

药用植物转录组研究现状与展望

严志祥, 罗茜, 张翼冠, 赵军宁* (四川省中医药转化医学中心, 四川省中医药科学院转化药理与临床应用研究所; 国家中医药管理局中药质量生物评价重点实验室, 四川省道地药材系统开发工程技术研究中心, 中药品质评价与创新中药研究四川省重点实验室, 成都 610041)

摘要:我国是世界上药用植物种质资源最丰富的国家之一。而基于我国数千年生产和实践的道地药材是优质中药材的代名词。研究药用植物药效相关的遗传基因, 尤其是为中医临床所公认的道地药材, 与保护植物种质资源、指导药材采集和生产、揭示药效功能分子机制提供了一个新的思路。随着高通量测序技术的普及和测序成本的降低, 药用植物基因转录组研究得到了较大的发展。笔者围绕转录组学在药用植物研究中的研究现状进行综述和展望, 指出了遗传资源的整合和多学科联合研究在药用植物研究中的重要性。

关键词:药用植物; 中药; 转录组; 遗传资源

doi:10.11669/cpj.2019.07.001 中图分类号:R282 文献标志码:A 文章编号:1001-2494(2019)07-0513-08

A Brief Review on Medicinal Plants Transcriptome Research

YAN Zhi-xiang, LUO Xi, ZHANG Yi-guan, ZHAO Jun-ning* (*Sichuan Institute for Translational Chinese Medicine, Sichuan Academy of Traditional Chinese Medicine Translational Pharmacology and Clinical Application Research Institute, Key Laboratory of Chinese Medicine Quality and Biological Evaluation of State Administration of Traditional Chinese Medicine, Sichuan Genuine Medicinal Materials System Development Engineering Technology Research Center, Sichuan Key Laboratory of Quality Evaluation and Innovation of Traditional Chinese Medicine, Chengdu 610041, China*)

ABSTRACT: China is one of the countries with the most abundant germplasm resources of medicinal plants in the world. Genuine medicinal materials, which originated from thousands of years of production and clinical practice in China, is synonymous with high-quality Chinese medicine. The study on the pharmacodynamic genes of medicinal plants, especially the traditional Chinese medicine, has provided a new idea for the germplasm resources protection, medicinal plant usage optimization and functional molecular discovery. With the advancement of high-throughput sequencing technology and the reduction of the cost of sequencing, the research on the medicinal plants transcriptome has been greatly developed. This article reviewed medicinal plant transcriptome study in recent years and present the importance of large-scale dataset from different field.

KEY WORDS: medicinal plants; traditional Chinese medicine; transcriptome; genetic resources

我国幅员辽阔, 自然条件复杂, 是世界上植物种质资源最丰富的国家之一, 全国已知植物约有 25 700 种, 其中很多植物都具有药用价值。根据 20 世纪 80 年代的系统资源调查发现, 我国有中药材资源 12 807 种, 其中药用植物资源种类有 383 科, 2 309 属, 11 146 种^[1], 在中药材资源总数中大约占 87%。《中国药典》2015 年版收载的植物类中药已达 500 多种, 其中为中医临床所公认的道地药材有 200 余种^[2]。

对药用植物遗传资源的研究是近年来的热点, 尤其是

对药用植物的转录组研究, 2016 年与 2006 年相比, 研究文献增加了 15 倍以上 (774 vs 46, 图 1)。转录组是指特定细胞或组织在某一功能状态或发育阶段下转录出来的所有 RNA 的集合, 包括非编码 RNA 和编码蛋白质的 mRNA^[3]。转录组测序技术 (RNA-Seq) 是利用深度测序技术进行转录组分析的新技术, 能够在单核苷酸水平对任意一个物种的整体转录活动进行检测^[4]。转录组与基因组相比, 在药用植物遗传资源的研究中具有以下特点: ①基因组的组装较为复杂, 成本高, 而转录组在一定程度上反

基金项目:国家重大科技专项“重大新药创制”课题资助 (2011ZX09401-304); 深圳市知识创新计划资助 (基 20160126); 四川省中医药管理局课题资助 (20017Z001)

作者简介:严志祥, 男, 博士 研究方向: 中医药基础研究
新药药理, 道地药材系统研究开发 Tel: (028) 85231378

*** 通讯作者:**赵军宁, 男, 博士, 研究员, 博士生导师 研究方向: 转化医学与
E-mail: zarmy@189.cn

映了基因的信息,且 RNA-Seq 技术不需知道基因组序列便可对任意一个物种的整体转录活动进行检测,而大多数药用植物还缺乏基因组信息,故该技术特别适用于药用植物遗传资源的研究。②转录组与基因组相比,还强调了时间和空间的限制,它反映的是一套基因在不同时空因素下的表达,而道地药材引种到不同地区种植,即使基因背景一致,但因其基因呈现差异性表达,导致次生代谢产物不一致,从而引起道地性差异,故转录组技术特别适用于发现和鉴定与道地药材药效功能相关的基因及研究基因表达的调控机制^[5-6]。

1 药用植物转录组研究现状

1.1 研究对象的概况

近年来,随着高通量测序技术的进步,药用植物功能基因组学研究得到了较大发展,药用植物基因组转录组的研究

团队主要集中在中国、印度、加拿大等国家,涉及包括青蒿 (*Artemisia annua*)、丹参 (*Salvia miltiorrhiza*)、红豆杉 (*Taxus chinensis*)、三七 [*Panax notoginseng* (Burk) F. H. Chen] 等诸多药用植物功能基因研究,国内外关于药用植物的转录组研究见表 1。

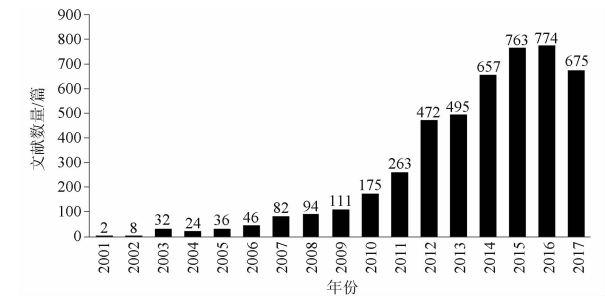


图 1 药用植物转录组文献数量统计图

表 1 国内外代表性药用植物转录组文章信息汇总

药用植物名称	拉丁学名	植物分类	研究团队	发表年份
银杏	<i>Ginkgo biloba</i>	银杏科, 银杏属	Ikhlas A. Khan, Shilin Chen	2011 ^[7]
红豆杉	<i>Taxus chinensis</i>	红豆杉科, 红豆杉属	Long-jiang Yu	2012 ^[8]
东北红豆杉	<i>Taxus cuspidata</i>	红豆杉科, 红豆杉属	Shilin Chen	2011 ^[9]
小叶买麻藤	<i>Gnetum parvifolium</i>	买麻藤科, 买麻藤属	Zeping Jiang, Shengqing Shi	2016 ^[10]
七叶一枝花	<i>Paris polyphylla</i> Smith var. <i>yunnanensis</i> (Franch.) Hand. -Mazz.	百合科, 重楼属	Shengchao Yang	2016 ^[11]
西藏延龄草	<i>Trillium govanianum</i>	藜芦科, 延龄草属	Ram Kumar Sharma	2017 ^[12]
铁皮石斛	<i>Dendrobium officinale</i> Kimura et Migo (Orchidaceae)	兰科, 石斛属	Shilin Chen	2013 ^[13]
魔芋	<i>Amorphophallus</i>	天南星科, 魔芋属	Ying Diao, Zhongli Hu	2013 ^[14]
鱼腥草	<i>Houttuynia cordata</i> Thunb.	三白草科, 蕺菜属	Xianjin Wu	2014 ^[15]
箭叶淫羊藿	<i>Epimedium sagittatum</i> (Sieb. Et Zucc.) Maxim	小檗科, 淫羊藿属	Ying Wang	2010 ^[16]
凤丹	<i>Paeonia suffruticosa</i> cv. FengDan	芍药科, 芍药属	Luqi Huang	2017 ^[17]
罂粟	<i>Papaver somniferum</i>	罂粟科, 罂粟属	Prabodh Kumar Trivedi	2013 ^[18]
博落回	<i>Macleaya cordata</i> and <i>Macleaya microcarpa</i>	罂粟科, 博落回属	Jianguo Zeng, An-Yuan Guo, Xingyao Xiong	2013 ^[19]
萝卜	<i>Raphanus sativus</i>	十字花科, 萝卜属	Maoteng Li	2013 ^[20]
大麻	<i>Cannabis sativa</i>	桑科, 大麻属	Timothy R Hughes, Jonathan E Page	2011 ^[21]
虎杖	<i>Polygonum cuspidatum</i>	蓼科, 虎杖属	HAO DaCheng, Shilin Chen	2015 ^[22]
小蓼	<i>Polygonum minus</i>	蓼科, 蓼属	Hoe-Han Goh	2017 ^[23]
白木香	<i>Aquilaria sinensis</i> (Lour.) Gilg	瑞香科, 沉香属	Jianhe Wei	2012 ^[24]
冬瓜	<i>Benicasa hispida</i>	葫芦科, 冬瓜属	Dasen Xie	2013 ^[25]
蒙古黄芪	<i>Astragalus membranaceus</i> Bge. var. <i>mongolicus</i> (Bge.) Hsiao	豆科, 黄芪属	Xuan Li, Peng Nan	2015 ^[26]
岩蔷薇	<i>Cistus creticus</i> subsp. <i>creticus</i>	半日花科, 岩蔷薇属	Angelos K. Kanellis	2008 ^[27]
沙棘	<i>Hippophae rhamnoides</i> L.	胡颓子科, 沙棘属	Priti Krishna	2012 ^[28]
			Prakash Chand Sharma	2012 ^[29]
			Prakash Chand Sharma	2013 ^[30]
			Prakash Chand Sharma	2014 ^[31]
刺毛黧豆	<i>Mucuna pruriens</i> (L.) DC.	豆科, 黧豆属	N. Sathyanarayana, Ashley N. Egan	2017 ^[32]
牛大力(美丽鸡血藤)	<i>Callerya speciosa</i> (Champ.) ScHot	豆科, 崖豆藤属	Zhiying Li	2016 ^[33]
唐古红景天	<i>Rhodiola algida</i> L.	景天科, 红景天属	Shilong Chen	2014 ^[34]
苦参	<i>Sophora flavescens</i>	豆科, 槐属	Kazuki Saito	2015 ^[35]
甘草	<i>Glycyrrhiza uralensis</i>	蝶形花科, 甘草属	André Steinmetz, Shi-Lin Chen	2010 ^[36]
喜树	<i>Camptotheca acuminata</i>	蓝果树科, 喜树属	Shilin Chen	2011 ^[37]
印度苦楝树	<i>Azadirachta indica</i> A. Juss (neem)	楝科, 苦楝属	Binay Panda	2012 ^[38]
西洋参	<i>Panax quinquefolius</i> L.	五加科, 人参属	Shilin Chen	2010 ^[39]
			Dan Brown	2013 ^[40]
			Tae-Jin Yang	2014 ^[41]

续表 1

药用植物名称	拉丁学名	植物分类	研究团队	发表年份
三七	<i>Panax notoginseng</i> (Burk) F. H. Chen	五加科, 有参属	Shilin Chen	2011 ^[42]
柴胡	<i>Radix bupleuri</i>	伞形科, 柴胡属	Jianhe Wei, Shilin Chen	2015 ^[43]
当归	<i>Angelica sinensis</i>	伞形科, 当归属	Lili Niu	2017 ^[44]
睡茄	<i>Withania somnifera</i>	茄科, 睡茄属	ParulGupta	2015 ^[45]
酸浆果	<i>Physalis peruviana</i>	茄科, 酸浆属	Leonardo Mariño-Ramírez	2012 ^[46]
枸杞	<i>Lycium chinense</i>	茄科, 枸杞属	Ying Wang	2015 ^[47]
胡黄连	<i>Picrorhiza kurroa</i> Royle ex Benth.	玄参科, 胡黄连属	Ravi Shankar, Sanjay Kumar	2012 ^[48]
地黄	<i>Rehmannia glutinosa</i>	玄参科, 地黄属	Xianen Li	2012 ^[49]
毛地黄	<i>Digitalis purpurea</i>	玄参科, 毛地黄属	Fengqing Wang	2017 ^[50]
洋车前子	<i>Plantago ovata</i>	车前科, 车前属	Shilin Chen, Shanfa Lu	2012 ^[51]
红花	<i>Carthamus tinctorius</i> L.	菊科, 红花属	Sanjana Kaul	2016 ^[52]
丹参	<i>Salvia miltiorrhiza</i>	唇形科, 鼠尾草属	Shangqin Hu	2012 ^[53]
			Shilin Chen	2010 ^[54]
			Wang Zhezhi	2011 ^[55]
			Changqing Yang	2013 ^[56]
			Xiu-Jie Wang, Reuben J Peters,	2014 ^[57]
			Luqi Huang	
			Xingfeng Li	2017 ^[58]
穿心莲	<i>Andrographis paniculata</i>	爵床科, 穿心莲属	Dashavantha R. Vudem	2016 ^[59]
贯叶连翘	<i>Hypericum perforatum</i>	藤黄科, 金丝桃属	Zhezhi Wang	2012 ^[60]
牛角瓜	<i>Calotropis procera</i> R. Br.	萝藦科, 牛角瓜属	Pahn-Shick Chang	2015 ^[61]
川西獐牙菜	<i>Swertia mussotii</i> Franch.	龙胆科, 獐牙菜属	Yue Liu, Yi Wang	2017 ^[62]
龙胆草	<i>Gentiana rigescens</i>	龙胆科, 龙胆属	Yuanzhong Wang	2015 ^[63]

另一方面, 药用植物的研究存在极大的不均衡发展状况。我们将《中国药典》2015 年版中所涉及的药用植物, 通过各药用植物的拉丁学名和 Transcriptome 为关键词, 在 Pubmed 上进行文献检索, 有些药用植物的转录组研究的文献较多, 功能基因及次级代谢产物的生物合成途径研究得较为深入, 例如人参、丹参、椒、蓖麻等药用植物出现的转录组文章有数十篇甚至百篇, 而例如山楂、川贝母、川芎、天山雪莲、天冬、天麻、五味子、温郁金、枸杞、半夏、决明、何首乌、苦参等药用植物仅有 1~5 篇, 而有转录组数据的药用植物的品种数量不到百种, 与现存有记录的万种药用植物品种相比仍存在巨大的数据生产空间, 可以说药用植物的遗传数据研究还处于原始积累阶段。

1.2 研究方向的概况

受限于药用植物的遗传资源的匮乏, 当前的转录组研究成果集中在几个主要方向(图 2)。

一个药用植物物种或品种的研究往往经历两个主要的过程, 其一为数据生产阶段, 人类第一次全面解码该物种的遗传信息, 以基因组和转录组数据生产为主, 由少数课题组率先公布遗传框架图和主要的基因预测结果, 其他研究者的后续研究则需要参考遗传框架图和基因功能信息; 其二是数据挖掘阶段, 大量的相近品种、个体的遗传信息得以测序, 为实现特定目的, 大量的实验设计和数据挖掘围绕该物种进行中, 文献举例如下。

1.2.1 转录组测序结果报告 多数药用植物尚未获得遗传信息, 各种药用植物在首次进行转录组测序后在 NCBI 上公布测序结果报告, 为该药用植物的遗传信息提供数据基础。

例: 木橘(*Aegle marmelos*) 是一种重要的药用植物, 许多药用成分都是从叶片中鉴定出来的, 具有抗糖尿病、抗炎、抗微生物、抗脂质和保肝活性等作用, 但转录组信息未见报道。Kaushika 等^[64]利用 Illumina HiSeq2500 平台首次获得了木橘叶片的转录组数据, 共获得 115.92 million 高质量 reads, 200 bp 及以上长度的转录本将用于进一步分析, 木橘叶片样品共产生了 133 616 份三位一体的转录本, 聚集在 46 345 份 unigenes 中, GC 所占的百分比为 40.50%。分析表明, 在被调查的 1 440 个植物核心基因中, 有 1 371 个(95.21%) 完全或部分存在, 而每个基因的邻位基因平均数量为 2.03 个, 且有 64.33% 的基因存在一个以上的邻位基因, 为木橘的遗传信息奠定数据基础。山茱萸(*Cornus officinalis*) 是我国等东亚国家数千年来广泛用于治疗肝肾心血管疾病和频繁排尿等疾病的药用植物之一, 但其遗传学及分子生物学研究尚不深入, 这阻碍了其代谢分子机制研究及有效利用。Hou 等^[65]利用 Illumina HiSeq4000 平台获得了山茱萸的叶片和果实组织转录组信息, 从叶片和果实中分别捕获 57 954 134 和 60 971 652 条 clean reads (GenBank number SRP115440)。这两个文库的 reads 被组装成 56 392 个 unigenes, 平均长度为 856 个 bp。其中, 41 146 个 unigenes 与 NCBI 非冗余蛋白数据库中的序列相匹配。GO 数据库分析, 有 24 336 个 unigenes 参与生物过程(约占 83.26%)、细胞组成(约占 53.58%) 和分子功能(83.93%)。此外, 由 KOG 数据库对 10 808 个 unigenes 进行 KOG 功能分类。通过对 KEGG pathway 数据库的检索发现, 18 435 个 unigenes 被映射到 371 个 KEGG pathway, 此次山茱萸转录组数据为解析遗传信息、分子机制研究奠定基础。

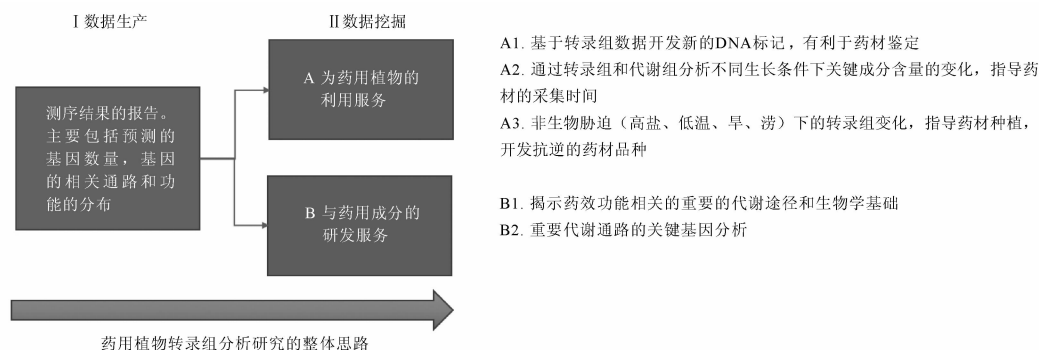


图2 药用植物转录组分析研究的整体思路

1.2.2 基于转录组数据开发分子标记 简单重复序列 (Simple sequence repeat, SSR) 标记具有多态性高、易检测、重复性好、无放射、共显性、覆盖面广、操作简单等优点,已在基因定位、遗传多样性分析、DNA 指纹图谱的构建、分子标记辅助育种等方面得到广泛应用。与传统的随机基因组微卫星标记相比,基于基因的微卫星标记物更受欢迎,因为快速而廉价的隔离方法和它们的跨物种可移植性。运用转录组序列开发 SSR 标记可帮助提供更多的信息,能够提高遗传多样性和分子标记辅助育种研究的精准性^[66]。Zeng 等^[16]用 454 gs-flx 焦磷酸测序技术对箭叶淫羊藿 (*Epimedium sagittatum*) 的 cDNAs 进行测序。从 EST 数据集确定总共有 2 810 个 EST-SSRs,随机选择 32 个 EST-SSRs 及合成引物对,测试其在 52 种淫羊藿中的可转移性。在引物设计得当的 21 对引物中,有 18 对引物对 (85.7%) 可以成功转移到淫羊藿物种,其中 16 个具有较高的遗传多样性,在理论上,2 810 个中的 56.2% (18/32),即约 1 580 个 EST-SSRs 标记物适用于淫羊藿属内遗传研究。Luo 等^[41]采用 454 焦磷酸测序技术对三七 (*Panax notoginseng*) 进行了转录组测序,获得了三萜皂苷生物合成的候选基因,包括 CYP450s 和 UGTs,同时共有 2 772 个简单重复序列 (SSR) 被确定,SSRs 的鉴定为三七的分子育种和遗传学应用提供了大量的分子标记信息。Zheng 等^[14]使用转录组测序在魔芋属 (*Amorphophallus*) 中发现了 10 754 个 SSR 标记,在 25 个个体中成功验证了 177 个多态标记,此研究中开发的大量遗传标记物应有助于研究魔芋属的遗传多样性和种质特性。

1.2.3 基于转录组数据分析关键成分的含量变化 药用植物的关键成分是药效发挥的主要来源,一直是医药学家重点关注的研究对象,通过转录组可以测定植物不同组织、不同时期、不同地区样本之间关键基因的表达差异,以此更加有效地指导选择样本的采摘、选择及分子育种等。印度藏茴香 [*Trachyspermum ammi* (L.)] 的有效成分麝香草酚是一种单萜类化合物,具有抗真菌、抗菌、抗病毒、抗炎等多种生物活性。Howyze 等^[67]通过转录组检测出萜类生物合成通路中间阶段的基因编码酶,分析发现,两种印度藏茴香品种的 4 个花序组织之间存在大量的 unigenes 存在差异表达,同时还发现编码脱氢酶、转录因子、细胞色素 P450s 的 unigenes 差异表达,可能与印度藏茴香萜类化合物多样性有

关,此次测序数据有助于印度藏茴香功能性育种的发展。银杏 (*Ginkgo biloba*) 俗称活化石,其叶片是黄酮类化合物的主要来源,采摘收获的时间对化合物提取优化起着重要作用。Ni 等^[68]对正午采摘和午夜采摘的两种银杏叶进行转录组分析发现,参与黄酮类生物合成的每一步骤的基因在午夜时间大概率表达量下调,并通过 Real-time PCR 进行验证。对两个时段的银杏叶黄酮类化合物进行含量测定,午夜时间总黄酮类化合物含量下降,进一步分析,其中大多数成分的含量呈现不同程度的下降,这项研究发现昼夜节律对银杏黄酮类化合物的含量有一定影响,为优化采摘时间提供数据基础。

1.2.4 非生物胁迫下的转录组变化 通过对比药用植物正常植株和非生物胁迫下的植株之间的基因表达差异,有助于发现耐受胁迫的应激反应基因。青蒿 (*Artemisia annua*) 是抗疟疾化合物青蒿素的植物来源,但青蒿素合成的含量有限。为了探究不同非生物胁迫条件下青蒿素含量的变化,Vashisth 等^[69]通过对 4 种非生物胁迫条件 (高盐、寒冷、干旱、水涝) 下的青蒿植株与正常植株进行含量对比发现,除干旱外,所有胁迫条件下青蒿素含量均有所增加。对各条件下的植株进行转录组测序,分别得到 89 362 份 (正常植株)、81 328 份 (高盐)、76 337 份 (寒冷)、90 470 份 (干旱) 和 96 493 份 (水涝) 转录本,应激基因在应激条件表现出上调或排他表达,并发现应激条件下的转录因子的变化,为青蒿多重非生物胁迫相关基因的功能研究和多重耐受性的潜在候选基因、应激耐受机制研究提供了新的方向。

1.2.5 基于转录组数据研究药效功能相关的代谢通路 许多药用植物的活性成分是其次生代谢产物,一个基因的表达会受到其他基因的影响,同时也会对其他基因的表达产生一定的影响,几者相互关联,相互作用,形成基因表达调控网络。通过转录组测序后的海量基因数据,可有助于挖掘代谢途径数据库及代谢网络,为及时发现未知的酶提供助力。美国人参 (西洋参) (*Panax quinquefolius* L.) 是世界上使用最广泛的草药之一,其主要的生物活性成分是三萜皂苷,即人参皂苷。然而,人们对西洋参中的人参皂苷的生物合成知之甚少,特别是在这一途径的晚期。Sun 等^[39]对西洋参转录组数据进行分析,发现了所有与人参皂苷合成相关的酶,均从乙酰辅酶 a 通过异戊二烯途径开始。为了促进对喜马拉雅山

脉上的一种濒临灭绝的药用草本植物——西藏延龄草(*Trilium govanianum*)的重要基因和药物重要生物合成途径的调控机制的基本了解,Singh 等^[12]对西藏延龄草的首次空间转录组测序进行了研究,发现了甾体皂苷生物合成及其他次生代谢途径中包含的基因,包括 brassinosteroid、类胡萝卜素、双萜类、类黄酮、苯丙类、甾类和萜类化合物的生物合成相关基因,以及重要的 TF 家族(bHLH、MYB 相关、NAC、FAR1、bZIP、B3、WRKY)。

1.2.6 药用植物功能基因的挖掘 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,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。通过转录组学研究可以筛选与药用活性成分相关的功能基因以及关键酶基因,对药用植物活性成分研究与高效利用具有重要的理论和实践意义。Chen 等^[26]利用 Illumina Hiseq2000 平台对 3 个不同的蒙古黄芪[*Astragalus membranaceus* Bge. var. *mongolicus* (Bge.) Hsiao]组织进行 RNA-测序,发现与异丙氨酸和三萜皂苷的生物合成相关的基因。涉及三萜皂苷生物合成的 MVA 和非 MVA 通路中的基因在 3 个检测组织中均有差异表达。Yang 等^[56]使用 454 GS-FLK 焦磷酸测序平台对丹参(*Salvia miltiorrhiza* Bunge)的根和叶组织进行转录组分析,鉴定出了 2 683 个在跟中高度表达的 unigenes,其中包括丹参素生物合成早期的编码酶基因,如柯巴基二磷酸合成酶(SmCPS),类贝壳杉烯合酶(SmKSL)和 CYP76AH1。

2 药用植物转录组研究展望

药用植物是植物中被人为定义出来的一类植物,并不是基于科学定义的分类法。植物为了抵抗昆虫、真菌、食草动物的侵害,会产生各种化学物质保护自己,这些化学物质有些成为治疗人类疾病的药物有效成分,这些植物就可作为药用植物。全世界的植物约有 27 万种^[70],高等植物约 25 万种,经学者调查,其中 17% 具有药用潜能^[71]。在植物界中,除了一些模式植物研究资源很丰富以外,大部分的植物遗传资源相对较少,而药用植物相关的遗传资源更少。因此,研究药用植物,除了方法体系需要参考植物界研究的经典方法外,在数据和知识层面也需要利用植物界已有的成果。2009 年美国、加拿大和中国等科学家联合发起的千种植物转录组计划(<http://www.onekp.com/>)计划完成 1 000 种植物的转录组测序^[72],覆盖了大部分的植物科级分类,是迄今为止最大规模的植物遗传资源计划,截止到 2018 年 10 月该计划已完成 1 400 余种植物的转录组数据测序、归档和分析工作。充分利用植物遗传资源大数据可以有效补充现有药用植物遗传资源的数据积累,而且药用植物只有与植物界的遗传资源进行对比参照研究才能在更大的视野层面上有新的发现。

基因组、转录组等遗传数据与代谢、蛋白质数据的联合分析逐渐在科学研究中成为新的实验设计和数据分析模式,

而与资源普查数据、生长所处地理环境数据的综合分析将是未来更全面理解药用植物特性的必然选择,基于现代统计学、机器学习的数据分析技术为海量、多层次的数据分析提供了技术可能性,将现代的 IT 架构和互联网技术应用到药用植物遗传数据共享和合作研究中将会为本领域的研究带来思想碰撞和新的研究模式。在药用植物领域中,“道地药材”这一概念非常重要而又颇多争议。我国的药用植物有 11 146 种^[1],其中道地药材约 200 余种,道地药材的用量占中药材总量的 80%^[2],说明道地药材已经成为中药行业中非常重要的一部分。道地药材之所以药效较非道地药材更好而被各医家推崇,除了生产技术、临床选择、文化传播、市场交通、社会政治等因素外,遗传变异和自然选择是道地药材形成的原因,换言之,道地药材的形成是特定的基因型,在特定的生境下受到复杂的调控,导致某些代谢过程的关键酶基因的表达产生了时空差异的产物^[73]。通过分析和总结,黄璐琦提出“道地药材的道地性越明显,其基因特化越明显”的假说来阐释道地药材的形成机制^[74]。道地药材的研究是十分典型的需要结合遗传、代谢、环境等数据进行综合分析的领域,而如此复杂的研究课题和数据构成,将更多的依赖于多学科的专家联合贡献,甚至在同一套数据基础上给出不同角度的解释。

随着生物大数据生产的规模化,机器学习等数据分析技术的应用,现代的生命科学研究已经进入大数据整合、多学科联合的时代,“药用植物遗传数据资源平台”是切合时代发展的需求。遗传数据资源共享平台需全面整合药用植物的样本采集数据,样本实验数据,图片资源及遗传分析数据等,并开发基于药用植物遗传资源数据的服务,开发快捷的数据查询、检索、数据上传和下载功能,建立数据全面、界面优良、使用便捷、系统稳定的药用植物资源和数据资源共享平台。为未来所有药用植物转录组资源研究者开放数据上传和使用通道。我们给出该平台的架构示意图供读者参考(图 3)。

综上所述,转录组测序能够全面揭示生物个体基因在特定时期和特定组织的表达情况,在分子标记、代谢物含量监控、功能基因的挖掘、药用植物活性成分的生物合成与调控、探索药材道地性分子机制方面提供了新的思路和方法。但是目前的药用植物遗传信息研究均是关注于单种或相近的少数几种药用植物,所用的数据仅为基因数据或联合代谢部分数据,而数据的共享、复用、联合等均未普遍涉及。我们

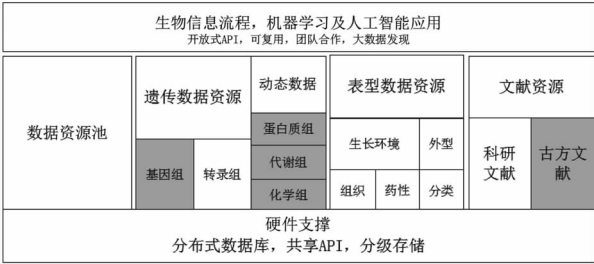


图3 中草药数据资源平台示意图

提出药用植物转录组分析的未来趋势在于大数据、多层次数据、多学科专家的联合,在基础的技术实现层面需要构建药用植物遗传数据资源平台,作为整体的数据存储和分析的集中地,也作为多学科专家掌握数据权限和建立共享合作模式的平台,为全球药用植物研究者提供前所未有的基础资源,将极大的促进药用植物药用活性成分、生物合成途径和基因调控等的纵向研究,以及药用植物的科研产出最大化,同时为提升我国的药用植物研究和经济发展打下坚实的基础,并将进一步奠定中国药用植物研究的国际地位。

REFERENCES

- [1] WANG C, WU Z C. Investigation on exploitation and utilization of medicinal biological resources in China[J]. *Bull Chin Acad Sci*(中国科学院院刊), 1996, 6:423-429.
- [2] State Administration of Traditional Chinese Medicine of the People's Republic of China. Q&a 41-50, outline of strategic planning for the development of traditional Chinese medicine (2016-2030) [EB/OL]. Peking, 2017, <http://www.satcm.gov.cn/fajiansi/zhengcewenjian/2018-03-24/2476.html>.
- [3] COSTA V, ANGELINI C, FEIS I D, *et al.* Uncovering the complexity of transcriptomes with RNA-Seq[J]. *J Biomed Biotechnol*, 2010:853916.
- [4] WANG Z, MARK G, MICHAEL S. RNA-Seq: a revolutionary tool for transcriptomics[J]. *Nat Rev Genet*, 2008, 10:57-63.
- [5] WANG Z, GERSTEIN M, SNYDER M, *et al.* RNA-Seq: a revolutionary tool for transcriptomics[J]. *Nat Rev Genet*, 2009, 10(1):57-63.
- [6] HUANG C, WU M H, LI G Y. The present and advance in transcriptomics of nasopharyngeal carcinoma[J]. *Prog Biochem Biophys*(生物化学与生物物理进展), 2007, 34(11):1129-1135.
- [7] LIN X H, ZHANG J, LI Y, *et al.* Functional genomics of a living fossil tree, *Ginkgo*, based on next-generation sequencing technology[J]. *Physiol Plantarum*, 2011, 143(3):207-218.
- [8] LI S T, ZHANG P, ZHANG M, *et al.* Transcriptional profile of *Taxus chinensis* cells in response to methyl jasmonate[J]. *BMC Genomics*, 2012, 13:295.
- [9] WU Q, SUN C, LUO H M, *et al.* Transcriptome analysis of *Taxus cuspidata* needles based on 454 pyrosequencing[J]. *Planta Med*, 2011, 77(4):394-400.
- [10] DENG N, CHANG E M, LI M H, *et al.* Transcriptome characterization of *Gnetum parvifolium* reveals candidate genes involved in important secondary metabolic pathways of flavonoids and stilbenoids[J]. *Front Plant Sci*, 2016, 7:174.
- [11] LIU T, LI X X, XIE S Q, *et al.* RNA-seq analysis of *Paris polyphylla* var. *yunnanensis* roots identified candidate genes for saponin synthesis[J]. *Plant Diversity*, 2016, 38(3):163-170.
- [12] SINGH P, SINGH G, BHANDAWAT A, *et al.* Spatial transcriptome analysis provides insights of key gene(s) involved in steroidal saponin biosynthesis in medicinally important herb *Trillium govanianum*[J]. *Sci Rep*, 2017, 7:45295.
- [13] GUO X, LI Y, LI C F, *et al.* Analysis of the *Dendrobium officinale* transcriptome reveals putative alkaloid biosynthetic genes and genetic markers[J]. *Gene*, 2013, 527(1):131-138.
- [14] ZHENG X F, PAN C, DIAO Y, *et al.* Development of microsatellite markers by transcriptome sequencing in two species of *Amorophallus* (Araceae)[J]. *BMC Genomics*, 2013, 14:490.
- [15] WEI L, LI S H, LIU S G, *et al.* Transcriptome analysis of *Houttuynia cordata* Thunb. by illumina paired-end RNA sequencing and SSR marker discovery[J]. *PLoS One*, 2014, 9(1):e84105.
- [16] ZENG S H, XIAO G, GUO J, *et al.* Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) Maxim [J]. *BMC Genomics*, 2010, 11:94.
- [17] XIE D M, YU N J, HUANG L Q, *et al.* Next generation sequencing and transcriptome analysis of root bark from *Paeonia suffruticosa* cv. Feng Dan[J]. *China J Chin Mater Med* (中国中药杂志), 2017, 42(5):2954-2961.
- [18] PATHAK S, LAKHWANI D, GUPTA P, *et al.* Comparative transcriptome analysis using high papaverine mutant of *Papaver somniferum* reveals pathway and uncharacterized steps of papaverine biosynthesis[J]. *PLoS One*, 2013, 8(5):e65622.
- [19] ZENG J G, LIU Y S, LIU W, *et al.* Integration of transcriptome, proteome and metabolism data reveals the alkaloids biosynthesis in *Macleaya cordata* and *Macleaya microcarpa* [J]. *PLoS One*, 2013, 8(1):e53409.
- [20] ZHANG L B, JIA H B, YIN Y T, *et al.* Transcriptome analysis of leaf tissue of *Raphanus sativus* by RNA sequencing[J]. *PLoS One*, 2013, 8(11):e80350.
- [21] BAKEL H V, STOUT J M, COTE A G, *et al.* The draft genome and transcriptome of *Cannabis sativa*[J]. *Genome Biol*, 2011, 12(10):R102.
- [22] HAO D C, MA P, MU J, *et al.* De novo characterization of the root transcriptome of a traditional Chinese medicinal plant *Polygonum cuspidatum*[J]. *Sci China Life Sci*, 2012, 55(5):452-466.
- [23] LOKE K K, RAHNAMAIE-TAJADOD R, YEOH C C, *et al.* Transcriptome analysis of *Polygonum minus* reveals candidate genes involved in important secondary metabolic pathways of phenylpropanoids and flavonoids[J]. *Peer J*, 2017, 5:e2938.
- [24] ZHANG Z, GAO Z H, WEI J H, *et al.* The mechanical wound transcriptome of three-year-old *Aquilaria sinensis*[J]. *Acta Pharm Sin*(药学报), 2012, 47(8):1106-1110.
- [25] JIANG B, XIE D S, LIU W R, *et al.* De novo assembly and characterization of the transcriptome, and development of SSR markers in wax gourd (*Benicasa hispida*)[J]. *PLoS One*, 2013, 8(8):e71054.
- [26] CHEN J, WU X T, XU Y Q, *et al.* Global transcriptome analysis profiles metabolic pathways in traditional herb *Astragalus membranaceus* Bge. var. *mongolicus* (Bge.) Hsiao[J]. *BMC Genomics*, 2015, 16(7):S15.
- [27] FALARA V, FOTOPOULOS V, MARGARITIS T, *et al.* Transcriptome analysis approaches for the isolation of trichome-specific genes from the medicinal plant *Cistus creticus* subsp. *creticus*[J]. *Plant Mol Biol*, 2008, 68(6):633-651.
- [28] FATIMA T, SNYDER C L, SCHROEDER W R, *et al.* Fatty acid composition of developing sea buckthorn (*Hippophae rhamnoides* L.) berry and the transcriptome of the mature seed[J]. *PLoS One*, 2012, 7(4):e34099.
- [29] SHARMA P C, JAIN A, CHAUDHARY S. Transcriptome analysis in seabuckthorn (*Hippophae rhamnoides* L.), a medicinally important plant [C]. Bangkok (Thailand): International Conference on Environmental and Biological Sciences, 2012:21-24.
- [30] GHANGAL R, CHAUDHARY S, JAIN M, *et al.* Optimization of de novo short read assembly of seabuckthorn (*Hippophae rhamnoides* L.) transcriptome[J]. *PLoS One*, 2013, 8(8):e72516.

- [31] JAIN A, CHAUDHARY S, SHARMA P C. Mining of microsatellites using next generation sequencing of seabuckthorn (*Hippophae rhamnoides* L.) transcriptome[J]. *Physiol Mol Biol Plants*, 2014, 20(1):115-123.
- [32] SATHYANARAYANA N, PITTALA R K, TRIPATHI P K, et al. Transcriptomic resources for the medicinal legume *Mucuna pruriens*: de novo transcriptome assembly, annotation, identification and validation of EST-SSR markers[J]. *BMC Genomics*, 2017, 18(1):409.
- [33] XU L, WANG J B, LEI M, et al. Transcriptome analysis of storage roots and fibrous roots of the traditional medicinal herb *Callerya speciosa* (Champ.) Schott[J]. *PLoS One*, 2016, 11(8):e0160338.
- [34] ZHANG F Q, GAO Q B, KHAN G, et al. Comparative transcriptome analysis of aboveground and underground tissues of *Rhodiola algida*, an important ethno-medicinal herb endemic to the Qinghai-Tibetan Plateau[J]. *Gene*, 2014, 553(2):90-97.
- [35] HAN R C, TAKAHASHI H, NAKAMURA M, et al. Transcriptome analysis of nine tissues to discover genes involved in the biosynthesis of active ingredients in *Sophora flavescens* [J]. *Biol Pharm Bull*, 2015, 38:876-883.
- [36] LI Y, LUO H M, SUN C, et al. EST analysis reveals putative genes involved in glycyrrhizin biosynthesis[J]. *BMC Genomics*, 2010, 11:268.
- [37] SUN Y Z, LUO H M, LI Y, et al. Pyrosequencing of the *Camptotheca acuminata* transcriptome reveals putative genes involved in camptothecin biosynthesis and transport [J]. *BMC Genomics*, 2011, 12:533.
- [38] KRISHNAN N M, PATTNAIK S, JAIN P, et al. A draft of the genome and four transcriptomes of a medicinal and pesticidal angiosperm *Azadirachta indica* [J]. *BMC Genomics*, 2012, 13:464.
- [39] SUN C, LI Y, WU Q, et al. De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis[J]. *BMC Genomics*, 2010, 11:262.
- [40] WU D, AUSTIN R S, ZHOU S J, et al. The root transcriptome for North American ginseng assembled and profiled across seasonal development[J]. *BMC Genomics*, 2013, 14:564.
- [41] JAYAKODI M, LEE S C, PARK H S, et al. Transcriptome profiling and comparative analysis of *Panax ginseng* adventitious roots [J]. *J Ginseng Res*, 2014, 38(4):278-288.
- [42] LUO H M, SUN C, SUN Y Z, et al. Analysis of the transcriptome of *Panax notoginseng* root uncovers putative triterpene saponin-biosynthetic genes and genetic markers[J]. *BMC Genomics*, 2011, 12:S5.
- [43] SUI C, CHEN M, XU J S, et al. Comparison of root transcriptomes and expressions of genes involved in main medicinal secondary metabolites from *Bupleurum chinense* and *B. scorzonifolium*, the two Chinese official *Radix bupleuri* source species[J]. *Physiol Plant*, 2015, 153(2):230-242.
- [44] ZHONG T, ZHANG H, DUAN X Y, et al. Anti-obesity effect of radix *Angelica sinensis* and candidate causative genes in transcriptome analyses of adipose tissues in high-fat diet-induced mice [J]. *Gene*, 2017, 599:92-98.
- [45] GUPTA P, GOEL R, AGARWAL A V, et al. Comparative transcriptome analysis of different chemotypes elucidates withanolide biosynthesis pathway from medicinal plant *Withania somnifera* [J]. *Sci Rep*, 2015, 5:18611.
- [46] GARZÓN MARTÍNEZ G A, ZHU Z I, LANDSMAN D, et al. The *Physalis peruviana* leaf transcriptome: assembly, annotation and gene model prediction[J]. *BMC Genomics*, 2012, 13:151.
- [47] KHALDUN A B, HUANG W J, LIAO S H, et al. Identification of microRNAs and target genes in the fruit and shoot tip of lycium chinense: a traditional Chinese medicinal plant[J]. *PLoS One*, 2015, 10(1):e0116334.
- [48] GAHLAN P, SINGH H R, SHANKAR R, et al. De novo sequencing and characterization of *Picrorhiza kurroa* transcriptome at two temperatures showed major transcriptome adjustments[J]. *BMC Genomics*, 2012, 13:126.
- [49] SUN P, SONG S H, ZHOU L L, et al. Transcriptome analysis reveals putative genes involved in iridoid biosynthesis in *Rehmannia glutinosa* [J]. *Int J Mol Sci*, 2012, 13:13748-13763.
- [50] WANG F Q, ZHI J Y, ZHANG Z Y, et al. Transcriptome analysis of salicylic acid treatment in *Rehmannia glutinosa* hairy roots using RNA-seq technique for identification of genes involved in acteoside biosynthesis[J]. *Front Plant Sci*, 2017, 8:787.
- [51] WU B, LI Y, YAN H X, et al. Comprehensive transcriptome analysis reveals novel genes involved in cardiac glycoside biosynthesis and miRNAs associated with secondary metabolism and stress response in *Digitalis purpurea* [J]. *BMC Genomics*, 2012, 13:15.
- [52] KOTWAL S, KAULS, SHARMA P, et al. De novo transcriptome analysis of medicinally important *Plantago ovata* using RNA-Seq [J]. *PLoS One*, 2016, 11(3):e0150273.
- [53] HUANG L L, YANG X, SUN P, et al. The first illumina-based *De novo* transcriptome sequencing and analysis of safflower flowers [J]. *PLoS One*, 2012, 7(6):e38653.
- [54] LI Y, SUN C, LUO H M, et al. Transcriptome characterization for *Salvia miltiorrhiza* using 454 GS FLX [J]. *Acta Pharm Sin (药学报)*, 2010, 45(4):524-529.
- [55] HUA W P, ZHANG Y, SONG J, et al. De novo transcriptome sequencing in *Salvia miltiorrhiza* to identify genes involved in the biosynthesis of active ingredients[J]. *Genomics*, 2011, 98(4):272-279.
- [56] YANG L, DING G H, LIN H Y, et al. Transcriptome analysis of medicinal plant *Salvia miltiorrhiza* and identification of genes related to tanshinone biosynthesis[J]. *PLoS One*, 2013, 6(11):e80464.
- [57] GAO W, SUN H X, XIAO H B, et al. Combining metabolomics and transcriptomics to characterize tanshinone biosynthesis in *Salvia miltiorrhiza* [J]. *BMC Genomics*, 2014, 15:73.
- [58] SONG Z Q, GUO L L, LIU T, et al. Comparative RNA-sequence transcriptome analysis of phenolic acid metabolism in *Salvia miltiorrhiza*, a traditional Chinese medicine model plant[J]. *Int J Genomics*, 2017:9364594.
- [59] CHERUKUPALLI N, DIVATE M, MITTAPELLI S R, et al. De novo assembly of leaf transcriptome in the medicinal plant *Andropogon paniculata* [J]. *Front Plant Sci*, 2016, 7:1203.
- [60] HE M, WANG Y, HUA W P, et al. De novo sequencing of *Hypericum perforatum* transcriptome to identify potential genes involved in the biosynthesis of active metabolites[J]. *PLoS One*, 2012, 7(7):e42081.
- [61] KWON C W, PARK K M, KANG B C, et al. Cysteine protease profiles of the medicinal plant *Calotropis procera* R. Br. revealed by de novo transcriptome analysis[J]. *PLoS One*, 2015, 10(3):e0119328.

- [62] LIU Y, WANG Y, GUO F X, *et al.* Deep sequencing and transcriptome analyses to identify genes involved in secoiridoid biosynthesis in the tibetan medicinal plant *Swertia mussotii* [J]. *Sci Rep*, 2017, 7:43108.
- [63] ZHANG X D, ALLAN A C, LI C X, *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.
- [64] KAUSHIK P, KUMAR S. Data of de novo assembly of the leaf transcriptome in *Aegle marmelos* [J]. *Data Brief*, 2018, 19:700-703.
- [65] 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.
- [66] HUANG H Y, DU H Y, TAN W Y, *et al.* Development of SSR molecular markers based on transcriptome sequencing of *Eucommia ulmoides* [J]. *Sci Silv Sin* (林业科学), 2013, 49 (5): 176-181.
- [67] HOWYZEH M S, NOORI S A S, VAHID S J, *et al.* Comparative transcriptome analysis to identify putative genes involved in thymol biosynthesis pathway in medicinal plant *Trachyspermum ammi* L. [J]. *Sci Reports*, 2018, 8:13405.
- [68] NI J, DONG L X, JIANG Z F, *et al.* Comprehensive transcriptome analysis and flavonoid profiling of *Ginkgo* leaves reveals flavonoid content alterations in day-night Cycles [J]. *PLoS One*, 2018, 13 (3): e0193897.
- [69] VASHITH D, KUMAR R, RASTOGI S, *et al.* Transcriptome changes induced by abiotic stresses in *Artemisia annua* [J]. *Sci Reports*, 2018, 8 (1): 3423.
- [70] TAN X M, ZHOU Y Q, CHEN J, *et al.* Advances in research on diversity of endophytic fungi from medicinal plants [J]. *Chin Pharm J* (中国药理学杂志), 2015, 50 (18): 1563-1580.
- [71] MAMEDOV N. Medicinal plants studies: history, challenges and prospective [J]. *Med Aromatic Plants*, 2012, 1: 8.
- [72] MATASCI N, HUNG L H, YAN Z X, *et al.* Data access for the 1 000 plants (1 kP) project [J]. *Giga Sci*, 2014, 3: 17.
- [73] HUANG L Q, GUO L P, HUA G D, *et al.* Attributes of Chinese geoherbs and its study strategy [J]. *Chin J Inform Tradit Chin Med* (中国中医药信息杂志), 2007, 14 (2): 44.
- [74] HUANG L Q, CHEN M L, XIAO P G. The modern biological basis and model hypothesis for the study of Chinese medicinal materials' pathogenicity [J]. *China J Chin Mater Med* (中国中药杂志), 2004, 29 (6): 494.

(收稿日期:2018-08-05)