传背景复杂、杂合度高，现阶段基因组测序仍然存在技术难度大、费用高等障碍，难以在药用植物上被普遍应用。相反，RNA-Seq 技术相对成熟，目前已有较多成功的应用实例和丰富的公开数据。

通过对目前 RNA-Seq 应用于药用植物研究进展的总结发现,就分析流程和数据处理而论, RNA-Seq 可分为无参考基因组和有参考基因组两大类。因其自身特点, RNA-Seq 常被广泛应用于基因信息十分匮乏的物种。通过对虎杖 *Polygonum cuspidatum* Sieb. et Zucc. 根<sup>[9]</sup>、杜仲 *Eucommia ulmoides* Oliver 果实和叶片<sup>[10]</sup>、丹参 *Salvia miltiorrhiza* Bunge 根<sup>[11]</sup>、铁皮石斛 *Dendrobium officinale* Kimura et Migo<sup>[12]</sup>和人参 *Panax ginseng* C.

A. Mey.<sup>[13]</sup>等进行了转录组测序,发现了一些参与药效成分生物合成相关的转录本和关键酶序列,以及次生代谢产物等。

以实验设计和预期目的而言,药用植物 RNA-Seq 的研究主要涉及以下 4 个方面:(1) 功能基因挖掘;(2) 分子标记开发和应用;(3) 基因网络分析、基因表达模式分析;(4) 遗传机制揭示。RNA-Seq 在药用植物不同领域的研究成果见表 1。

表 1 已报道的药用植物不同领域 RNA-Seq 研究成果

Table 1 Reported research results of RNA-Seq in different fields of medicinal plants

应用领域	物种	测序平台	研究结果	参考文献	
功能基因挖掘	铁皮石斛 <i>Dendrobium officinale</i>	Illumina HiSeq 2000	145 791 unigenes	14	
	丹参 <i>Salvia miltiorrhiza</i>	454GSFLX	13 980 unigenes	15	
	甘草 <i>Glycyrrhiza uralensis</i>	Illumina HiSeq 2000	94 828、305 100 unigenes	16	
	梅冬青 <i>Ilex asprella</i>	Illumina HiSeq 2000	39 269 unigenes	17	
	珠子参 <i>Panax japonicus</i>	Illumina HiSeq 2000	188 914 unigenes	18	
	玄参 <i>Scrophularia ningpoensis</i>	Illumina HiSeq 4000	73 983 unigenes	19	
	锁阳 <i>Cynomorium songaricum</i>	Illumina HiSeq 2000	106 626 unigenes	20	
	银杏 <i>Ginkgo biloba</i>	Illumina Genome Analyzer Iix	32 032 unigenes	21	
	杜仲 <i>Eucommia ulmoides</i>	Illumina HiSeq 2000	96 469 unigenes	22	
	山茱萸 <i>Cornus officinalis</i>	Illumina HiSeq 2000	56 392 unigenes	23	
	冬凌草 <i>Rabdosiarubescens</i>	Illumina HiSeq 4000	3 7 961 unigenes	24	
	阳春砂 <i>Amomum villosum</i>	Illumina HiSeq 2000	144 020 unigenes	25	
	灯盏花 <i>Erigeron breviscapus</i>	Illumina HiSeq 2000	182 527 unigenes	26	
	鹅掌草 <i>Anemone flaccida</i>	Illumina HiSeq 2000	46 962 unigenes	27	
	甘青青蓝 <i>Dracocephalum tanguticum</i>	Illumina HiSeq 4000	151 463 unigenes	28	
	当归 <i>Angelica sinensis</i>	Illumina HiSeq 2000	30 432 unigenes	29	
	分子标记开发和应用	铁皮石斛	Illumina HiSeq 2000	9 892 个 SSR 位点	30
		丹参	Illumina HiSeq 2000	获得 6 058 个能够用于设计引物的 SSR 位点	31
		黄芩 <i>Scutellaria baicalensis</i>	Illumina HiSeq 2500	12 775 个 SSR 位点	32
厚朴 <i>Magnolia officinalis</i>		Illumina HiSeq 2000	8 635 个 SSR 位点	33	
穗花杉 <i>Amentotaxus argotaenia</i>		Illumina HiSeq 2500	2 827 个 SSR 位点	34	
牡丹 <i>Paeonia suffruticosa</i>		Illumina HiSeq 4000	9 939 个 SSR 位点	35	
玄参		Illumina HiSeq 4000	11 659 个 SSR 位点	19	
绞股蓝 <i>Gynostemma pentaphyllum</i>		Illumina HiSeq 2000	3 891 个 SSR 位点	36	
独行菜 <i>Lepidium apetalum</i>		Illumina HiSeq 2000	6 304 个 SSR 位点	37	
半夏 <i>Pinellia ternate</i>		Illumina HiSeq 2000	14 468 SSR 位点	38	
基因网络分析、基因表达模式分析	川牛膝 <i>Cyathula officinalis</i>	Illumina HiSeq 2000	关键化合物合成通路	39	
	甘薯 <i>Dioscorea esculenta</i>	Illumina HiSeq 2000	影响块根花青素含量的重要基因列表	40	
	丹参	Illumina HiSeq 2000	假定酶	41	
遗传机制揭示	龙胆 <i>Gentiana scabrus</i>	Illumina HiSeq 2000	亲缘关系	42	
	铁皮石斛	Illumina HiSeq 2000	基因组	43	
	白及 <i>Bletilla striata</i>	Illumina HiSeq 2000	遗传多样性	44	

### 3.1 功能基因挖掘

功能基因挖掘是挖掘参与调控生长发育和环境相应的转录因子基因以及参与药效成分生物合成相关的转录本和关键酶序列。通过测序平台,基于这些组织已有转录组测序资料,进一步比对数据库中的已知序列,从而挖掘新基因并预测其功能<sup>[16]</sup>。李滢等<sup>[15]</sup>通过 BLAST 与基因本体 (gene ontology, GO) 分析获得了 27 条可能参与丹参酮合成的序列 (编码 15 个关键酶), 29 条参与丹酚酸合成的序列 (编码 11 个关键酶), 70 条细胞色素 P450 序列, 577 条转录因子序列。张春荣等<sup>[16]</sup>应用 Illumina HiSeq 2000 进行转录组双末端测序, 分析甘草根响应适度干旱胁迫的差异表达基因。基于这些组织已有转录组测序资料, 进一步比对数据库中的已知序列, 从而挖掘新基因并预测其功能。朱孝轩等<sup>[45]</sup>通过 Illumina 测序共得到 140 787 121 条短序列, 经 *de novo* 拼接获得 191 167 条 unigenes, 平均长度 674 bp。KEGG 注释鉴定出 604 个与萜类合成有关的基因及 12 个与吲哚生物碱合成有关的基因。张绍鹏等<sup>[18]</sup>采用 Illumina HiSeq 2000 高通量测序获得珠子参根茎的转录组数据。珠子参根茎部表达的 19 个 unigenes 与三萜碳环骨架合成相关; 69 个 unigenes 与介导三萜碳环骨架的氧化和羟基化等修饰的细胞色素 P450 (CYP450) 相关。18 个 unigenes 与负责三萜碳环骨架的糖基化转移酶相关。潘媛等<sup>[19]</sup>应用新一代测序技术 Illumina HiSeq (TM) 4000 在转录水平上对药用植物玄参根部进行测序, 通过 Swiss-Prot、GO、KEGG、COG 比对注释, 发现 520 条编码玄参次生代谢途径关键酶基因和 191 个相关转录因子。张楠等<sup>[21]</sup>采用高通量测序对银杏叶细胞转录组进行测序, 共生成 69 286 个序列, 56 387 个支架, 32 032 个平均长度为 636 bp 的 unigenes。从间隙分布、GC 含量、基因覆盖率等方面评价了 unigene 的质量。朱昀昊等<sup>[46]</sup>采用新一代高通量测序技术对山茱萸果实进行转录组测序, 发现共有 153 条 unigenes 参与山茱萸次生代谢产物的氧化/羟基化; 72 条 unigenes 参与次生代谢产物的糖基化。冯晓宇等<sup>[47]</sup>通过 Illumina 技术对未出土和出土后 2 个时期的花序和茎进行转录组测序, 推断无色花青素还原酶 (LAR) 和二羟黄酮醇-4-还原酶 (DFR) 可能对锁阳儿茶素的合成起着重要的调控作用。李铁柱等<sup>[48]</sup>所构建的杜仲果实和叶片转录组数据库获得了 54 471 338 条序列数据, 获得了包括脂类代

谢、DNA 剪切、植物激素生物合成、苯丙氨酸生物合成、萜类化合物与类固醇类化合物合成等有关转录组。陈延清等<sup>[24]</sup>利用 Illumina HiSeq 4000 高通量测序技术, 得到冬凌草叶和茎转录组有 60 条 unigenes 参与萜类化合物骨架合成、6 条 unigenes 参与各种萜类化合物合成、26 条 unigenes 参与二萜化合物合成途径。

### 3.2 分子标记开发和应用

大多数的药用植物简单重复序列 (SSR) 或单核苷酸多态性 (SNP) 标记相对较为匮乏, 通过 RNA-Seq 获得的数万条 unigene 为 SSR 标记的开发, 以及高测序深度时为 SNP 标记的挖掘均提供了可能。表达序列标签 SSR (expressed sequence tag SSR, EST-SSR) 多态性分子标记, SSR 序列的功能分析为药用植物基因挖掘和分子标记辅助育种提供了有利的工具, 为进一步克隆、研究其功能提供了基础数据, 并为药效成分的生物合成研究奠定了基础。利用 ESTs 技术可发现药效成分相关的新基因<sup>[49]</sup>。Xu 等<sup>[50]</sup>通过微卫星搜索工具 (microsatellite searching tool, MISA) 挖掘, 从 15 433 个 unigenes 中开发出 25 935 对高质量 EST-SSR 引物。在 100 个随机选择的引物中有 87 个被有效扩增; 其中 63 个产生了预期大小的 PCR 产物, 25 个在 4 个地方品种中产生了大量的多态性产物。对 6 个与蝴蝶兰有密切关系的蝴蝶兰属和石斛属的个体进行了 25 个多态性标记的可转移性测试, 在这些群体中, 总共有 5 个标记可以成功扩增。开发的 EST-SSR 引物可促进系统发育研究和育种。王德富等<sup>[51]</sup>利用 Illumina HiSeq TM 2500 对黄芩进行转录组测序和分析, 从黄芩转录组中识别出与黄芩苷生物合成相关的关键基因, 并利用 MISA 软件查找黄芩转录组测序数据, 共发现了 5 658 个 SSR 位点, 而且发现黄芩转录组 SSR 重复碱基以双碱基型重复最多, 占所有 SSR 的 51.75%。代娇等<sup>[33]</sup>利用 Primer03 进行引物设计, 并利用 Blast 软件对含有 SSR 的 unigene 进行功能分析。在 16 369 条厚朴转录组序列中共获得 8 635 个 SSR 位点, 发现了厚朴高通量转录组序列的 SSR 位点具有关型丰富、特异性强和潜能高等特点。王帅等<sup>[34]</sup>利用 Illumina HiSeq2500 技术对穗花杉的茎叶进行转录组测序发现了 20 个与萜类合成有关的 unigenes 和 2 827 个 SSR 位点。谢冬梅等<sup>[35]</sup>采用 Illumina Hi Seq 4000 高通量测序平台对 5 年生“凤丹”(牡丹)根皮转录组进行测序, 在 72 997

条 unigenes 中共检测到 9 939 个 SSR 序列, 其中二核苷酸重复的 SSR 标记占 20.75%。潘媛等<sup>[19]</sup>应用新一代测序技术 Illumina HiSeq (TM) 4000 在转录水平上对药用植物玄参根部进行测序, 利用 MISA 软件在所有 unigenes 中共搜索到 11 659 个 SSR 位点, 重复类型以二核苷酸为主。Zhou 等<sup>[52]</sup>利用 454 焦磷酸测序技术分离了银杏中 30 个新的多态性微卫星位点, 这些多态性的 SSR 将有助于评价银杏种群遗传多样性和资源保护。董林林等<sup>[53]</sup>基于 RNA-Seq 技术检测出三七抗病品种包含 12 个特异 SNP 位点, 经验证 519 688 位点与三七抗根腐病相关, 包含此位点的基因片段可作为抗病品种的遗传标记辅助三七系统选育。逢洪波等<sup>[54]</sup>利用新一代测序技术, 利用 MISA 软件在鬼针草 *Bidens pilosa* L. 中寻找 17 140 个含有 SSR 的重复单元。分子标记是以个体间遗传物质(核苷酸序列)变异为基础的遗传标记, 是 DNA 水平遗传多态性的直接反映。常用的分子标记包括 SSR 和 SNP 等, SNP 被称为第 3 代分子标记, 在基因的精确定位、基因组比较等相关研究中广泛应用。

### 3.3 基因网络分析、基因表达模式分析

全基因组关联分析是指以生物体中的连锁不平衡(linkage disequilibrium, LD)为研究基础, 利用数以百万计的 SNP 标记构建的高密度遗传图谱, 从而探究目标性状和遗传变异间的关系<sup>[40]</sup>。针对个性化的科学研究需要, 对不同组织、不同时期或不同处理下的样本提取 RNA, 构建文库并进行 RNA-Seq 测序, 将原始测序数据比对到参考序列, 获得该转录本各基因的表达量的数字化注释<sup>[55]</sup>。王茹媛等<sup>[40]</sup>对甘薯花青素显著相关转录本(significantly associated transcripts, SATs)进行 KEGG 分析和共表达网络分析, 得到了一系列关键候选基因, 并且通过非同义突变率与同义突变率的比值(KA/KS)分析发现整个花青素通路在物种间受到强烈的纯化选择, 构建了甘薯花青素调控、合成、转运的分子途径并获得了影响块根花青素含量的重要基因列表。杨杰等<sup>[29]</sup>利用 Illumina HiSeq 2000 第 2 代高通量测序技术检测出非模式物种大量功能注释基因, 进一步探测基因组研究缺乏物种功能基因变化的网络模式。对当归功能基因序列进行代谢路径注释, 发现参与次生代谢产物生物合成路径的功能基因序列为 1 159 个, 可能参与当归药用活性物质代谢调控。Luo 等<sup>[41]</sup>利用 Illumina 深度 RNA 测序对茉莉酸

甲酯(MeJA)处理的丹参叶片与模拟处理叶片的转录组进行差异转录谱分析, 结果显示了丹参中的转录复合物, 并提供了有关 MeJA 反应的更多信息; 发现了在 MeJA 诱导下编码具有不同表达模式的关键酶的基因, 总共从约 2 100 万个读数中获得 37 647 个独特序列, 并且基于针对公共数据库的 BLAST 搜索, 注释了 25 641 个(71.53%)这些序列。基于与丹参中柯巴基焦磷酸合酶 1/类考氏合成酶 1(SmCPS1/SmKSL1)和迷迭香酸合酶(SmRAS)基因的共表达模式分析, 从转录组数据中挖掘出的几个候选 P450 基因为丹参酮和酚酸生物合成途径提供了假定酶。闫焱<sup>[39]</sup>通过高通量测序及后续数据处理及分析, 建立起了一套数据预处理技术, 探索了不同数据组装方法对川牛膝小深度全基因组从头测序拼装结果的影响, 并在最终数据中对影响川牛膝品质的关键化合物合成通路上的相关基因进行了电子克隆。基因网络分析通过关联表型数据, 结合功能注释, 对特异表达基因进行筛选, 进而为功能基因挖掘等深入研究奠定基础。这也是当下 RNA-Seq 研究的主体<sup>[55]</sup>。

### 3.4 遗传机制揭示

利用 RNA-Seq 测序技术可对物种进行亲缘关系分析、遗传多样性与居群遗传结构检测、遗传作图、基因定位、分子标记辅助育种等方面的研究。Zhang 等<sup>[42]</sup>利用 Illumina HiSeq 2000 获得了龙胆根和叶转录组的序列和转录丰度数据。在差异表达的基因中, 有 130 多个被标注为与萜类生物合成有关。还有来自不同途径的在根和叶之间分别表达的 3 306 个 unigenes。这一结果代表了龙胆的遗传资源可能为进一步研究龙胆及其亲缘关系提供基础。严亮<sup>[43]</sup>以云南普洱铁皮石斛为材料, 结合第 2 代 Illumina HiSeq 2000 测序技术和第 3 代 PacBio 测序技术, 得到序列  $N_{50}=25.1$  kb, scaffold  $N_{50}=76.4$  kb 的基因组草图, 在基因家族扩张收缩分析中, 与真菌共生和抗旱性相关的基因家族发生了明显扩增。发现与多糖合成相关的蔗糖磷酸合成酶(SPS)和蔗糖合成酶(SUSY)基因发生大规模的复制。黎君等<sup>[44]</sup>基于 Illumina 测序技术构建白及基因组文库和微卫星文库, 设计白及微卫星引物, 表明白及在物种水平均有较高的遗传多样性[等位基因数( $N_a$ )=3.85, Shannon's 信息指数( $I$ )=1.07, Nei's 多样性指数( $H$ )=0.614 7], 白及种群遗传分化强烈[遗传分化系数( $G_{st}$ )=0.43], 居群间的基因流

较弱 [基因流 ( $N_m$ ) = 0.867 6], 居群聚类分析结果均表明地理距离较近的居群具有较近的遗传关系。陈中苏直等<sup>[56]</sup>采用 SSR 分子标记对滇重楼 5 个不同居群共 115 份样品进行遗传多样性分析, 利用 UPGMA 聚类分析, 将 5 个居群分为 2 类, 表明滇重楼遗传多样性水平较高, 居群内和居群间具有一定的遗传分化。运用 RNA-Seq 测序技术, 构建基因文库, 得到基因家族进化树, 从而揭示物种的遗传机制。

#### 4 RNA-Seq 在药用植物研究中的应用展望

本文在简要介绍 RNA-Seq 技术的基础上, 从功能基因发掘、基因网络分析、遗传机制揭示、分子标记开发 4 个方面综述了国内外近年来在药用植物研究中的应用成果。RNA-Seq 主要研究内容有: (1) 检测新的转录本, 包括未知转录本和稀有转录本; (2) 基因转录水平研究, 如基因表达量、不同样本间差异表达; (3) 非编码区域功能研究, 如微小 RNA (microRNA)、非编码长 RNA (lncRNA)、RNA 编辑; (4) 转录本结构研究<sup>[57]</sup>; (5) 转录本结构变异研究, 如可变剪接、基因融合; (6) 开发 SNPs 和 SSR; (7) 转录起始位点的预测; (8) 链特异性测序等<sup>[58]</sup>。RNA-Seq 技术在药用植物应用中有 2 点缺陷: 一是 SSR 出现频率不确定性, 得到的 unigene 序列长度较短, 难以获得同源性比对, 这在一定程度上增加了基因功能注释的难度; 二是基因数据库中的生物信息暂时缺乏, 一些表达不丰富的基因可能无法获得准确的功能注释。

结合 RNA-Seq 的技术优点和其在其他物种中的广泛应用, 还可以深入挖掘该技术在药用植物研究如下几个领域中的应用潜力。

##### 4.1 种质资源保护修复与生产研究

随着环境的变化, 现在许多药用植物的生存环境被严重破坏, 土地被占有使得有效种植地变少; 由于大量竞相采挖, 某些生长期较长的药用植物的野生资源也骤减; 药用价值显著, 但市场缺口较大, 因此, 需要对药用植物进行扩大生产。某些人工栽培药用植物也存在种源混杂、种质退化和栽培技术体系不完善等弊端, 需要对其有效成分含量以及形成机制进行预测与研究, 并对药用植物资源分布以及环境进行保护与修复。因此, 通过 RNA-Seq 技术可获得物种的转录组序列, 以这些转录组序列为基础可对地方品种种质资源进行遗传多样性分析, 从而有针对性地对种质资源进行保护修复。并且可从转录组学角度研究种质资源特性及品质差异, 克隆

并分析生物合成途径中的关键基因, 进而筛选出物种的优质资源, 进行扩大生产。只有品种优良才能在一定区域范围内表现出品质好、长势好、抗逆性强、适应性广、有效成分含量高等优点<sup>[59]</sup>, 才能为药用植物的研究做好铺垫。

##### 4.2 系统基因组学研究

同一生物体中, 不同基因分布在不同部位, 基因家族分布广泛, 种类也很多, 它们对植物的生长发育、胁迫、调控等都有着十分重要的作用。基因家族中基因组织特异性表明它们在不同组织有着不同的功能, 但目前药用植物在生物合成途径和调控方面还多为单基因的研究, 基因网络是依据现有的数据库和已有实验数据构成的网络模型, 所以并不能够反映全部, 存在不同程度的局限性。因此, 可利用 RNA-seq 技术对转录本进行分析, 研究其相关性, 并建立完整的基因网络体系。

##### 4.3 代谢组学与转录组学的结合

植物体内的代谢与调控是一个网络关系, 植物体响应环境因子变化的调节过程不仅作用于单条代谢途径, 也会影响整个代谢网络的平衡。因此, 高通量的组学研究技术可以更系统地观察植物生理变化过程, 对发现新基因、探索生物合成机制起重要的作用<sup>[60]</sup>。Duressa 等<sup>[61]</sup>通过 RNA-seq 分析鉴定出约 50% 显著差异表达的基因, 这些数据对大丽轮枝菌的生物发生和成熟过程中的基因表达和分子过程提供了进一步证据。Kim 等<sup>[62]</sup>使用 RNA-seq 比较了在 2 种不同的培养状态 (孵育后 10 h 和 24 h, 对应指数期和静止期) 下伯克氏病菌 BSR3 与 QS 缺陷的伯克氏病菌 BSR3 突变体之间依赖于 QS 的基因表达。Zhang 等<sup>[63]</sup>揭示了由于 MeJA 的干扰引起的主要代谢和转录组调节, 并构建了一个监管网络。同时, 进行了辅酶 A 连接酶 (4CL3) 的功能研究, 以验证网络预测功能。这些发现将有助于更好地了解植物中 MeJA 诱导代谢的机制。由于某些化合物结构复杂, 空间构象特殊, 难以通过一般化学合成获得。并且药用植物的基因组复杂、杂合度高、重复序列含量高。因此, 将高通量转录和代谢分析结合, 可以更好地了解中药中重要生物合成途径的调控机制。

##### 4.4 优良品种选育与分子标记辅助育种

目前药用植物的品种选育包括系统选育、染色体倍性育种、杂交育种、分子育种、诱变育种等。系统选育是品种选育的有效方法, 在药用植物中应

用最广泛,通常采用连续不断地从栽培植物的大田中或杂种后代的试验地里,选择具有优良性状的单株或群体,并用人工的方法加以定向培育,形成优质高产品种<sup>[64]</sup>。优良品种选育是研究药用植物的重要一步,因此在选育时需产量、质量相平衡。某些药用植物虽可大量生产,但其优良品种稀少且遗传多样性,因此还可采用分子标记技术辅助育种。药用植物 DNA 标记辅助育种是以 DNA 多态性为基础,依据分子杂交、聚合酶链式反应、高通量测序等技术,筛选与高产、优质、抗逆等表型关联的 DNA 片段作为标记,辅助新品种的选育<sup>[53]</sup>。利用分子标记技术,探讨药用植物的遗传多样性和遗传分化,可为药用植物物种资源的保护和合理利用提供一些理论依据。随着科学技术的发展,需要不断丰富药用植物的理论体系,不断利用先进的技术与方法,为药用植物的研究发展提供支撑。因此, RNA-Seq 技术将在药用植物优良品种选育与分子标记辅助育种的应用中起到重要作用。

#### 4.5 药用植物道地性机制研究

药用植物的有效成分主要是其次生代谢产物,由于环境、遗传差异引起了不同的转录基因表达,因此,研究药用植物的道地性十分有必要。在药用植物道地性的研究中,关键酶以及基因网络分析是研究的关键,通过道地基因工程改良药用植物品种是研究和应用的热点。但是次生代谢产物生物合成途径较复杂,需要应用生物统计和化学分析等方法进行深入研究。随着科学技术的发展,需要不断丰富药用植物的理论体系,不断加入先进的技术与方法如 RNA-Seq,才能更好地阐明药用植物道地性产生的机制,为药用植物的研究发展提供技术支撑。

#### 5 结语

RNA-Seq 技术作为新一代测序技术,必将成为今后应用最广的分析技术<sup>[65]</sup>。多数药用植物的研究并不局限于单独使用 RNA-Seq 技术,因此,整合转录组、代谢组等多重组学研究,将会推动中药的现代化进程。

#### 参考文献

[1] 时圣明,潘明佳,王洁,等. 分子鉴定技术在中药中的应用 [J]. 中草药, 2016, 47(17): 3121-3126.  
 [2] 周德贵,赵琼一,付崇允,等. 新一代测序技术及其对水稻分子设计育种的影响 [J]. 分子植物育种, 2008, 6(4): 619-630.  
 [3] 陈泽斌,李冰,王定康,等. 应用 Illumina MiSeq 高

通量测序技术分析玉米内生细菌多样性 [J]. 现代食品科技, 2016, 32(2): 113-120.

- [4] 向亚男,黄蕊蕊,顾婷婷,等. 基于 RNA-Seq 的拟南芥不定芽再生过程的基因表达谱分析 [J]. 南京农业大学学报, 2018, 41(2): 308-320.  
 [5] 许波,张伟强,冯晓曦,等. 转录组测序技术在玉米中的应用研究进展 [J]. 玉米科学, 2014, 22(1): 67-72.  
 [6] 张春兰,秦孜娟,王桂芝,等. 转录组与 RNA-Seq 技术 [J]. 生物技术通报, 2012(12): 51-56.  
 [7] 赵振宇,王仕玉,郭凤根,等. 转录组测序及其在药用植物上的应用 [J]. 基因组学与应用生物学, 2017, 36(2): 820-825.  
 [8] Qian X, Ba Y, Zhuang Q F, et al. RNA-Seq technology and its application in fish transcriptomics [J]. OMICS, 2014, 18(2): 98-110.  
 [9] 欧阳蒲月,刘芬,夏黎,等. 基于高通量测序技术的虎杖 EST-SSRs 和 EST-SNPs 的开发及特征分析 [J]. 中药材, 2015, 38(6): 1164-1167.  
 [10] Zhang W R, Li Y, Zhao J, et al. Isolation and characterization of microsatellite markers for *Eucommia ulmoides* (Eucommiaceae), an endangered tree, using next-generation sequencing [J]. Genet Mol Res, 2016, doi: 10.4238/gmr.15027789.  
 [11] Xu H B, Song J Y, Luo H M, et al. Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza* [J]. Mol Plant, 2016, 9(6): 949-952.  
 [12] Liang Y, Xiao W, Hui L, et al. The genome of *Dendrobium officinale* illuminates the biology of the important traditional Chinese orchid herb [J]. Mol Plant, 2015, 8(6): 922-934.  
 [13] Liu S, Wang S, Liu M, et al. De novo sequencing and analysis of the transcriptome of *Panax ginseng* in the leaf-expansion period [J]. Mol Med Rep, 2016, 14(2): 1404-1412.  
 [14] Zhang J, He C, Wu K, et al. Transcriptome analysis of *Dendrobium officinale* and its application to the identification of genes associated with polysaccharide synthesis [J]. Front Plant Sci, 2016, doi: 10.3389/fpls.2016.00005.  
 [15] 李滢,孙超,罗红梅,等. 基于高通量测序 454 GS FLX 的丹参转录组学研究 [J]. 药学报, 2010, 45(4): 524-529.  
 [16] 张春荣,桑雪雨,渠萌,等. 基于转录组测序揭示适度干旱胁迫对甘草根基因表达的调控 [J]. 中国中药杂志, 2015, 40(24): 4817-4823.  
 [17] Zheng X, Xu H, Ma X, et al. Triterpenoid saponin biosynthetic pathway profiling and candidate gene mining of the *Ilex asprella* root using RNA-Seq [J]. Int J Mol Sci,

- 2014, 15(4): 5970-5987.
- [18] 张绍鹏, 金健, 胡炳雄, 等. 珍稀药用植物珠子参的转录组测序及分析 [J]. 中国中药杂志, 2015, 40(11): 2084-2089.
- [19] 潘媛, 陈大霞, 宋旭红, 等. 基于高通量测序的玄参根部转录组学研究及萜类化合物合成相关基因的挖掘 [J]. 中国中药杂志, 2017, 42(13): 2460-2466.
- [20] 斯日格. 基于转录组分析锁阳花序不同发育期花青素合成途径 [D]. 呼和浩特: 内蒙古大学, 2016.
- [21] 张楠, 孙桂玲, 戴均贵, 等. 银杏细胞转录组高通量测序及分析 [J]. 中国生物工程杂志, 2013, 33(5): 112-119.
- [22] Feng Y, Zhang L, Fu J, *et al.* Characterization of glycolytic pathway genes using RNA-Seq in developing kernels of *Eucommia ulmoides* [J]. *J Agric Food Chem*, 2016, 64(18): 3712-3731.
- [23] Hou D Y, Shi L C, Yang M M, *et al.* *De novo* transcriptomic analysis of leaf and fruit tissue of *Cornus officinalis* using Illumina platform [J]. *PLoS One*, 2018, 13(2): e0192610.
- [24] 陈延清, 胡志刚, 刘大会, 等. 药用植物冬凌草高通量转录组测序与分析 [J]. 中国现代中药, 2018, 20(12): 1476-1482.
- [25] 于安民. 基于 RNA-Seq 的阳春砂果实发育过程中糖和萜类代谢的研究 [D]. 广州: 广州中医药大学, 2014.
- [26] Chen R B, Liu J H, Xiao Y, *et al.* Deep sequencing reveals the effect of MeJA on cutellarin biosynthesis in *Erigeron breviscapus* [J]. *PLoS One*, 2015, 10(12): e0143881.
- [27] Zhan C, Li X, Zhao Z, *et al.* Comprehensive analysis of the triterpenoid saponins biosynthetic pathway in *Anemone flaccida* by transcriptome and proteome profiling [J]. *Front Plant Sci*, 2016, doi: 10.3389/fpls.2016.01094.
- [28] Li H, Fu Y, Sun H, *et al.* Transcriptomic analyses reveal biosynthetic genes related to rosmarinic acid in *Dracocephalum tanguticum* [J]. *Sci Rep*, 2017, 7(1):74.
- [29] 杨杰, 王金权, 丁维俊, 等. 基于 Illumina HiSeq 2000 测序技术对当归根的转录组特性研究 [J]. 中草药, 2015, 46(8): 1216-1222.
- [30] 吴超, 彭娟, 向林, 等. 基于高通量测序的铁皮石斛叶片转录组分析 [J]. 分子植物育种, 2016, 14(12): 3334-3346.
- [31] 郭林林. 丹参基因组 SSR 标记的开发及其在连锁图谱构建的应用 [D]. 济南: 山东农业大学, 2016.
- [32] 齐琳洁, 龙平, 蒋超, 等. 黄芩基因组 SSR 分子标记的开发及遗传多样性分析 [J]. 药学学报, 2015, 50(4): 500-505.
- [33] 代娇, 时小东, 顾雨熹, 等. 厚朴转录组 SSR 标记的开发及功能分析 [J]. 中草药, 2017, 48(13): 2726-2732.
- [34] 王帅, 邵芬娟, 李论, 等. 穗花杉的转录组测序及其转录组特性分析 [J]. 林业科学研究, 2017, 30(5):759-764.
- [35] 谢冬梅, 俞年军, 黄璐琦, 等. 基于高通量测序的药用植物“凤丹”根皮的转录组分析 [J]. 中国中药杂志, 2017, 42(15): 2954-2961.
- [36] Zhao Y M, Zhou T, Li Z H, *et al.* Characterization of global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers in two species of *Gynostemma* (Cucurbitaceae) [J]. *Molecules*, 2015, 20(12): 21214-212131.
- [37] 周茜, 赵惠新, 李萍萍, 等. 独行菜种子转录组的高通量测序及分析 [J]. 中国生物工程杂志, 2016, 36(1): 38-46.
- [38] Zhang G H, Jiang N H, Song W L, *et al.* *De novo* sequencing and transcriptome analysis of *Pinellia ternata* identify the candidate genes involved in the biosynthesis of benzoic acid and ephedrine [J]. *Front Plant Sci*, 2016, doi: 10.3389/fpls.2016.01209.
- [39] 闫燊. 川牛膝全基因组高通量测序及初步数据分析 [D]. 成都: 四川农业大学, 2014.
- [40] 王茹媛. 甘薯块根花青素含量的关联转录组学研究及 MYB 基因家族的初步分析 [D]. 徐州: 江苏师范大学, 2017.
- [41] Luo H, Zhu Y, Song J, *et al.* Transcriptional data mining of *Salvia miltiorrhiza* in response to methyl jasmonate to examine the mechanism of bioactive compound biosynthesis and regulation [J]. *Physiol Plantarum*, 2014, 152(2): 241-55.
- [42] Zhang X, Allan A C, Li C, *et al.* *De novo* assembly and characterization of the transcriptome of the Chinese medicinal herb, *Gentiana rigescens* [J]. *Int J Mol Sci*, 2015, 16(5):11550-11573.
- [43] 严亮. 中国传统兰科药用植物铁皮石斛基因组及其生物学特性研究 [D]. 长春: 吉林大学, 2014.
- [44] 黎君, 杨恒, 周天华. 白芨 SSR 引物筛选及群体遗传多样性研究 [J]. 西北植物学报, 2016, 36(7): 1343-1350.
- [45] 朱孝轩. 长春花转录组与萜类吲哚生物碱代谢途径研究 [D]. 北京: 北京协和医学院中国医学科学院, 清华大学医学部, 2015.
- [46] 朱响昊, 董诚明, 郑晓珂, 等. 基于转录组测序的山茱萸次生代谢生物合成相关基因的挖掘 [J]. 中国中药杂志, 2017, 42(2): 213-219.
- [47] 冯晓宇. 锁阳转录组测序与儿茶素合成途径相关基因的分析 [D]. 呼和浩特: 内蒙古大学, 2016.



- [48] 李铁柱, 杜红岩, 刘慧敏, 等. 杜仲果实和叶片转录组数据组装及基因功能注释 [J]. 中南林业科技大学学报, 2012, 32(11): 122-130.
- [49] 吴春颖, 宋经元, 陈士林. 表达序列标签在药用植物研究中的应用 [J]. 中草药, 2008, 39(5): 778-782.
- [50] Xu D L, Chen H B, Aci M, *et al.* *De novo* assembly, characterization and development of EST-SSRs from *Bletilla striata* transcriptomes profiled throughout the whole growing period [J]. *PLoS One*, 2018, 13(10): e0205954.
- [51] 王德富, 成媛媛, 杨 锋, 等. 黄芩高通量转录组测序数据组装和分析 [J]. 山西农业科学, 2016, 44(8): 1065-1072.
- [52] Zhou Q, Wang X X, Xu M, *et al.* Development and characterization of novel microsatellite markers for *Ginkgo biloba* using 454 pyrosequencing [J]. *Genet Mol Res*, 2015, 15(1): 15017385.
- [53] 董林林, 陈中坚, 王 勇, 等. 药用植物 DNA 标记辅助育种 (一): 三七抗病品种选育研究 [J]. 中国中药杂志, 2017, 42(1): 56-62.
- [54] 逢洪波, 解元坤, 张 璐, 等. 基于转录组测序的鬼针草 SSR 标记开发及其应用 [J]. 分子植物育种, 2018, 16(16): 5359-5368.
- [55] 冯 超, 朱长青, 徐昌杰, 等. RNA-Seq 在果树学研究中的应用 [J]. 果树学报, 2014(1): 115-124.
- [56] 陈中苏直, 田 波, 蔡传涛. 基于 SSR 分子标记的滇重楼遗传多样性研究 [J]. 中草药, 2017, 48(9): 1834-1838.
- [57] 祁云霞, 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): 1191-1202.
- [58] 王康宇, 王 义, 孙春玉, 等. RNA 测序的研究进展 [J]. 江苏农业科学, 2014, 42(10): 12-16.
- [59] 张杨红. 浅析道地药材与地理标志保护 [J]. 黑龙江医药, 2013(6): 1034-1036.
- [60] 常家东. 基于转录组与代谢组学研究增加 CO<sub>2</sub> 和升高温度对紫金牛生理和酚类合成途径的影响 [D]. 杭州: 浙江理工大学, 2017.
- [61] Duressa D, Anchieta A, Chen D, *et al.* RNA-seq analyses of gene expression in the microsclerotia of *Verticillium dahliae* [J]. *BMC Genomics*, 2013, 14(1): 607.
- [62] Kim S, Park J, Choi O, *et al.* Investigation of quorum sensing-dependent gene expression in *Burkholderia gladioli* BSR3 through RNA-seq analyses [J]. *J Microbiol Biotechnol*, 2014, 24(12): 1609-1621.
- [63] Zhang L, Chen J, Zhou X, *et al.* Dynamic metabolic and transcriptomic profiling of methyl jasmonate-treated hairy roots reveals synthetic characters and regulators of lignan biosynthesis in *Isatis indigotica* Fort [J]. *Plant Biotechnol J*, 2016, 14(12): 2217-2227.
- [64] 华国栋, 郭兰萍, 黄璐琦, 等. 药用植物品种选育的特殊性及其对策措施 [J]. 资源科学, 2008, 30(5): 754-758.
- [65] 李 慧, 马德志, 姜 明, 等. 传统药用植物转录组研究进展 [J]. 中医药信息, 2018, 35(6): 114-120.