

## 转录组测序及其在野生大豆基因资源发掘中的应用

吴倩, 张磊, 黄志平, 王大刚, 胡国玉

(安徽省农业科学院作物研究所/安徽省农作物品质改良重点实验室, 安徽 合肥 230031)

**摘要:**野生大豆是栽培大豆的原始祖先种,其遗传多样性远远超过栽培大豆,并且蕴藏着多种优异抗性基因。目前关于野生大豆优异基因挖掘和利用的研究很少,本研究概述了转录组测序的基本原理、实验流程、数据分析和应用现状,并展望了将转录组测序技术应用于野生大豆的研究,可能对野生大豆资源基因发掘和利用及其在大豆种质创新与品种设计育种中产生的影响。将野生大豆研究深入到分子水平,对野生大豆资源的深入研究具有重要理论意义和实践价值。

**关键词:**野生大豆;基因资源;转录组测序;挖掘和利用

**中图分类号:**S565.1      **文献标识码:**A      **文章编号:**1000-9841(2013)06-0845-07

## Transcription Sequencing and Its Application on Discovering the Gene Resources of Wild Soybean

WU Qian, ZHANG Lei, HUANG Zhi-ping, WANG Da-gang, HU Guo-yu

(Crop Institution of Anhui Academy of Agricultural Science/Key Lab of Crop Quality Improvement of Anhui Province, Hefei 230031, China)

**Abstract:** As the original ancestor of the cultivated soybean, wild soybean has abundant genetic diversity and holds a variety of excellent genes. The molecular and genomic research of the wild soybean genes was less documented. This study summarized the basic principle, experiment process, data analysis and the current application status of transcription sequencing, and also prospected the influence on the gene resources exploitation and germplasm enhancement and molecular design breeding of soybean based on the application of transcription sequencing to wild soybean. The in-depth research of wild soybean to the molecular level has great theoretical significance and practical value to excavation and utilization of the wild soybean resources.

**Key words:** Wild soybean; Gene resource; Transcription sequencing; Excavate and utilize

大豆起源于中国,一年生野生大豆是栽培大豆的原始祖先种,较栽培大豆拥有更高的遗传多样性,是我国乃至世界的宝贵植物遗传资源<sup>[1]</sup>。野生大豆具有多种优良性状,如高蛋白、多花、多荚、多粒等产量品质类性状;耐荫性、部分材料对光周期的不敏感性等生理类性状;抗大豆胞囊线虫病(SCN)、抗大豆花叶病(SMV)、抗蚜虫、耐旱、耐涝、耐盐碱等抗病虫及耐逆境性状;核不育、核质不育等育性性状<sup>[2]</sup>。栽培大豆亲本遗传基础狭窄,缺乏丰富的基因源是大豆育种难以取得突破性进展的重要原因<sup>[3]</sup>。野生大豆蕴藏的多种优异基因是大豆改良的重要基因来源,野生大豆与栽培大豆在分类上属于同一物种,无种间隔离现象,是研究遗传育种和生物技术必不可少的载体,可以用其来拓宽栽培大豆遗传基础、增加有益性状的变异范围。

目前我国野生大豆基因资源的研究和利用仍处于初级阶段,加快野生大豆基因组学和分子生物学的研究显得尤为重要。转录组调控是生物体最

重要的调控方式之一,也是功能基因组学的重要研究方向。随着新一代高通量测序技术的日益成熟,越来越多的人开始关注转录组测序,该技术已成为挖掘基因资源的重要手段。转录组也称为“转录组”,广义上指特定组织或细胞在相同环境(或生理条件)下的所有转录产物(RNA)的集合,包括mRNA和非编码RNA(non-coding RNA, ncRNA);狭义上指所有mRNA的集合。转录组测序技术(RNA-seq)是通过深度测序对组织或细胞内的所有RNA反转录成的cDNA进行测序,通过对相关读段(reads)数统计不仅可以计算出RNA表达量的差异,还能够发现新转录本以及转录水平的SNP<sup>[4]</sup>。运用转录组测序技术可进行基因差异表达、等位基因特异性表达、新基因和SNP发现、RNA编辑、可变剪接和分子进化等方面的相关研究<sup>[5-6]</sup>。

转录组测序技术(RNA-seq)已经成功应用于多项研究中, Eveland等<sup>[7]</sup>成功地将转录组测序和数字基因表达谱技术结合应用于玉米雌穗发育的研

收稿日期:2013-07-13

基金项目:国家高技术研究发展计划“863计划”(2011AA10A105);现代农业产业技术体系建设专项(CARS-04-PS07)。

第一作者简介:吴倩(1984-),女,博士,助理研究员,主要从事大豆分子育种研究。E-mail:woshiwqq@126.com。

通讯作者:张磊(1956-),男,研究员,主要从事大豆遗传育种研究。E-mail:leizh66@163.com。

究,发现了一批调控花序结构的基因。Wang 等<sup>[8]</sup>利用数字基因表达谱测序技术(DGE)研究了野生型和突变型棉花的毛状体纤维细胞的分裂和延长过程,测序结果发现纤维素合成酶基因、磷酸(脂)酶基因和脱氢酶基因在纤维细胞发育过程中差异表达量最高,因此这些基因很可能参与了纤维细胞发育的调控。莫德林等<sup>[9]</sup>应用 Solexa 测序技术对脂肪型蓝塘猪和瘦肉型长白猪不同时期背最长肌进行转录谱测序,通过差异基因筛选从不同发育阶段共鉴定出 595 个可能与肌肉发育有密切关系的差异表达基因,并发现一些与脂肪发育有关的基因,如 CEBP 家族, FABP3 在猪背最长肌中高表达,而且在蓝塘猪体内表达量更高。

我国野生大豆资源丰富,占世界总数的 90%。来永才等<sup>[10]</sup>开展了野生大豆利用研究,筛选出一批品质性状特异、高抗逆性、高繁殖系数的种质资源。杨振宇等<sup>[11]</sup>对野生大豆抗蚜性进行重复鉴定,发现了具有生物学抗性的材料 85-32,高抗蚜虫。马晓萍等<sup>[12]</sup>对 989 份野生大豆进行鉴定,筛选出 8 份高抗大豆胞囊线虫 3 号生理小种,5 份高抗大豆花叶病毒病,381 份高抗大豆食心虫的野生大豆材料。加快野生大豆资源基因发掘和利用是一个重要的研究领域,也是一个挑战。转录组测序技术凭借其优势在水稻育种中已广泛应用于重要农艺性状相关基因的发掘,而利用新一代高通量测序技术研究野生大豆基因资源挖掘与利用鲜有报道。本文对基于新一代高通量测序的转录组测序技术基本原理、实验流程、数据分析及其技术优势与应用进行综述,并对转录组测序在野生大豆资源发掘及种质创新与应用前景进行展望,旨在为我国野生大豆资源的开发提供新的研究思路,同时为栽培大豆的改良提供参考。

## 1 转录组测序的方法及数据分析

### 1.1 测序平台的选择

转录组测序技术一般由模板准备、测序和成像、序列组装和比对等部分组成。高通量测序平台 Roche 454、Illumina Solexa 和 ABI SOLiD 均可进行转录组测序,但是 3 种平台测序原理不同,产出数据量、数据质量和运行成本也各不相同,各有优缺点。Roche 454 测序平台的优点是单序列读长可以达到 600~1 000 bp,适用于未知基因组序列物种的转录组分析,读长最长且便于拼接,能获得更好的转录本数据,缺点是通量较低,对连续单碱基重复区判断准确度不高。Illumina/Solexa 平台测序通量

大,但读长比 454 测序短(通常 100 bp),单次测序获得的数据量大,覆盖率高,能够检测到更多低丰度的转录本,且成本较低,在已知基因组序列的物种转录组分析中更具优势。ABI/SOLiD 测序平台读长最短(50 bp),双碱基编码的应用降低了测序错误率,更适用于高质量的参考基因组序列,特别是 SNP 检测,但该技术的广泛应用受片段长度问题的限制<sup>[13-14]</sup>。在实际研究中,可以根据所选材料、研究目的以及不同测序平台的特点、运行成本,选择合适的测序平台和试验方案。对有参考基因组序列的物种进行转录组测序,选择 Illumina 测序平台更加经济,对于没有参考基因组的物种,选择 Roche 454 测序平台能够降低后续数据处理中拼接难度,从而获得更为精确的数据<sup>[15]</sup>。

### 1.2 样品分离和文库构建

转录组测序工作第一步是 RNA 样品的分离和制备。提取样品总 RNA 后,真核生物用带有 Oligo(dT)的磁珠富集 mRNA,若为原核生物,还需用试剂盒去除 rRNA,此外为了防止基因组 DNA 的污染,还需要用 DNase 去除基因组 DNA。选择不同的测序平台,对 RNA 浓度也有不同的要求。Roche GS FLX 测序平台要求样品的总 RNA 质量浓度大于 400 ng·μL<sup>-1</sup>,总量大于 15 μg;而 Illumina Solexa 平台要求样品的总 RNA 质量浓度大于 300 ng·μL<sup>-1</sup>,总量大于 6 μg<sup>[14]</sup>。以总 RNA 为模板利用反转录试剂盒合成 cDNA 第一条链,再加入缓冲液、dNTPs、Rnase H 和 DNA polymerase I 合成第二条 cDNA 链,再利用 PCR 纯化试剂盒去除反应体系中小于 300 bp 的片段。通过多次扩增、纯化、浓缩,最终收集到双链 cDNA。对双链 cDNA 进行末端修复并连接测序接头,然后用琼脂糖凝胶进行片段大小选择。采用 454GS FLX Titanium 平台,双链 cDNA 需要被打断成 300~800 bp 的片段后,两端添加特异性衔接子 A 和 B,并变性为单链连接到磁珠上,经 emPCR 富集后上机测序;而采用 Illumina Solexa 平台测序,双链 cDNA 则需要被打断成 150 bp 左右的片段,经处理才能上机测序。对构建的文库还要进行质量检测,以确保文库的大小、纯度及浓度。

### 1.3 测序结果的处理和分析

1.3.1 数据处理 测序得到原始图像数据经 Base calling 软件处理转化为相应的核苷酸序列数据,称之为原始读段(raw reads)并以 FASTQ 格式储存。对获得原始读段的数量、长度、碱基数以及 GC 含量等数据信息进行质量评估和可信度分析,评估其是否满足信息分析要求。再利用软件如 SeqClean 去除衔接子区域和低质量序列,屏蔽 cDNA 文库的

PCR 引物,将所获得的高质量短序列(clean reads)拼接、组装。根据样品是否有基因组测序数据,可分为有参考基因组的重测序读长定位和无参考基因组的从头测序组装。对于有基因组数据可参考的物种,将读长排序并通过序列定位到参考基因组上,经比对挑选匹配好的读长进行基因定位和进一步分析,常用的组装定位软件有 BWA、SOAP、SAM-tools、MAQ 和 ZOOM。重测序为研究已测序模式植物的转录组提供了一个有效的方法,并已得到了广泛应用。对没有基因组数据可以参考的物种样品就要采用从头测序组装(de novo sequence assembly)方法,目前从头测序组装的常用软件有 SOAPdenovo、Velvet、Oases、Abyss 和 ALLPATH 等<sup>[16]</sup>。以 SOAPdenovo (Version 1.03) 软件进行从头测序组装为例,首先将具有一定长度 overlap 的读段拼接成不含 N 的组装片段,这些片段称之为 Contig (重叠群)。在 Contig 数据的基础上,进一步拼接获得 scaffold (框架) 片段,并用“N”填补中间空隙。最终得到含 N 最少、两端不能再延长的非冗余序列,称之为 Unigene (单一基因簇)。自 Li 等<sup>[17]</sup>第一次对大型物种熊猫进行全基因组 de novo 测序,从头测序组装已经广泛地应用于大量无基因组参考的非模式动植物中。

1.3.2 基因注释及分类 测序结果拼接、组装后得到大量 unigenes,经过数据筛选得到 clean reads, 并进行评估。下一步需要对 unigenes 进行基因功能注释和分类。根据“基因结构相似,功能同源”的原理,结合生物信息学方法,运行 BLAST 程序将得到的 unigene 与公共核酸或蛋白质序列数据库进行相似性比对,对已有序列的基因结构进行优化,可变剪接分析,新转录本预测以及 SNP 分析等。根据基因表达注释和样品间的表达差异以及 GO 功能显著性富集分析和 Pathway 显著性富集分析,能够推测未知基因的功能或者发现新的基因。常用的核酸和蛋白质数据库: GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>)、EMBL (<http://www.edi.ac.uk/embl/>)、SWISS-PROT (<http://www.ebi.ac.uk/swissprot/>)、SMART 数据库 (<http://smart.embl-heidelberg.de/>)、Gene Ontology (<http://www.geneontology.org/>) 和 KEGG (<http://www.genome.jp/kegg/pathway.html>)。目前常用的基因功能分类方法有 Gene Ontology (简称 GO) 和 Kyoto Encyclopedia of Genes and Genomes (简称 KEGG)。GO 是一个国际标准化的基因功能分类体系,适用于各个物种,按照参与的生物学过程、分子功能和细胞组分三大亚类对基因进行分类,可全面描述生物体基因和基因

产物的属性。GO 充分利用了比较基因组方法和 BLAST 相似性搜索的优点,有效提高了自动注释准确性的同时还提供了形象的可视化信息,为科研人员提供了快速有效的指导信息。KEGG 是一个分析基因产物及其功能的数据库,能够系统地对细胞中基因产物的代谢途径、功能以及不同基因间如何相互协调完成其生物学功能进行分析。通过它可将基因组中一系列基因用细胞内分子相互作用形成的网络来展现更高一级的生物学功能,有助于研究者把基因及表达信息作为一个整体网络进行研究,进一步了解基因在生物学上的复杂行为,从而准确预测基因的功能<sup>[18-20]</sup>。

## 2 转录组测序技术优势及应用

### 2.1 技术优势

转录水平调控是生物体最重要的调控方式,转录组研究已成为功能基因组学的一个重要研究方向。转录组测序(RNA-seq)是利用新一代测序技术对片段化处理的 mRNA 反转录生成的 cDNA 进行高通量测序,拼接并统计相关序列测序次数,从而获得不同转录本(mRNA)的表达信息。基因芯片是目前应用较广的转录组检测技术,然而基因芯片技术局限于已知序列,无法检测新的转录本,且杂交灵敏度有限,难以检测低丰度的目标基因。相比传统芯片技术,转录组测序不存在传统微阵列杂交的荧光模拟信号带来的交叉反应和背景噪音问题,而且不需要任何基因序列信息,能够全面检测所有基因的表达,已逐渐取代基因芯片技术成为开展大规模研究转录组的一种有效方法<sup>[21]</sup>。RNA-Seq 技术除了具有高通量、低成本、灵敏度高、定量准、重复性好以及检测范围宽等特点,它还是一个具有高度灵活性的平台,不局限于物种基因组是否测序,能够在单核苷酸水平对任意物种的转录活动进行检测,不需要克隆的步骤和预先设计特异性探针,操作简单,应用领域广<sup>[22]</sup>。对于真核生物基因组研究,转录组测序不需要测内含子序列,与全基因组测序相比在经济方面也具有无可比拟的优势<sup>[23]</sup>。对于有参比基因的物种,将 RNA-seq 结果与基因组 DNA 序列数据对比,能够得到基因表达、基因结构优化、可变剪切、SNP 位点以及新基因发现等结果。对于没有参比基因的物种,可以进行 de novo 测序,即对得到的测序读长片段进行 de novo 组装,得到该物种 Unigene 数据并进行功能注释、蛋白编码区预测(CDS)、差异表达分析、Unigene gene ontology、代谢通路分析和富集性分析等。

## 2.2 应用

转录组测序具有如此多的优势,是深入研究全转录组的强大工具,已在动植物及人类的转录组研究中得到了广泛应用。

**2.2.1 转录本结构和可变剪接研究** 真核生物中,可变剪接的现象普遍存在,也是转录组和蛋白质组复杂多样性的重要来源。通过对转录组测序结果拼接,并与现有基因注释的比对,可以对基因结构进行优化并进行可变剪切分析。翟焕趁等<sup>[24]</sup>对稻瘟病菌菌株 FJ81278 进行转录组测序并比对,将 5 163 个基因进行了延长,其中 2 043 个基因的 5' 端区域得到延长,另外 3 120 个基因的 3' 端区域得到延长,有 1 464 个基因存在可变剪切,约占所预测全部基因的 13%,主要的可变剪接模式是内含子保留,占发生可变剪接基因的 44%。Filichin 等<sup>[25-26]</sup>通过转录组测序,发现拟南芥中 42% 以上有内含子基因存在可变剪接,这个数值远远高于以前基于 cDNA/EST 测序方法得到的预测值。

**2.2.2 基因表达水平** 运用转录组测序技术,可以研究物种在某一特定生物过程、发育阶段或处理后的基因组织特异性表达和表达水平。通过对大量样品同时测序,获得样品之间的表达差异,从而捕获低表达的基因,转录组测序已经越来越多地用于差异表达基因的筛选及鉴定等方面<sup>[14]</sup>。魏利斌等<sup>[27]</sup>通过比较不同芝麻材料间转录物序列及表达水平,初步确定在种子形成过程中有 1 277 条序列表达量下调达 10 倍以上,990 条序列仅在芝麻植株中表达而在籽粒中未有表达;660 条序列在种子形成过程中表达量上调 10 倍以上,296 条序列可能在种子形成过程中特异性表达。程立宝等<sup>[28]</sup>利用转录组测序技术研究莲藕根状茎膨大过程 34 个与淀粉代谢相关基因的表达,发现 *LrGBSS*、*LrSBEI*、*LrSBEII*、*LrSBEIII* 这 4 个基因在根状茎膨大后期表达量显著增加,其他基因表达量变化较小,表明这 4 个基因对根状茎膨大后期淀粉合成具有重要的作用。

**2.2.3 新基因发现** 转录组测序技术的一个重要用途就是用于发现新基因。Zhang 等<sup>[29]</sup>利用转录组测序技术,在水稻茎尖、根尖、叶片、稻穗和愈伤组织检测到 7 232 个新转录本和 1 356 个融合基因,同时鉴定了 234 个候选嵌合转录本。Sara 等<sup>[30]</sup>通过转录组测序技术鉴定出葡萄 (*Vitis vinifera*) 果实 glutathione-S-transferase (GST) 家族的 53 个新基因和 MYB 转录因子家族的 28 个新基因,而这些基因由于表达量低、缺少相应的探针序列使用芯片技术无法检测到。为了阐明与传染病有关的转录组反

应,Zoltán<sup>[31]</sup>利用转录组测序技术对分枝杆菌感染后的斑马鱼转录组变化进行研究,发现了经分枝杆菌感染诱导后产生的新基因和新的转录机制。此外,研究人员还利用 454 测序技术在油橄榄 (*Olea europaea* L.)、玉米 (*Zea mays* L.)、巨桉 (*Eucalyptus grandis*)、羊草 (*Leymus chinensis* Trin. Tzvel.) 等植物的转录组中发现了新基因<sup>[32-35]</sup>。

**2.2.4 非编码区域功能研究** 小分子 RNA 近几年受到了科学界的广泛关注,它是一类长约 20 ~ 30 个核苷酸的非编码 RNA 分子,其介导的转录后基因调控是生物体的一种新型基因调控机制,它在生物体的生长发育和适应环境胁迫的过程中起着重要的作用。由于小 RNA 序列短、同源性高,使用普通方法很难检测,高通量转录组测序技术解决了这一难题,并已经成功地应用于水稻<sup>[36]</sup>、拟南芥<sup>[37]</sup>和小麦<sup>[38]</sup>等生物的小 RNA 研究。Morin 等<sup>[39]</sup>利用 Solexa 测序技术在人类胚胎干细胞的小分子 RNA 谱带中发现了 81 个差异表达的小分子 RNA 和 23 种未报道的 miRNA。Huang 等<sup>[40]</sup>利用 Solexa 测序技术检测了荷斯坦奶牛不同时期(出生 3 d 和 2.5 岁)睾丸和卵巢组织 miRNA 的差异,在睾丸和卵巢中分别鉴定出 100 和 104 种新的前体 miRNA,分别编码 122 和 136 种成熟的 miRNA,发现 6 个 miRNA 是牛所特有的,246 个 miRNA 在睾丸和卵巢中共表达,其中 21 个在睾丸中特异性表达,9 个在卵巢特异性表达,与睾丸组织比较,有 65 个在卵巢中是明显上调的,这些结果表明有大量的 miRNA 在荷斯坦奶牛繁殖器官中起着重要的调节作用。Wang 等<sup>[8]</sup>在研究陆地棉品种徐州 142 和无纤维无短绒突变体的胚珠转录组时,发现 7 个与纤维形成相关的 miRNA 及其作用的靶基因,并鉴定出 36 个新的棉花 miRNA。

**2.2.5 SNP 发现及分子标记筛选** 在正向遗传学中,大量工作常常通过数量性状位点 (quantitative trait loci, QTLs) 分析来研究各种农作物复杂农艺性状的遗传机制,但其受到分子标记密度 (如 SSRs 和 RFLPs) 较低的限制。新一代测序技术的应用能够解决这一难题,并可在此基础上构建出超高密度的遗传图谱。Graham 等<sup>[41]</sup>对青蒿材料进行转录组测序,鉴定出青蒿素合成的关键基因和分子标记,从而构建出青蒿遗传图谱,并定位了 3 个与青蒿素产量相关的 QTL,其中 *LG1* 与 *LG9* 上的 QTL 与青蒿素产量正相关,*LG4* 上的 QTL 与青蒿素产量负相关。Parchman 等<sup>[42]</sup>通过对海滩松 (*Pinus contorta* Dougl.) 转录组的研究,得到了大量可利用的 SNP 和 SSR 分子标记,并为此设计了高质量的 PCR 扩增引物,为深入研究海滩松奠定了基础。Shu 等<sup>[43]</sup>通

过 454 测序系统在不同大豆 (*Glycine max* L. Merrill) 种系间的转录组中鉴定出 3 899 个 SNP, 这些 SNP 广泛分布在大豆基因组中, 参与多种生理生化过程, 从而影响重要的农艺性状, 例如位于基因 *Glyma07g034903-UTR* 的分子标记 CAPS282 就与大豆的百粒重相关。这些研究结果表明基于高通量测序的 de novo 转录组测序技术运用于动植物中, 尤其可以有效地用于基因组大且复杂的物种的新基因发现和新分子标记的开发<sup>[15]</sup>。

2.2.6 代谢途径的确定 转录组测序数据结合 GO 和 KEGG 数据库可对基因进行功能注释和所属代谢途径进行分类。Wang 等<sup>[44]</sup> 对甘薯 (*Ipomoea batatas*) 块状根进行高通量转录组测序, 从头组装出 56 516 个 unigenes (单一基因序列), 通过与已知蛋白的相似性比对, 鉴定出 35 051 个基因, 使用 KEGG 分析有 11 056 个属于代谢途径的基因, 以参与碳水化合物和次生代谢物质生物合成途径的基因为主。Sun 等<sup>[45]</sup> 利用 454 测序技术在美国西洋参的根中获得 31 088 个 unigenes, 通过与 NCBI 等公共数据库进行序列比对和功能注释等分析, 共发现了 150 个细胞色素 P450 和 235 个糖基转移酶 unigenes。运用 Real-time PCR 技术进行组织特异性表达和茉莉酸甲酯诱导表达的分析与验证, 确定了人参皂苷合成途径最关键的候选基因: 1 个 CYP450 和 4 个 UDP 基因。Hao 等<sup>[46]</sup> 结合转录组测序和 DGE 技术对红豆杉进行研究, 鉴定出 23 515 个 unigenes 和根、茎、叶 3 种组织的差异表达基因, 并经过 GO 和 KEGG 分析鉴定出紫杉烷生物合成途径的相关基因和一批组织特异性的功能基因。

2.2.7 分子进化研究 随着分子生物学的逐步深入, 研究者开始从分子水平解释生物进化的问题。随着转录组测序技术的快速发展, 采用比较基因组方式研究进化和功能成为了可能。通过保守序列推测进化历史, 用同源关系提示保守的生理和生化功能, 将有助于基因组注释。Timme 和 Delwiche<sup>[47]</sup> 结合 Sanger 与 454 测序技术对圆形鞘毛藻和草原水绵进行转录组测序, 通过聚类分析得到草原水绵的 12 000 条 unigenes 和圆形鞘毛藻的 19 000 条 unigenes。序列比对发现, 与其他自养的绿藻相比, 在转录组水平上这两种植物与几种可能在进化过程中非常重要的陆生植物更加相似。Wei 等<sup>[48]</sup> 对飞蝗的小 RNA 进行了测序, 与 mirBase 数据库比对鉴定出 50 个保守 miRNA 家族, 经生物信息分析确定了飞蝗特有的 185 个 miRNAs 家族, 并发现无脊椎动物、体腔动物、虫类是 miRNAs 集中产生的 3 个主要生物进化阶段, 此外还发现一些来自转座子

的内源小 RNAs 和 piRNAs 也参与很多基因的表达调控。

### 3 利用转录组技术研究野生大豆基因资源的应用前景

野生种和作物近缘种是植物种质资源的重要组成部分。野生大豆经过了上千年的自然演变, 具有蛋白质含量高、单株荚数多、氨基酸种类丰富、抗病虫以及环境耐受力强的特点, 是栽培大豆种质改良中的重要基因来源。我国野生大豆资源丰富, 挖掘野生大豆遗传资源潜力, 利用优异基因资源丰富大豆育种基因库, 拓宽大豆育种遗传基础, 对改良栽培大豆将起到不可估量的作用。许多研究者将野生大豆和现有大豆品种进行杂交, 通过将野生大豆有益基因融合到栽培大豆中, 促进野生种质向栽培大豆遗传渗透, 以达到改良栽培大豆的目的<sup>[48]</sup>。然而, 目前野生大豆的研究主要集中在传统的杂交育种方面, 关于野生大豆基因组学和分子生物学方面的研究较少, 加快野生大豆性状鉴定和优异基因挖掘成为大豆育种工作中迫切需要解决的问题。转录组测序技术凭借诸多的优势已被迅速地广泛应用于全转录组的深入研究, 为全面开展作物分子生物学研究提供了新思路。通过转录组测序技术挖掘野生大豆和栽培大豆的差异基因, 可揭示大豆基因组在驯化过程中发生的变化, 从而找到在人工驯化过程中流失的性状优异的基因, 尤其是那些与抗性相关的基因。转录组测序获得的大量转录本信息, 能够为野生大豆分子生物学研究提供重要的基础和依据, 更重要的是可进一步用于基因克隆、表达谱分析、新基因发现、分子标记开发、遗传图谱构建和功能基因分离等方面, 能够有力推动大豆分子标记辅助育种、转基因育种和分子设计育种逐步成为现实。总之, 利用转录组测序技术挖掘野生大豆优良基因资源, 可有效丰富大豆育种基因库, 从而育成优良大豆品种投入生产实践, 为我国农业可持续发展提供宝贵的物质基础, 对我国国民经济发展和粮食、资源安全具有重大意义。

### 参考文献

- [1] 王克晶, 李福山. 我国野生大豆 (*G. soja*) 种质资源及其种质创新利用[J]. 中国农业科技报, 2000, 2(6): 69-72. (Wang K J, Li F S. General situation of wild soybean (*G. soja*) germplasm resources and its utilization of introgression into cultivated soybean in China[J]. Review of China Agricultural Science and Technology, 2000, 2(6): 69-72.)
- [2] 胡小梅, 张必弦, 朱延明, 等. 野生大豆资源的研究与利用[J].

- 安徽农业科学,2011,39(22):13311-13313. (Hu X M,Zhang B X,Zhu Y M,et al. The studies and utilization of wild soybean(*Glycine soja*) [J]. Journal of Anhui Agricultural Sciences,2011,39(22):13311-13313. )
- [3] 齐宁,林红. 利用野生大豆资源创新优质抗病大豆新种质[J]. 植物遗传资源学报,2005,6(2):200-203. (Qi N,Lin H. Using wild soybean resources to develop the new soybean germplasm of high quality and diseases resistance[J]. Journal of Plant Genetic Resources,2005,6(2):200-203. )
- [4] Wang Z H, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics[J]. Nature Reviews Genetics,2009,10:57-63.
- [5] Wilhem B T, Landry J R. RNA-Seq-quantitative measurement of expression through massively parallel RNA sequencing[J]. Methods,2009,48:249-257.
- [6] Velculescu V E, Zhang L, Zhou W, et al. Characterization of the yeast transcriptome[J]. Cell,1997,88:243-251.
- [7] Eveland A L, Satoh-Nagasawa N, Goldshmidt A, et al. Digital gene expression signatures for maize development[J]. Plant Physiology,2010,154:1024-1039.
- [8] Wang Q Q, Liu F, Chen X S, et al. Transcriptome profiling of early developing cotton fiber by deep-sequencing reveals significantly differential expression of genes in a fuzzless/lintless mutant[J]. Genomics,2010,96:369-376.
- [9] 莫德林,赵晓,李安宁. 应用深度测序技术对脂肪型猪和瘦肉型猪的不同发育时期的骨骼肌进行比较转录谱研究[C]. 第十六次全国动物遗传育种学术讨论会论文集,2011:340. (Mo D L,Zhao X,Li A N. The research of compared transcription between lean and obese swine in different development stages through application the depth sequencing technology [C]. The Sixteenth National Animal Genetic Breeding of Academic Symposium,2011:340. )
- [10] 来永才,林红,方万程,等. 野生大豆资源在大豆种质拓宽领域中的应用[J]. 沈阳农业大学学报,2004,35(3):184-188. (Lai Y C,Lin H,Fang W C,et al. Wild soybean natural resources and its application in broadening of soybean germplasm[J]. Journal of Shenyang Agricultural University,2004,35(3):184-188. )
- [11] 杨振宇,本多健一郎,王曙明,等. 中国东北抗蚜野生大豆重复鉴定的研究[J]. 吉林农业科学,2004,29(5):3-6. (Yang Z Y,Kanichilo Honda,Wang S M,et al. The repetitive identification on Aphid-resistant wild soybean in northeast of China[J]. Journal of Jilin Agricultural Sciences,2004,29(5):3-6. )
- [12] 马晓萍,杨光宇,杨振宇,等. 野生大豆在大豆育种中的应用[J]. 作物研究,2009,23(2):12-13. (Ma X P,Yang G Y,Yang Z Y. Application of wild soybean in breeding[J]. Crop Research,2009,23(2):12-13. )
- [13] 李庆岗,陶立. 高通量测序技术及其在生命科学中的应用[J]. 畜牧与饲料科学,2012,33(2):25-28. (Li Q G,Tao L. High-throughput sequencing technology and its application in life science[J]. Animal Husbandry and Feed Science,2012,33(2):25-28. )
- [14] 井赵斌,魏琳,余靓,等. 牧草转录组测序及其在牧草基因资源发掘中的应用前[J]. 草业科学,2010,28(7):1364-1369. (Jing Z B,Wei L,Yu L,et al. Transcription sequencing and its application prospective on discovering the gene resources of forages[J]. Pratacultural Science,2010,28(7):1364-1369. )
- [15] 岳桂东,高强,罗龙海,等. 高通量测序技术在动植物研究领域中的应用[J]. 中国科学,2012,42(2):107-124. (Yue G D,Gao Q,Luo L H,et al. The application of high-throughput sequencing technology in plant and animal research[J]. Scientia Sinica Vitae,2012,42:107-124. )
- [16] 周华,张新,刘腾云,等. 高通量转录组测序的数据分析与基因发掘[J]. 江西科学,2012,30(5):607-611. (Zhou H,Zhang X,Liu T Y,et al. Data processing and gene discovery of high-throughput transcriptome sequencing[J]. Jiangxi Science,2012,30(5):607-611. )
- [17] Li R,Fan W,Tian G,et al. The sequence and de novo assembly of the giant panda genome[J]. Nature,2010,463:311-317.
- [18] Ashburner M,Ball C A,Blake J A, et al. Gene ontology: tool for the unification of biology[J]. The Gene Ontology Consortium, Nature Genetics,2000,25:25-29.
- [19] Altermann E,Klaenhammer T R. Pathway voyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database[J]. BMC Genomics,2005,6:60.
- [20] Wixon J,Kell D. The Kyoto encyclopedia of genes and genomes-KEGG[J]. Yeast,2000,17:48-55.
- [21] 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. (Qi Y X,Liu Y B,Rong W H. RNA-Seq and its application: a new technology for transcriptomics[J]. Hereditas,2011,33(11):1191-1202. )
- [22] 王兴春,杨致荣,王敏,等. 高通量测序技术及其应用[J]. 中国生物工程杂志,2012,32(1):109-114. (Wang X C,Yang Z R,Wang M,et al. High-throughput sequencing technology and its application[J]. China Biotechnology,2012,32(1):109-114. )
- [23] 梁焯,陈双燕,刘公社. 新一代测序技术在植物转录组研究中的应用[J]. 遗传,2011,33(12):1317-1326. (Liang Y,Chen S Y,Liu G S. Application of next generation sequencing techniques in plant transcriptome [J]. Hereditas,2011,33(12):1317-1326. )
- [24] 翟焕趁,连璧,王杏杏,等. 稻瘟病菌菌株 FJ81278 中新发现的一个基因注释及其表达分析[J]. 福建农林大学学报,2011,40(6):570-575. (Zhai H C,Lian B,Wang X X,et al. Annotation and expression analysis of a novel unigene in *Magnapothae oryzae* isolate FJ81278[J]. Journal of Fujian Agriculture and Forestry University,2011,40(6):570-575. )
- [25] Filichkin S A,Priest H D,Givan S A, et al. Genome-wide mapping of alternative splicing in *Arabidopsis thaliana* [J]. Genome Research,2010,20:45-58.
- [26] Campbell M A,Haas B J,Hamilton J P, et al. Comprehensive analysis of alternative splicing in rice and comparative analyses with *Arabidopsis* [J]. BMC Genomics,2006,7:327.
- [27] 魏利斌,苗红梅,张海洋. 芝麻发育转录组分析[J]. 中国农业科学,2012,45(7):1246-1256. (Wei L B,Miao H M,Zhang H Y. Transcriptome analysis of sesame development[J]. Scientia Agricultura Sinica,2012,45(7):1246-1256. )
- [28] 程立宝,李淑艳,李岩,等. 莲藕根状茎膨大过程中淀粉合成相关基因的表达[J]. 中国农业科学,2012,45(16):3330-3336. (Chang L B,Li S Y,Li Y,et al. Expression of genes related to starch synthesis during rhizome swelling of lotus root(*Nelumbo nucifera Gaertn*) [J]. Scientia Agricultura Sinica,2012,45(16):3330-3336. )

- [29] Zhang G, Guo G, Hu X, et al. Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome [J]. *Genome Research*, 2010, 20:646-654.
- [30] Sara Z, Alberto F, Enrico G, et al. Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq [J]. *Plant Physiology*, 2010, 152:1787-1795.
- [31] Hegedüs Z, Zakrzewska A, Ágoston V C, et al. Deep sequencing of the zebrafish transcriptome response to mycobacterium infection [J]. *Molecular Immunology*, 2009, 46:2918-2930.
- [32] Alagna F, D'Agostino N, Torchia L, et al. Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development [J]. *BMC Genomics*, 2009, 10:399.
- [33] Julio V, Enrique I, Beatriz J, et al. Deep sampling of the Palomero maize transcriptome by a high throughput strategy of pyrosequencing [J]. *BMC Genomics*, 2009, 10:299.
- [34] Novaes E, Drost D R, Farmerie W G, et al. High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome [J]. *BMC Genomics*, 2008, 9:312.
- [35] Liu G, Li X, Chen S, et al. Transercriptomics and gene discovery in *Leymus chinensis* via next generation sequence technology [C]. The 6th International Symposium on the Molecular Breeding of Forage and Turf, 2010.
- [36] Sunkar R, Zhou X, Zheng Y, et al. Identification of novel and candidate miRNAs in rice by high throughput sequencing [J]. *BMC Plant Biology*, 2008, 8:25-41.
- [37] Lu C, Tej S S, Luo S, et al. Elucidation of the small RNA component of the transcriptome [J]. *Science*, 2005, 309:1567-1569.
- [38] Yao Y, Guo G, Ni Z, et al. Cloning and characterization of microRNAs from wheat (*Triticum aestivum* L.) [J]. *Genome Biology*, 2007, 8:96-108.
- [39] Morin R D, Oconnor M D, Griffith M, et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells [J]. *Genome Research*, 2008, 18:96-108.
- [40] Huang J M, Ju Z H, Li Q L, et al. Solexa sequencing of novel and differentially expressed microRNAs in testicular and ovarian tissues in Holstein cattle [J]. *International Journal of Biological Sciences*, 2011, 7:1016-1026.
- [41] Graham I A, Besser K, Blumer S, et al. The genetic map of *Artemisia annua* L. identifies loci affecting yield of the antimalarial drug artemisinin [J]. *Science*, 2010, 327:328-331.
- [42] Parchman T L, Geist K S, Grahnen J A, et al. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery [J]. *BMC Genomics*, 2010, 11:180.
- [43] Shu Y J, Li Y, Zhu Z L, et al. SNPs discovery and CAPS marker conversion in soybean [J]. *Molecular Biology Reporter*, 2011, 38:1841-1846.
- [44] Wang Z, Fang B, Chen J, et al. De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*) [J]. *BMC Genomics*, 2010, 11:726.
- [45] 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.
- [46] Hao D C, Ge G B, Xiao P G, et al. The first insight into the tissue specific taxus transcriptome via illumina second generation sequencing [J]. *PLoS One*, 2011, 6:e21220.
- [47] Timme R E, Delwiche C F. Uncovering the evolutionary origin of plant molecular processes: comparison of *Coleochaete* (*Coleochaetales*) and *Spirogyra* (*Zygnematales*) transcriptomes [J]. *BMC Plant Biology*, 2010, 10:96.
- [48] Wei Y, Chen S, Yang P, et al. Characterization and comparative profiling of the small RNA transcriptomes in two phases of locust [J]. *Genome Biology*, 2009, 10:R6.
- [49] 田清震. 中国野生大豆与栽培大豆 AFLP 指纹分析及生态群体遗传关系研究 [D]. 南京:南京农业大学, 2000. (Tian Q Z. AFLP fingerprint analysis and genetic relationship among eco-types of *G. soja* and *G. max* in China [D]. Nanjing: Nanjing Agricultural University, 2000.)

## 欢迎订阅 2014 年《北方园艺》

全国自然科学(中文)核心期刊  
中国农业核心期刊  
全国优秀农业期刊

中国北方优秀期刊  
黑龙江省优秀科技期刊  
美国化学文摘社(CAS)收录期刊

主管:黑龙江省农业科学院

主办:黑龙江省农业科学院、黑龙江省园艺学会

刊号:ISSN 1001-0009 CN 23-1247/S

邮发代号:14-150 半月刊 每月 15、30 日出版

单价:7.00 元 全年:168 元

全国各地邮局均可订阅,或直接向编辑部汇款订阅。

本刊内容丰富、栏目新颖、技术实用、信息全面。涵盖园艺学的蔬菜、果树、瓜类、花卉、植保等研究领域的新成果、新技术、新品种、新经验。竭诚欢迎全国各地科研院所人员、大专院校师生,各省、市、县、乡、镇农业技术推广人员、农民科

技示范户等踊跃订阅。

栏目:试验研究、研究简报、设施园艺、栽培技术、园林花卉、生物技术、植物保护、贮藏保鲜加工、食用菌、中草药、新品种选育、土壤与肥料、产业论坛、专题综述、经验交流、农业经纬等。

地址:黑龙江省哈尔滨市南岗区学府路 368 号《北方园艺》编辑部

邮编:150086

电话:0451-86674276

网址:bfyy.haasep.com

信箱:bfyybjb@163.com