

转录组学在药用植物研究中的应用*

□吴 琼** 孙 超** 陈士林 罗红梅 李 滢 孙永珍 牛云云

(中国医学科学院 药用植物研究所 北京 100193)
(北京协和医学院)

摘 要:药用植物转录组学研究已经成为药用植物基因组研究中最活跃的领域之一。以转录组学的产生、发展和研究方法为基础,介绍了药用植物中转录组学的应用现状,重点介绍了转录组学当前采用的技术手段以及在西洋参、丹参、乌拉尔甘草功能基因组学中的研究进展。同时对转录组学在药用植物中的发展前景进行了讨论。

关键词:药用植物 转录组学 功能基因组学 454 GS FLX 表达序列标签

doi: 10.3969/j.issn.1674-3849.2010.03.030

药源短缺成为限制临床治疗和新药研发的最大瓶颈,因此,开发新的药物来源途径是迫切需要解决的问题。利用现代生物技术和次生代谢工程手段是未来生产药物最有潜力的发展方向。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。例如人参属 (*Panax*) 和红豆杉属 (*Taxus*) 分别作为最重要的人参皂苷和紫杉醇天然来源植物,功能基因的研究亦相当有限。阐明和完善重要药用植物次生代谢产物合成途径及其调控研究,是一项紧迫的任务。功能基因组学方法,尤其转录组学研究在发现药用植物次生代谢产物生物合成关键酶基因、阐明次生代谢途径及调控机制方面显示了重要的应用价值。本文就目前转录组学的研究方法、转录组学在药用植物基因组研究中的应用情况作一介绍。

一、转录组学研究方法

功能基因组学 (Functional Genomics) 是利用结构基因组学提供的信息和产物,以高通量、大规模实验方法,应用统计科学和计算机分析,在基因组或系统水平全面研究功能基因、蛋白质和代谢物,并探索它们之间的相互联系和规律^[1-2]。功能基因组学将生物学研究从单个基因或蛋白质上升至基因组,同时对多个基因和蛋白质进行系统研究。功能基因组学的研究内容,包括基因组多样性研究、基因功能发现、表达调控的研究、突变检测、蛋白质组学、代谢组学研究等等。可以采用高通量模式,在基因组或系统整体水平上对生命活动规律进行阐述。

功能基因组学的一个重要方面是转录组学 (Transcriptomics),此外,还包括蛋白质组学 (Proteomics) 和代谢组学 (Metabolomics)。转录组 (Transcriptome) 的概念由 Velculescu VE 等^[3]首先提出。转录组是指特定细

收稿日期: 2010-06-04

修回日期: 2010-06-13

* 国家自然科学基金面上项目 (30772735): 濒危药用植物川贝母 EST 文库的建立和 hmgr 基因克隆的研究,负责人:陈士林。

** 联系人:吴琼,博士研究生,主要研究方向:药用植物分子生物学,Tel:010-62899727,E-mail:qw@implad.ac.cn;孙超,副研究员,主要研究方向:药用植物分子生物学,Tel:010-62899729,E-mail:csun@implad.ac.cn。

胞在某一功能状态下全部表达的基因总和,代表了每一个基因的身份和表达水平。同一细胞在不同的生长时期及生长环境下,基因表达情况是不完全相同的,具有特定的空间性和时间性特征^[4]。与基因组具有静态实体的特点不同,转录组是受外源和内源因子调控的。因此,它是物种基因组和外部物理特征的动态联系,是反映生物个体在特定器官、组织或某一特定发育、生理阶段细胞中所有基因表达水平的数据。可用来比较不同组织或生理状况下基因表达水平差异,发现与特定生理功能相关的基因,推测未知基因。转录组学是功能基因组学的重要组成部分,是一门在整体水平上研究某一阶段特定组织或细胞中全部转录本的种类、结构和功能,以及转录调控规律的科学^[5]。转录组学从一个细胞或组织基因组的全部 mRNA 水平研究基因表达情况,它能够提供全部基因的表达调节系统和蛋白质的功能、相互作用的信息。转录组学研究作为一种整体的方法,改变了单个基因的研究模式,将基因组学研究带入了高速发展的时代^[6]。

第一个得到描述的转录组是酿酒酵母细胞,共分析了 60,633 个转录本,揭示了 4,665 个基因,其中,1,981 个基因有已知的功能,而 2,684 个基因从未被鉴定过^[3]。随着研究的深入,人们建立了一系列的方法和技术用于转录组学的研究,目前研究转录组主要有以下几种方法:表达序列标签(Expressed Sequence Tag, EST)、基因表达序列分析(Serial Analysis of Gene Expression, SAGE)、基因芯片、cDNA-AFLP、高通量测序(High-Throughput Sequencing)、生物信息学(Bioinformatics)等等^[7-8]。

表达序列标签(EST)是由 mRNA 逆转录产生,从不同组织来源的 cDNA 文库中随机挑取克隆,对其进行大规模测序所获得的部分 cDNA 的 5'或 3'序列,长度一般为 300~500bp。在缺乏全基因组信息的情况下,EST 已经成为一种进行基因序列比较、发现和鉴定表达基因的快捷途径,并广泛用于药用植物功能基因的研究^[7]。EST 序列分析,可以获得药用植物特定组织、特定发育时期的基因表达种类,EST 的数目也反映出特定基因表达的拷贝数。基因表达的丰度越高,相应的 EST 数目也越多。通过对某一 cDNA 文库进行大规模的 EST 测定和聚类分析,可以构建相应的组织和发育阶段的基因表达谱。通常它们适于对高、中丰度的基因进行表达研究。EST 分析的结果

受 cDNA 文库构建中逆转录效率、文库质量和代表性、大规模测序中的测序量及成本等的影响。

基因表达序列分析(SAGE)是一种定量的高通量基因表达分析技术,利用每个转录本 3'末端特定位置的单一序列标签(9~14 个碱基对)为识别标记,代表一种基因,通过酶切分离每个转录本的序列标签,随后将这些短序列连接、克隆和测序,根据特定序列标签出现的次数估计基因表达的丰度。由于每个 SAGE 测序克隆中都含有几十个序列标签,因此其单位测序成本所获得的信息量高,多数为低丰度转录本。但是 SAGE 文库构建流程长,技术要求高,单条短序列标签所含的基因信息很少,不适于对遗传信息缺乏的药用植物作转录组研究。

基因芯片是一种大规模检测基因表达的技术,它是利用固定探针与样品进行分子杂交,根据杂交图谱荧光信息的强弱测定研究基因的表达丰度。按照采用的探针种类和芯片点制方法的不同,基因芯片可分为 cDNA 芯片(Microarray)和原位合成寡核苷酸芯片(Gene Chip)。cDNA 芯片可以直接从 cDNA 文库中克隆探针,EST 是其主要探针来源。寡核苷酸芯片探针设计需要已知基因组序列,序列筛选和寡核苷酸合成较为费时。目前,基因芯片技术在成本和重复性上仍有待改进。

cDNA-AFLP 是在 AFLP (Amplified Fragment Length Polymorphism)的基础上发展起来的 RNA 指纹图谱技术。通过对 cDNA 限制性酶切片段进行选择性的扩增,获得扩增片段的基因表达信息^[9]。cDNA-AFLP 可用于基因序列信息匮乏的药用植物基因组研究、基因表达定量分析、基因的发现与克隆等方面。研究结果表明假阳性低,可获得转录衍生片段,提供更多的编码区信息^[10]。但获得的转录本序列不可能全部都具有 2 个限制性内切酶识别位点,序列不全面以及技术繁琐是其未来要克服的缺点。

高通量测序是对传统测序方法的一次革命性变革。相对于传统的 96 道毛细管测序,高通量测序的特点是测序通量极高,通过一次运行可以获得 0.45Gb 到 200Gb 碱基数据。例如,最新一代高通量罗氏测序仪(454 GS FLX)以焦磷酸测序方法为原理、依靠生物发光进行 DNA 序列分析。运行一个测序反应仅需要约 10 个小时,一次测序可产生 100 万条序列,序列平均长度约 400bp,在第 400 个碱基处的准确度可达 99%以上,数据总量约 500M。具有高通量、

低成本、快速、准确的特点。应用 454 GS FLX 高通量测序技术进行转录组研究,大大提高了测序的效率,降低了实验成本,使我们能够进行药用植物转录组和次生代谢相关基因的研究。

二、转录组学在药用植物中的应用

目前,1/3 以上的临床用药来源于植物提取物或其衍生物。随着分子生物学向各个学科领域的渗透及蛋白质组学和生物信息学的应用,阐明药用植物天然活性成分生物合成途径及其关键酶,实现关键酶基因的克隆与体外高效表达,利用现代生物技术手段及次生代谢工程,大规模生产药用植物的有效成分将成为未来发展方向之一。

通过对物种的转录组进行描述,能够对理解物种的生物学和生物化学的各个方面提供新的信息。虽然药用植物有效成分的化学和药理学研究已经具有良好的基础,但是在天然活性成分生物合成途径和调控方面研究还很薄弱。目前,该领域的研究主要集中在长春花、青蒿、红豆杉和甘草等少数物种,这些研究多采用单基因研究策略。例如 Collu G. 等^[11]将候选细胞色素 P450 基因转化长春花的悬浮细胞,验证了其具有 10-香叶醇羟化酶的催化功能;Seki H. 等^[12]利用昆虫外源体内共表达和酵母体外表达检验酶活的方法鉴定了甘草中的三萜甘草酸合成关键酶基因;Wildung MR & Croteau R. A 采用同源探针筛选太平洋紫杉 cDNA 文库并结合 GC-MS 的方法,鉴定了紫杉醇生物合成中的关键酶 Taxadiene Synthase 基因^[13]。

功能基因组学在模式生物中发展较快,具有结构基因组数据,并且遗传背景清晰。相比之下,药用植物的序列资源缺乏,遗传背景并不清楚。亟需采用高通量方式在全基因组或整体水平上全面分析基因功能,建立基因组和功能表达之间的联系。近年来,药用植物功能基因组学研究得到了较大发展,国内外已经开展了西洋参^[14-16]、甘草^[17]、丹参^[18]等多种药用植物的大规模转录组研究。转录组研究已经成为分离与克隆新基因及基因功能研究的重要手段之一。目前,在多数植物无法进行全基因组序列测定的情况下,EST 成为进行基因序列比较、发现和鉴定表达基因的一种快捷途径。通过构建 cDNA 文库进行 EST 分析的方法,对药用植物不同组织部位、不同生长时期的 EST 进行研究,可有效发掘和鉴定次生代谢产

物生物合成酶的编码基因及其代谢调控相关的基因。Wu Q. 等^[15]通过分析西洋参 3 种不同组织部位的 EST,获得转录组信息,通过序列对比及功能注释,得到了西洋参组织特异性基因表达谱。发现了 24 条可能参与人参皂苷生物合成相关酶的基因,结合 Real-time PCR 等方法从中克隆并鉴定了甲羟戊酸(MVA)途径中的限速酶 3-羟基-3-甲基戊二酰辅酶 A 还原酶(HMGR)基因,建立了西洋参中人参皂苷生物合成途径,为后续完成基因异源表达,采用 LC/MS 鉴定酶活性,明确基因功能奠定基础(图 1)。另外,获得多个参与调控生长发育和环境相应的转录因子基因,以及 EST-SSR 多态性分子标记。Sun C. 等^[16]采用罗氏 454 GS FLX 测序仪测定了西洋参根 20 余万条 EST 序列,序列拼接获得了 8 倍的覆盖度。通过生物信息学分析,获得了人参皂苷生物合成中大部分酶的基因。同时,发现了 150 条 CYP450 和 235 条糖基转移酶候选序列,它们可能参与各种类型人参皂苷的生物合成。通过甲基茉莉酸诱导和空间表达分析锁定了 1 条 CYP450 和 4 条糖基转移酶基因。李滢等^[18]应用新一代高通量测序技术 454 GS FLX Titanium 对 2 年生丹参根的转录组进行测序,研究其基因表达谱、挖掘其功能基因。获得 4 万 6 千余条表达序列标签,所得序列与 GenBank 丹参 EST 合并拼接,获得 1 万 8 千余条 Unigene,其中,454 高通量测序发现了 13980 条新的 Unigene。通过生物信息学分析获得了可能参与丹参酮合成的序列 27 条(编码 15 个关键酶),参与丹酚酸合成的序列 29 条(编码 11 个关键酶),细胞色素 P450 序列 70 条,转录因子序列 577 条。这些基因的发现为丹参酮和丹酚酸类化合物生物合成研究奠定了基础。Li Y. 等^[17]采用 454 GS FLX Titanium 高通量测序平台,获得了 5 万 9 千余条乌拉尔甘草 EST。结合 GenBank 现有的 5 万条乌拉尔甘草 EST,序列拼接共获得了 2 万 7 千余条 Unigene。通过生物信息学的全面分析,发现了甘草甜素生物合成途径中 16 个酶的基因,初步建立了甘草甜素生物合成途径(图 2)。

药用植物功能基因的研究,主要是为了发现药用植物天然活性成分合成功能基因及其表达规律,确定有效药用活性成分的生物合成途径,了解其调控机制,同时研究其基因组多样性,并将所得的序列信息用于品种鉴定、资源保护和扩大、种质繁育等多个方面。针对具有重要经济价值的药用植物和代表

不同次生代谢途径的模式药用植物开展基因组系列研究,例如转录组序列测定、组装和生物信息学分析,对于目前药用植物功能基因研究还很不完善,大多数药用植物缺乏基因组信息的现状具有很大的现实意义。功能基因组学改变了经典的单个基因、蛋白、代谢产物的纵向的研究方式,取而代之的是全面、整体的基因、蛋白质、代谢产物的横向研究方式^[1]。通过分析大多数(如果不是全部)参与特定生物学过程的分子,功基因组学尤其是转录组学研究,提供了生命活动的整体、全面的描述。

三、展 望

随着中药现代化的逐步推进,药用植物功能基因组学研究成为一个新的研究热点。功能基因的研究有望提供扩大天然药物来源的新途径,促进种质资源鉴定、保存、扩大与优良种质选育。但是由于药

用植物功能基因组学的起步较晚,遗传信息相当薄弱,基础数据和基因组背景研究几乎为空白。药用植物基因组学的研究进展缓慢,如何为中药材功能基因组学研究提供一套完整的研究思路,加快药用植物功能基因组学研究的步伐是当前面临的主要问题。

新一代高通量测序技术(如 454 GS FLX)是对传统测序方法的一次巨大的变革。该技术可以取代传统通过构建文库获得表达序列标签的繁琐过程,以双链 cDNA 为样品,通过片断化处理、Emulsion PCR 等简便的步骤进行测序反应,可以在较短的时间内完成海量数据的获取,并且有配套的商业化数据处理软件提供数据的读取和处理。这一技术对于药用植物转录组学研究具有极其重要的意义。在 GS FLX 高通量测序的基础上,结合 EST 分析、分子生物学方法和化学分析技术,阐明药用植物次生代谢化合物

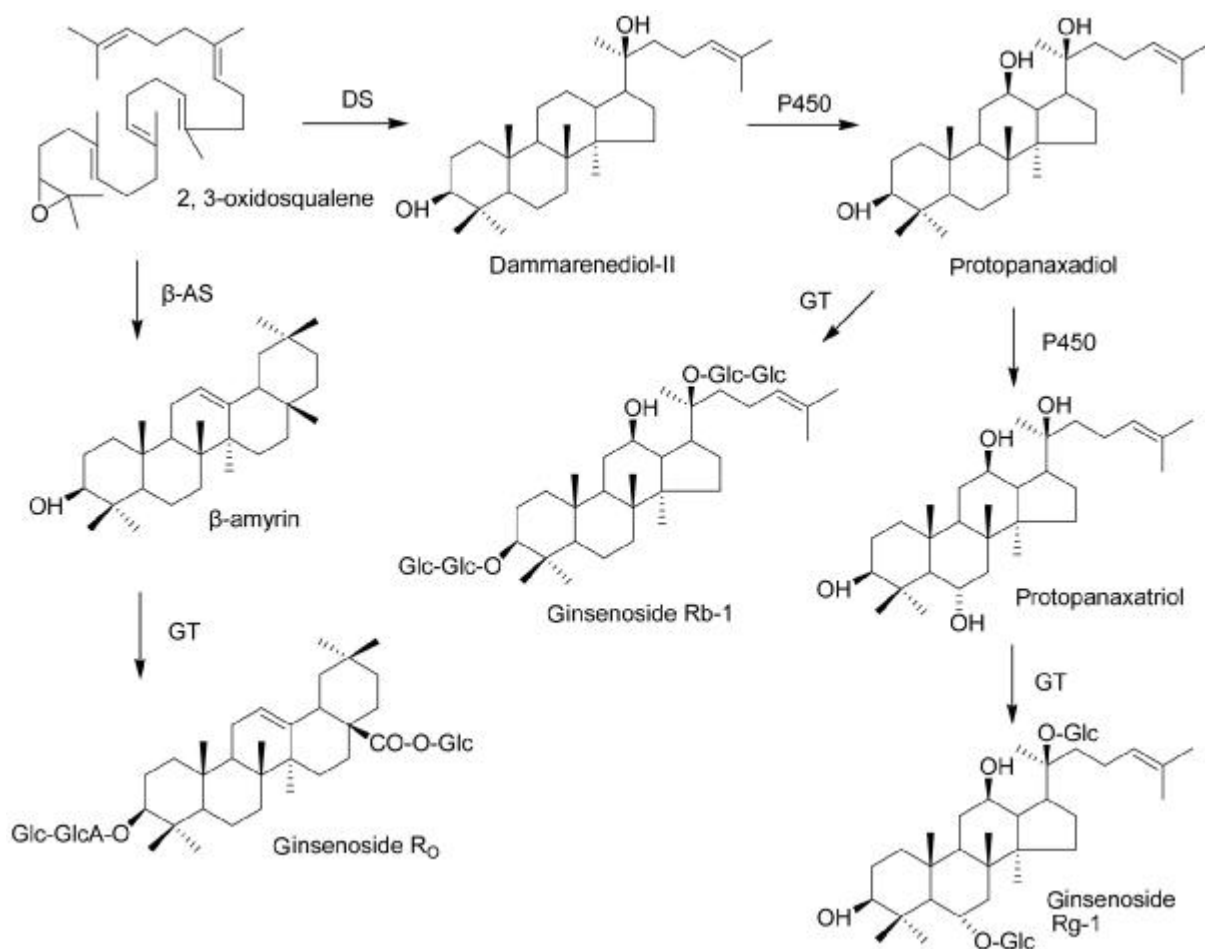


图1 基于 EST 大规模分析构建的人参皂苷生物合成途径

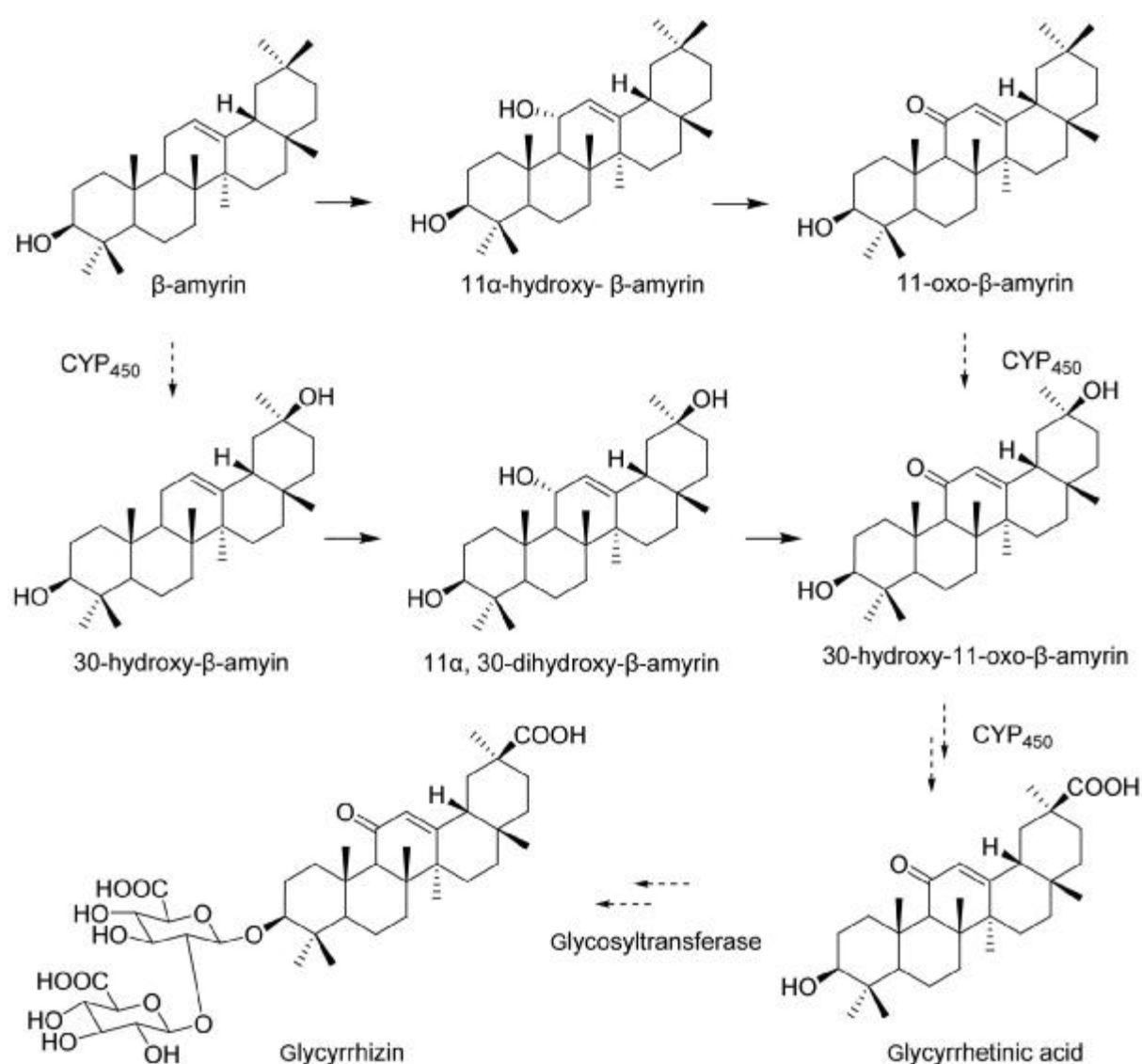


图2 通过高通量测序获得乌拉尔甘草甘草甜素生物合成途径(虚线代表推测的途径)

的生物合成途径,鉴定关键酶的基因,研究酶产物特征,为进一步研究次生代谢调控奠定了基础。对物种的转录组进行描述,能够对理解物种的生物学和生物化学的各个方面提供新的信息,将有力地促进药用植物资源更好地发展和利用。

因此,药用植物转录组研究在基础理论和生产实践中均具有重要的意义。它提供了一种快速、高通量、全面解读药用植物基因组信息的全新技术手段。为功能基因挖掘、药用植物活性成分的生物合成与调控、药用植物种质资源评估与扩大、探索药材道地性分子机制提供了新的思路和方法。通过生物合成途径和基因调控的研究,推动药用植物次生代谢工

程的发展,提高中药材的品质,为中药的良种选育、规范化种植和质量控制提供技术支撑。

参考文献

- 1 Colebatch G, Trevaskis B, Udvardi M. Functional genomics: tools of the trade. *New Phytologist*, 2002,153:27-36.
- 2 Kanehisa M, Goto S, Hattori M, et al. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res*, 2006, 34: D354-D357.
- 3 Velculescu VE, Zhang L, Zhou W, et al. Characterization of the yeast transcriptome. *Cell*, 1997, 88:243-251.
- 4 Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*, 2001, 409(6822):860-921.
- 5 黄琛,武明花,李桂源. 鼻咽癌转录组学研究的现状与进展. *生物化*

- 学与生物物理进展, 2007, 34(11): 1129~1135.
- 6 薛建江, 邱景富. 病原菌感染宿主的转录组学研究进展. 河北北方学院学报(医学版), 2007, 24(5): 63~66.
 - 7 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. 中草药, 2008, 39(5): 778~782.
 - 8 胡松年. 基因表达序列标签(EST)数据. 杭州: 浙江大学出版社, 2005.
 - 9 王勇波, 刘忠, 赵爱华, 等. 功能基因组学方法在药用植物次生代谢物研究中的应用. 中国中药杂志, 2009, 34(1): 6~10.
 - 10 于光, 段金殿. cDNA-AFLP 技术及其在药用植物基因表达研究中的应用展望. 中国现代中药, 2008, 10(8): 5~7.
 - 11 Collua G, Unvera N, Peltenburg-Loomana AMG, et al. Geraniol 10-hydroxylase1, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. FEBS Letters, 2001, 508: 215~220.
 - 12 Seki H, Ohya K, Sawai S, et al. Licorice β -amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. PNAS, 2008, 105(37): 14204~14209.
 - 13 Wildung MR, Croteau R. A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of taxol biosynthesis. J Biol Chem, 1996, 271(16): 9201~9204.
 - 14 陈士林, 孙永巧, 宋经元, 等. 西洋参 cDNA 文库构建及表达序列标签(EST)分析. 药学学报, 2008, 6(43): 657~663.
 - 15 Wu Q, Song JY, Sun YQ, et al. Transcript profiles of *Panax quinquefolius* from flower, leaf and root bring new insights into genes related to ginsenosides biosynthesis and transcriptional regulation. Physiol Plantarum, 2010, 138: 134~149.
 - 16 Sun C, Li Y, Wu Q, et al. De novo sequencing and analysis of American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. BMC Genomics, 2010, 11: 262.
 - 17 Li Y, Luo H, Sun C, et al. EST analysis reveals putative genes involved in glycyrrhizin biosynthesis. BMC Genomics, 2010, 11: 268.
 - 18 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. 药学学报, 2010, 45(4): 524~529.

Application of Transcriptomics in the Studies of Medicinal Plants

Wu Qiong, Sun Chao, Chen Shilin, Luo Hongmei, Li Ying, Sun Yongzhen, Niu Yunyun

(Institute of Medicinal Plant Development, Peking Union Medical College & Chinese Academy of Medical Sciences, Beijing 100193, China)

Abstract: Transcriptomics study concerning medicinal plants has become the most active area in medicinal plant genome research. Here we report the application status of transcriptomics in medicinal plants based on emergence, development and methodology of transcriptomics. We focus on the technical means used in transcriptomics and highlight the progress in the functional genomics in *Panax quinquefolius*, *Salvia miltiorrhiza*, *Glycyrrhiza uralensis*. Finally we discuss the developmental trend of transcriptomics in medicinal plants.

Keywords: Medicinal plant, Transcriptomics, Functional genomics, 454 GS FLX, EST

(责任编辑:王 瑀, 责任译审:张立崑)