

植物功能基因组研究进展

邢俊杰^{1,2}, 梁 铃³, 曹孟良²

(1. 福建农林大学, 中国福建 福州 350002; 2. 湖南杂交水稻技术研究中心, 中国湖南 长沙 410125;
3. 广西大学, 中国广西 南宁 530005)

摘要:随着植物基因组计划的深入, 植物基因组学研究的重点已经转变为基因组功能的研究, 即利用基因组序列的信息和高通量的系统分析技术, 在基因组水平研究植物结构和组织与植物功能在细胞、有机体和进化上的关系, 对功能基因组学研究的内容、方法以及最新研究进展进行了综述。

关键词:植物; 功能基因组学; 生物芯片; 蛋白质组; 生物信息学

中图分类号: Q756

文献标识码: A

文章编号: 1007-7847(2005)S0-0022-05

Advances in Plant Functional Genomics

XING Jun-jie^{1,2}, LIANG Ling³, CAO Meng-liang²

(1. Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China; 2. Hunan Hybrid Rice Research Center, Changsha 410125, Hunan, China; 3. Guangxi University, Nanning 530005, Guangxi, China)

Abstract: Following the progress in the plant genomic project, the emphases of plant genomic research have changed into genomic functional research, that is, by using the information in genomic sequence and high-fluxing system analyzing technology research the relation between the tissue structure and its function in the cell, organism, evolution that are based on the genomic level. The content, methods and recent research progress in genomic functional research are reviewed.

Key words: plant; functional genomics; biochip; proteome; bioinformation

(*Life Science Research*, 2005, 9(2): 022 ~ 026)

随着部分模式植物基因组测序的完成, 功能基因组的研究愈发受到重视。功能基因组学是利用结构基因组学提供的信息, 运用高通量序列分析技术、大规模实验技术、计算机统计分析技术和生物信息学来研究基因功能。植物的生长和发育是一个有机体或有机体的一部分形态建成和功能按一定次序而进行的一系列生化代谢反应的总合。反应在分子水平上, 它要求相应的遗传代谢途径必须按照特定的时空次序严格进行以保证正常发育。植物功能基因组研究就是要利用植物全基因组序列的信息, 通过发展和应用系统基因组水平的实

验方法来研究和鉴别基因组序列的作用; 研究基因组结构和组织与植物功能在细胞、有机体和进化上的关系以及基因与基因间的调控关系; 从表达时间、部位和表达水平 3 个方面对目的基因在植物中的精细调控进行系统研究。

1 功能基因组的研究内容

在功能基因组的研究内容方面, 国内外都注重尽可能获取更多的全长基因的表达序列、饱和突变体库的创建、高通量的转基因遗传分析和比较基因组学研究, 以便通过改变基因的表达状况

收稿日期: 2005-03-16; 修回日期: 2005-06-17

作者简介: 邢俊杰(1981-), 男, 山东烟台人, 硕士研究生, 主要从事功能基因组研究。Tel: 0731-8122127。

寻找基因与性状的联系,而最终阐明基因的功能。“功能基因组学”研究的基本策略是以整体的观点,对转录组和蛋白组以及代谢组进行系统的高通量的分析。即从一次研究一个基因或一个蛋白、特定的组分转变到一次系统分析所有基因或蛋白质以及多种代谢途径及其代谢产物的变化。在植物功能基因组学的研究中,拟南芥和水稻是两种最常用的模式植物,对模式植物功能基因组的研究将有助于整个植物基因组学的研究。

功能基因组主要包括基因组(genomics)、转录组(transcriptome)、蛋白质组(Proteome)以及代谢组等几个方面。基因组(genome)是指某一生物的所有DNA。基因组学是一门研究生物体的整个基因组的科学;转录组是指某一生物或细胞所有基因表达的RNA。转录组学是一门对某一生物或细胞所有基因表达的RNA进行全面分析的科学;蛋白质组是指某一生物或细胞在各种不同环境条件下表达的所有蛋白质。蛋白质组学是一门对某一生物或细胞在各种不同环境条件下表达的所有蛋白质进行定性和定量分析的科学。转录组学和蛋白质组学是分别在基因的转录和转录后的蛋白质翻译与修饰两个水平上研究基因的功能。一个细胞的基因转录表达水平能够精确而特异地反映其类型、发育阶段以及应激状态,是功能基因组学的主要研究内容之一。目前的功能基因组研究主要包括以下几个方面:

- 1) 全长 cDNA 克隆与测序;
- 2) 获得 DNA 芯片等基因转录图谱;
- 3) 突变体库的构建;
- 4) 高通量的遗传转化鉴定系统;
- 5) 生物信息技术平台与相应数据库的构建;
- 6) 研究基因组表达的全部蛋白质及其相互作用为主要内容的蛋白质组学。

2 功能基因组研究的主要策略

植物功能基因组学是一个崭新的研究领域,其研究方法正在不断发展和完善。目前,获得一段DNA序列的功能信息的最简单的方法是将该DNA序列与GenBank中公布的基因序列进行同源性比较,如利用BLASTn和BLASTx两种软件分别进行核苷酸和氨基酸序列同源性比较等。上述同源性检索分析方法仅仅为该DNA片段的功能提供了间接的证据,对基因功能的直接证据还需要实验上的数据。Bouchez和Hofte将所需要的

实验证据归纳如下:

- 1) 通过研究基因的时空表达模式确定其在细胞学或发育的功能,如在不同细胞类型、不同发育阶段、不同环境条件下以及病原菌感染过程中mRNA和蛋白质的表达的差异等;

- 2) 研究基因在亚细胞内的定位和蛋白质的翻译后调控等;

- 3) 利用基因敲除(knock-out)技术进行功能丧失(loss of function)分析或通过基因的过量表达(转基因)进行功能获得(gain of function)分析,进而研究目的基因与表型性状间的关系;

- 4) 通过比较研究自发或诱发突变体与其野生型植株在特定环境条件下基因表达的差异来获取基因功能的可能信息。

为了功能基因组的深入研究,人们相继建立了一系列新的技术和方法,包括功能基因表达克隆(functional cloning)、基因芯片(gene chip)或基因微阵列技术(microarray)、基因表达系统分析(Serial analysis of gene expression, SAGE)或表达序列标签(expressed sequence tag, EST)、蛋白质组技术(proteomics)、质谱测序技术、生物信息学(bioinformatics)、插入突变(insertion mutagenesis)等等,这些前沿技术为功能基因组学的研究提供了强有力的技术发挥。

2.1 表达序列标签测序技术(EST)

通过从cDNA文库中随机挑取的克隆进行测序所获得的部分cDNA的5'或3'端序列称为表达序列标记(EST),一般长300~500bp左右,利用EST作为标记所构建的分子遗传图谱被称为转录图谱。目前植物EST计划主要集中在拟南芥和水稻上。这些EST不仅为植物基因组遗传图谱的构建提供了大量的分子标记,而且来自不同组织和器官的EST也为基因的功能研究提供了有价值的信息:当一个已知功能和特征的基因从一个植物中克隆出来后,就可以通过EST数据库去鉴定另一种植物中的具有同样特征的直向同源基因。此外,EST计划还为基因的鉴定提供了候选基因(candidates)。EST的一个不足之处在于通过随机测序有时难以获得那些低丰度表达的基因和那些在特殊环境条件下(如生物胁迫和非生物胁迫)诱导表达的基因。随着基因组测序计划的完成,这个缺点可以得到弥补。

2.2 插入突变

通常应用于植物方面的插入突变有两种类

型:转座因子(transposons)和农杆菌介导的转移DNA(Transfer DNA, T-DNA)。机理是:外源序列的插入,导致内源基因表达受阻,出现突变体表现型。

转座子是存在于染色体DNA上的一段可自我复制和位移的DNA序列。转座子的转位插入作用,使被插入的目的基因发生突变失去活性,而转座子的删除又可使目的基因恢复活性。因此利用转座子的转位插入作用可对被插入的目的基因进行基因功能的研究,这就是转座子标签技术。转座子标签技术已成为植物基因鉴定、分离的有效方法。

土壤根瘤农杆菌可将自身携带的Ti质粒上的一段T-DNA,通过侵染植物伤口而转移到植物基因组中。T-DNA的插入可引起植物基因突变,它一般是通过阻断正常基因的表达而表现为敲除突变。通过对T-DNA两侧基因序列作特异扩增等方法可分离野生型的基因。这种方法可在预先不知道基因的产物和表达的情况下分离基因。

插入性诱变最大的技术障碍是:转化过程中经常产生体细胞克隆变异。这种变异不是插入诱变的结果,而是组织培养转化体系诱变的产物。但随着转化方法的改进,将降低非标记突变体的出现。

2.3 基因表达系统分析(SAGE)

Velculescu等人1995年发明了基因表达系列分析技术。这是一种高通量且快捷有效的基因表达研究技术,可用于研究任何一种由细胞转录变化引起的生物现象,而无须对基因性质和生物系统预先有所了解。

SAGE是以转录子(cDNA)上特定区域9~10bp的寡核苷酸序列作为标签(tag)来特异性代表该转录子。然后通过连接酶将多个标签(一般为20~60个)随机串联并克隆到载体中,建立SAGE文库。通过对双标签的序列分析,可获得基因转录的分布以及表达丰度的情况(尤其是可检测到低丰度表达的基因),从而可充分了解基因转录组的全貌。SAGE技术已广泛用于全面获取生物基因的表达信息,定量比较不同状态下的基因表达,寻找新基因等研究领域。

2.4 基因芯片

基因芯片技术,是指采用特殊的手段,将大量基因探针分子按一定的方式密集排列并固定在面积有限的特殊材料表面,通过未知样品中的生物分子与芯片上所载的已知分子间的杂交,反应

结果用同位素法、化学发光法或酶标法显示,然后用精密的扫描仪或CCD摄像技术记录,通过计算机分析,形成可读信息,实现对细胞、蛋白质、DNA以及其他生物组分的准确、快速、大信息量的检测。通常比较典型的DNA芯片制备方法有4种:

第1种方法是Affymetrix公司开发的光引导原位合成法,该方法是微加工技术中光刻工艺与光化学合成法相结合的产物;第2种方法是Incyte Pharmaceutical公司采用的化学喷射法,该方法是将合成好的核苷酸探针定点喷射到芯片上并加以固定化来制作DNA芯片;第3种方法是斯坦福大学研制的接触式点涂法。在DNA芯片制备中通过高速精密机械手的精确移动让移液头与玻璃芯片接触,而将DNA探针涂敷在芯片上;第4种方法是通过使用4支分别装有A, T, G, C核苷的压电喷头在芯片上并行地合成出DNA探针。

根据制备方法的不同可分为:DNA芯片(DNA chip)、微点阵(Microarray)和电子芯片。

DNA芯片是利用原位合成法或将一系列寡核苷酸以预先设定的排列方式固定在固相支持介质表面(硅片、玻片、尼龙膜等),形成高密度的寡核苷酸的阵列,以用于杂交。由于链内互补序列及Tm值等因素的限制,其中寡核苷酸的长度一般为25bp左右。此种芯片可用于DNA的序列测定、突变的检测及基因转录表达分析。

微点阵,多指通过点样法制备的中、低密度点阵芯片。所用支持物一般为玻片、尼龙膜等。玻片需经特殊的处理,多是经过多聚赖氨酸等进行包被或进行其他形式的化学修饰。点样于其上的探针可以是cDNA片段,基因组片段、寡核苷酸或核酸类似物如PNA等。微阵数据分析的一个主要用途是在特定的组织和代谢过程中鉴定有相关作用的基因。

电芯片是带有阳电荷的硅芯片、芯片经热氧化,制成1mm×1mm的阵列,每个阵列含多个微电极,每个电极上通过氧化硅沉积和蚀刻制备出样品池。将连接链亲和素的琼脂糖覆盖在电极上,电场作用下生物素标记的探针即可结合在特定电极上。

基因芯片技术的主要应用领域为基因表达谱分析、新基因发现、基因突变及多态性分析以及监测基因组在不同时期断面上的整体转录表达状况。因此,基因芯片技术已成为功能基因组学研究中一项非常重要和关键的实验技术。

2.5 蛋白质组技术

由于基因芯片技术只能反映从基因组到 RNA 的转录水平上的表达情况, 而从 RNA 到蛋白质须经过许多中间环节的影响, 因此仅凭基因芯片技术我们还不能最终掌握生物功能具体执行者——蛋白质的整体表达状况。

蛋白质组是指基因组表达的全部蛋白质及其存在方式。蛋白质组学旨在阐明生物体全部蛋白质的表达模式及功能模式, 其内容包括鉴定蛋白质的表达、存在方式(修饰形式)、结构、功能和相互作用等。利用蛋白质组研究植物功能基因组可以得到以下 3 方面信息: 从基因序列预测的基因产物的翻译情况; 基因产物的相对浓度; 基因产物翻译后的修饰程度。蛋白质组学将基因表达的数据与植物代谢和植物表型的问题紧密连在一起, 既可以用于研究植物生理机制, 又可以用于研究未知功能的蛋白质。蛋白质组学研究技术和方法很多, 并不断发展和出现新的技术。

2.5.1 双向聚丙烯酰胺凝胶电泳(2D-PAGE)

迄今为止, 它仍然是分离蛋白质的最有效的方法。与基因组研究不同的是, 蛋白质组学并没有类似于 PCR 反应的扩增方法, 因此, 分离样品的精确性就成了至关重要的问题。随着荧光染料应用, 以及 2D-PAGE 相关技术, 如高度敏感的质谱(Mass-spectrometric)和 EST 数据库的发展, 使分离和鉴定蛋白质的工作进一步得到提高。

2.5.2 质谱(Mass-spectrometric)技术

质谱技术是近年来蛋白质组学研究最重要的技术突破之一, 其原理是将样品分子离子化, 根据离子间质荷比(m/z)的差异来分离并确定质量, 是高灵敏度高特异性地快速鉴定生物分子的技术。

质谱鉴定有两条主要途径。一是“肽链质量图谱”途径。测量质谱的方法是基质辅助激光解吸附/电离法, 通过测定一个蛋白质酶解混合物中肽段的电离飞行时间来确定其分子量等数据, 最后通过相应的数据库搜索鉴定蛋白质; 二是串联质谱途径, 将胰蛋白酶消化后的蛋白质单个肽链直接从液相经“电喷离子化”而被电离, 分解为氨基酸或含有 C 末端的片段, 片段化离子被喷射到“串联质谱仪”进行质量测定, 以得到序列信息。

质谱技术还可用于蛋白质磷酸化、硫酸化、糖苷化以及其它一些修饰的研究。

2.5.3 蛋白质芯片(Protein chips)技术

蛋白质芯片技术是一种高通量、微型化和自动化的蛋白质分析技术。在蛋白质芯片技术途径中, 首先将一系列的蛋白质(如抗体)按照一定的排列格式固定在经特殊处理的材料表面上。然后以我们感兴趣的样品为探针来探查该表面, 那些与相应的抗体相结合的蛋白质就会被吸附在表面上。而后把未与抗体结合的蛋白质洗掉, 把结合的蛋白质洗脱下来, 经凝胶电泳之后通过质谱法进行鉴定。这种技术实际上是一种大规模的酶联免疫分析。蛋白质芯片制作过程中最重要的是保持蛋白质的活性。

蛋白质之间的相互作用是蛋白质组研究中的一个关键问题。蛋白质芯片能够同时分析上千种蛋白质的变化情况, 使得在基因组水平研究蛋白质的功能(如酶活性、抗体的特异性、配体-受体交互作用以及蛋白质与蛋白质或核酸或小分子的结合)成为可能。

2.6 生物信息学

生物信息学是分子生物学和计算机信息处理技术结合后研究生命现象、组织和分析呈指数增长的生物学数据的一门学科。生物信息学由数据库、计算机网络和应用软件 3 大部分组成。它是利用计算机科学、数学和统计学的方法, 通过搜索、分析、比较大量的基因和蛋白质序列数据, 建立理论模型, 进行基因结构的鉴定、分子设计、蛋白质结构预测、蛋白质结构与功能关系的研究。该学科以计算机和生物电子设备为主要工具, 发展各种软件, 对日益增长的 DNA 和蛋白质的序列结构进行收集、整理、储存、发布、提取、加工、分析和发现。生物信息学的研究目标是揭示“基因组信息结构的复杂性及遗传语言的根本规律”。

生物信息学研究的重要内容有: 大规模基因组测序中的信息分析; 大规模基因功能表达谱的分析; 新基因和新 SNP (single nucleotide polymorphisms) 的发现; 与鉴定密码起源和生物进化的研究等。

随着功能基因组研究的深入, 生物信息学发挥的作用越来越大, 这将推动功能基因组的快速发展。

3 小结

随着人类基因组计划、植物基因组计划和微生物基因组计划的出现, 生物科学的研究已进入

后基因组时代. 基因组学的研究也从结构基因组学转向功能基因组学研究. 结构基因组学是以建立生物体遗传、物理和转录图谱为主, 是基因组研究的初级阶段. 功能基因组学是利用结构基因组学提供的信息, 运用高通量序列分析技术、大规模实验技术、计算机统计分析技术和生物信息学来研究基因功能, 是基因组研究的高级阶段. 植物功能基因组学的研究将不仅会使我们全面、客观地了解基因的功能, 而且将有助于我们更好地利用这些研究结果定性地改造植物性状, 使其更好地服务于人类社会. 相信随着对基因调控规律的进一步揭示和蛋白质结构与功能的进一步研究, 植物功能基因组学的研究一定能够在农业的高产、优质、超高产育种以及人类重大疾病的防治等方面做出重大贡献.

参考文献 (References):

- [1] BOUCHEZ D, HOFTE H. Functional genomics in plants[J]. *Plant Physiol*, 1998, 118: 725-732.
- [2] DONSON J, FANG Y, ESPIRITU-SANTO G, *et al.* Comprehensive gene expression analysis by transcript profiling[J]. *Plant Molecular Biology*, 2002, 48: 75-97.
- [3] ROBERTS J K. Proteomics and a future generation of plant molecular biologists[J]. *Plant Molecular Biology*, 2002, 48: 143-154.
- [4] SOMERVILLE C, SOMERVILLE S. Plant functional genomics [J]. *Science*, 1999, 285: 383-385.
- [5] KAISER J. From genome to functional genomics[J]. *Science*, 2000, 288: 1715-1716.
- [6] ZIVY M, VIENNE D. Proteomics: a link between genomics, genetics and physiology[J]. *Plant Molecular Biology*, 2000, 44: 575-580.
- [7] 梁宇. 植物蛋白质组学研究进展[J]. *植物生态学报*, 2004, 28(1): 114-125.
- [8] 李子银, 陈受宜. 植物的功能基因组学研究进展[J]. *遗传*, 2000, 22(1): 57-60.
- [9] 黎裕, 王天宇. 植物功能基因组学的发展现状与发展趋势[J]. *生物技术通报*, 2000, 4: 6-10.
- [10] 周晓骥, 肖乃什, 白云峰, 等. 植物功能基因组学的研究策略[J]. *中国生物工程杂志*, 2002, 22(6): 13-17.
- [11] 于玲, 王莱, 牛吉山, 等. 植物功能基因组研究进展[J]. *西北师范大学学报*, 2003, 39(1): 104-116.
- [12] 吴鹏, 阎松. 生物芯片技术的发展与应用[J]. *现代电子技术*, 2003, 149(6): 1-4.
- [13] 陈力学. 生物信息学在基因组研究中的应用[J]. *国外医学临床生物化学与检验学分册*, 2003, 24(6): 339-340.
- [14] MARTIENSSEN R A. Functional genomics: probing plant gene function and expression with transposons[J]. *Proc Natl Acad Sci USA*, 1998, 95: 2021-2026.
- [15] RAMSAY G. DNA chips: state of the art[J]. *Nature biotechnology*, 1998, 16: 40-44.