

学科发展

植物功能基因组研究^{*}

韩 斌

(上海生命科学研究院国家基因研究中心 上海 200233)

摘要 植物功能基因组研究是利用基因组序列的信息和高通量的系统分析技术,在基因组水平研究其结构和组织与植物功能在细胞、有机体和进化上的关系。水稻作为最重要的模式作物,应该成为我国植物功能基因组研究的主攻方向。通过建立从结构基因组学到功能基因组学,直至将功能基因组的信息和知识用于开发改良植物的完整研究体系和队伍,使我国在水稻功能基因组研究的整体水平上取得与我国头号水稻种植大国相称的地位。

关键词 功能基因组,水稻,拟南芥



1 植物功能基因组研究概述

植物功能基因组研究是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。分析的重点已从单个基因转向整个生物体的全部基因系统。同时,植物功能基因组研究还将从表达时间、部位和表达水平三个方面对目的基因在植物中的精细调控进行系统研究。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。

植物是地球上高等生物赖以生存的基础,整个生物圈都依赖于植物提供食物和氧气,而且有超过

25 万种开花植物点缀了地球,它是地球生态环境的主要组成部分。植物基因组研究较之人的基因组研究有不同的特点。植物通过光合作用这一地球上最重要的化学反应,将简单的无机物(CO_2 和 H_2O)合成为碳水化合物,为自身生长发育提供物质和能量来源,同时也为生物提供食物来源;植物的生长还需要不断从土壤中吸收水分、矿物质和有机质养分;植物在自然界的生长又一直受到各种生物(病原微生物和昆虫)和非生物(旱、涝、寒冷、炎热、盐碱等)因素的胁迫和干扰。因此,研究与植物生长发育的基本代谢和与重要农作物的优质、高产、抗病、抗逆密切相关基因的功能及调控关系,是植物功能基因组研究的核心内容。

鉴于上述,植物王国里最先被测序并详细分析的基因组就成了里程碑式的成就。2000 年 12 月 *Nature* 发表了十字花科植物拟南芥(*Arabidopsis thaliana*)的全基因组序列分析结果^[1];2002 年,作为另一种模式植物和人类主要粮食作物——水稻的全基因组序列测定也已基本完成,即:华大基因研究中心和美国 Syngenta 公司完成了对籼稻 93-11 和粳稻日本晴的全序列草图^[2-3];由日本、中国和美国等 10 个国家和地区参加的国际水稻基因组测序计划(IRGSP)完成了水稻粳稻日本晴基因组精确测

^{*} 收稿日期:2003 年 6 月 20 日

序及第 1、4、10 等几条染色体的完成图^[4-6]。大规模基因组测序属结构基因组学 (Structural Genomics)，而从结构基因组的庞大数据中发掘出有用信息和通过各种实验鉴定基因的功能及基因间的相互功能则是功能基因组学 (Functional Genomics) 的内容了。这两个分别代表双子叶和单子叶模式植物的全基因组序列的完成，拉开了植物功能基因组研究的帷幕。

获得基因组序列只是更艰巨任务的开始。首先必须对基因组序列进行注解 (Annotation)，预测和鉴定基因组的全部基因信息；其次必须得到所有基因的转录物 mRNA (不同组织、不同时期及生长状况下的转录谱)；最后研究翻译产物，阐明其表达、代谢、调控的途径及与其它生物分子的相互作用。拟南芥全基因组序列的注解显示它共有 25 498 个预测基因，但仅有 10% 的基因可通过实验的方法获得证实，而其它 90% 的基因或者不知其功能，或者其推测的功能尚未得到实验完全证实^[1]。因此，确定这些序列和基因的功能，并在此基础上揭示各种生命现象所涉及到的基因及其表达调控的途径，最终阐明整个基因组的功能就成了目前的主要任务。基因的功能主要包括生物化学、细胞、发育及适应性等^[8]。在拟南芥中，预测的全部基因中约 54% 的新基因与已在其它物种上认知的基因具有序列相似性。对已知序列进行数据库比较是最简单的获知其功能信息的办法。然而仅仅数据分析还不足以确定基因的功能，还需通过实验证实。例如，细胞发育功能的间接信息可以通过基因的时序表达模式来获得；基因敲除或过度表达使基因可以与表型联系起来；在不同环境中比较野生型与突变株可以获知其适应性功能等。

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 (Proteomics)。由于基因组是相对稳定的，但转录物及蛋白质却在不断变化中，而且由于转录后调控及翻译后修饰，通过 DNA 芯片检测出的 mRNA 表达

水平与相应蛋白质的表达水平并不完全相同，因此，仅有 mRNA 的表达信息还不够，蛋白质的表达似乎更重要。

2 国内外发展现状及趋势

国际上在植物功能基因组研究方面已成规模。各国近年的努力主要集中在构建功能基因组研究技术平台，以争取为大规模功能基因组研究打下坚实的基础。随着研究工作的进展，功能基因组技术平台的内容还将不断扩展。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寻找基因与性状的联系，而最终阐明基因的功能。

在模式植物拟南芥中，国际上已用多种途径建立了 T-DNA (或转座子) 插入饱和的突变体库^[7]。在水稻突变体库的创建方面，国际竞争十分激烈，如荷兰的国际植物研究公司、韩国 Pohang 科技大学和日本农业资源研究所等均已获得了数以万计的突变体。由于水稻是一种重要的粮食作物，加之各研究单位在研究中均不同程度地得到了大的生物技术公司的资助，这些材料难以实现全球共享。

由于 cDNA 对功能基因组研究的极端重要性，国内近年来在各种项目的支持下，武汉、北京、上海等几家单位已启动了突变体的创建工作。在表达序列的 (cDNA 和 EST) 测序方面国际上也已做了大量的工作^[8]，但这些序列大部分出自日本水稻基因组研究计划 (RGP)，而且他们不会大批量地对外提供 cDNA 克隆。在功能基因组研究技术平台的其它方面，如 DNA 芯片技术、生物信息技术及高效遗传转化技术等，目前国内均建立起较好的工作基础。在重要性状或重要生命现象的功能基因组研究方面，国际上对模式植物拟南芥已开展了较多的工作。如通过对突变体库的筛选分离功能基因，通过 cDNA 芯片技术建立性状或诱导特异性基因表达谱、分离功能基因，已取得很好的进展。而在水稻方面，这些工作目前还只处于起步阶段，但这些工作较多地集中在我国的几家研究单位，并形成一定的规模和优势。

目前主要的研究技术和方法如下：

(1)植物转化(Plant Transformation)。通过根农杆菌介导的转化系统,将外源目的基因转入到宿主植物中去,或在目的基因处插入组织专一性启动子使其增强表达。水稻的遗传转化系统在禾本科植物中最为成熟,已经广泛用于大规模创制 T-DNA 插入突变体库和功能基因组研究。

(2)增强子捕获(Enhancer trapping)。将转录激活因子与一个剪切过的微弱启动子相连接,置于 T-DNA 的左边界,当载体插入到某个增强子(或启动子)下游,该增强子就会按其自身的组织特异性,增强插入基因的表达,而该基因又会识别并激活上游序列,使其下游的报告基因得以表达。通过对报告基因产物的筛选,则可知道增强子的存在,故将这一策略称之为“增强子捕获”。

(3)插入突变(Transposon /T-DNA Insertion)。要获得未知基因的功能信息,最直接的办法就是打断目的基因与其表型的联系,制造出功能缺失突变株并研究突变株的表型。能够随机插入染色体中的 DNA 元件(如 Transposon,T-DNA)可以用来在植物中产生功能缺失突变株。由于插入序列已知,用 PCR 可以方便地得到被插入序列的信息。对插入突变株系进行筛选,从表型或其它特征的改变即可获得被插入基因的功能。这里的关键问题是插入饱和水平(即在每个基因至少有一个插入突变的概率,它取决于基因组的大小、结构及插入的随机性),对于 Arabidopsis,估计 120 000 次独立插入可以达到 95%的饱和度。要获得一个突变的作用,细致的分析是必须的。有时获得的插入突变株在通常条件下无明显的表型异常,因此鉴定出合适的环境条件对于表型的正确表达极为重要。另外,在植物基因组中有大量的基因扩增现象,许多基因以多拷贝(或多个同源基因)存在。而且有些基因在细胞生长发育的各个阶段都有重要的功能,一旦被突变掉就会导致细胞死亡,基因与表型之间的联系就将彻底中断。

(4)基因沉默(Gene Silencing)。RNA 在生命活动中具有重要的作用,它和蛋白共同负责基因的表达和调控。近年来的研究表明,植物中有一些小的双链 RNA 能高效、特异地阻断体内基因表达,促进

RNA 降解,可在细胞内发挥基因敲除作用,称为 RNA 干预(即 RNAi)。与高等动物的免疫系统类似,它是植物抵御外在感染(主要是植物病毒及转座子 DNA)的一种天然防御机制。在植物中主要有两种 RNAi 技术系统,即转录后基因沉默(Post Transcriptional Gene Silencing, PTGS)和病毒诱导的基因沉默(Virus Induced Gene Silencing, VIGS)。

(5)异位表达(Ectopic Expression)。通过基因的异位表达(Ectopic Expression)也可以获得基因与表型的联系,如利用可诱导性的启动子,通过外加物质的诱导作用增强转基因的表达。在自然界的生物中普遍存在着一种热休克系统,在高温条件下能够迅速表达。将目的基因上游连接一个热休克蛋白的启动子序列然后转入生物体内,用小的激光束短时间处理任意组织器官甚至是单个细胞,就可以使目的基因在该处表达。

3 我院研究的现状

我院的水稻基因组研究在国内有明显的优势。2001 年启动的院战略行动计划重大项目——水稻基因组测序和重要农艺性状功能基因组研究,大大推动了我院在该领域的研究,也确立了中国科学院在水稻功能基因组研究方面的优势。标志性的成果有中国科学院遗传与发育生物学研究所、国家杂交水稻工程技术研究中心及北京华大基因研究中心三方合作的“中国超级杂交水稻基因组测序和功能开发利用”已于 2002 年 4 月完成了水稻(籼稻)93-11 全基因组的工作框架图;2002 年 11 月,中国科学院国家基因中心联合其它单位在国际上率先完成的水稻(粳稻)日本晴 4 号染色体的精确测序。这两项研究成果分别发表在 *Science* 和 *Nature* 杂志上,而且都在杂志封面上予以突出报道,成为植物基因组研究里程碑性的工作;2003 年 4 月 *Nature* 发表了中国科学院遗传与发育生物学研究所和中国水稻研究所合作克隆和鉴定控制水稻分蘖基因的论文^[9],这是我国在水稻重要农艺性状基因研究方面里程碑性的工作。

“九五”期间,国家“973”计划、“863”计划、国家自然科学基金委员会、中国科学院等支持开展了一定规模的植物功能基因组研究,如水稻突变体的

筛选鉴定、水稻重要性状的功能基因组学等研究,并取得了一批重要成果。通过图位克隆、高通量基因表达分析技术(如 cDNA 芯片等)获得了一些具有潜在应用价值的基因,如与植物脂肪代谢、营养高效利用、育性、抗性、品质等相关的基因。2001 年开始的科技部“十五”计划水稻功能基因组重大专项研究,使我国基因组研究由过去的分散作业,开始进入联合、高通量、大规模和全面系统的新阶段。科技部、中国科学院和农业部等单位分别在上海、北京和武汉等地建立了基因组研究中心,逐步形成了我国基因组研究的重要力量和技术平台,为克隆自主知识产权的重要功能基因奠定了坚实的基础。重要生物全基因组测序和分析能力是一个国家生命科学研究总体实力的具体反映,系统开展水稻籼粳稻亚种间的比较基因组研究^[9],同时,参与国际水稻基因组计划并开展我国籼稻全基因组序列测定及后基因组研究将会大大提高我国在国际生命科学领域的竞争和创新能力,将全面带动转基因技术和分子育种在生产上的应用,为我国水稻、小麦、玉米、大豆等农业品种的更新换代、实现农业可持续发展做出重要贡献。

我院在功能基因组研究方面虽然取得了一定的成绩,但从总体来看形势依然非常严峻,如真正有自主知识产权的技术和成果还很少;水稻及其它植物的栽培和遗传育种分析的高通量还有待于加强;关键的技术,如高效遗传转化载体、有效分析软件、高通量分析平台技术等都要依赖国外实验室提供;技术成果的转化还不够及时等。

4 学科发展对策和建议

如果说以全基因组测序为主要研究内容的结构基因组学研究有明确的时间节点——以完成全基因组精确测序为标志,那么,功能基因组研究将是一项长期和更加艰巨的工作,也更是需要多学科、多种技术不断融合的研究过程。我们应根据该学科的研究特点和现已成熟的技术方法,确定功能基因组研究的几个主要阶段,构建好几个关键的技术平台,合理组织一支高效的学科和技术齐全的研究队伍。功能基因组研究依赖于高通量的实验和数据处理、系统全面的分析,而且其具有可计划性、可

操作性和可预见性等基因组研究的特点。

鉴于上述,我院在发展植物功能基因组研究时,首先要建立从结构基因组学到功能基因组学,直至将功能基因组的信息和知识用于开发改良植物和以植物为基础的新型产品的完整研究开发体系和创新团队,这样才能把瞄准国际前沿领域的研究和国家的重大需求有效地结合在一起,才能保证研究和开发应用的连续性。

其次,研究方向应主攻水稻的功能基因组。近十年来,国际上和我国在水稻基因组测序和结构分析研究上投入了大量人员和经费,已经完成了全基因组测序,现在正是需要充分利用这些数据和条件,系统开展水稻功能基因组研究的最好时期。水稻是研究作物优质高产和抗病抗逆的最佳模式植物,也是研究植物生长发育的模式植物。同时,由于区域种植的原因,水稻的生物学研究也是欧美发达国家投入力量相对较少的研究领域,中国科学家应该抓住有利时机做好水稻功能基因组研究和应用的大文章。

第三,该团队应集合植物分子生物学、基因组学、分子育种学、植物生理学和生物信息学等方面一批科研骨干,研究队伍要加强院内外的密切合作,既注重生命科学不同层次的学科交叉,也要重视与生物信息学、计算数学、化学和物理学的交叉,这些学科的交叉必须有明确的切入点和交叉点,有明确的研究目标。只有通过这样的交叉融合,才能对植物的生长代谢(包括光合作用等生命的基本代谢)和基因的网络关系有深入和突破性的认识,大大促进植物功能基因组的研究。另外,积极争取和承担国家的重大科技攻关项目,并完成好这些任务。

总之,我国的水稻功能基因组研究应努力在激烈的国际竞争中争取到更多的基因资源,并开发利用好这些资源。为达到这一目标,建议有关部门在现有的基础上进一步采取措施加大投入力度,集中支持和全面开展水稻功能基因组研究。相信我国科学家完全有能力在水稻功能基因组研究的整体水平上取得与我国头号水稻种植大国相称的地位。

主要参考文献

1 The Arabidopsis Genome Initiative. Analysis of the genome

sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000, 408, 796-815.

2 Yu J, Hu S, Wang J et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, 2002,296: 79-92.

3 Goff S A, Ricke D, Lan T H et al. Adraft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*, 2002, 296: 92-100.

4 Sasaki T et al. The genome sequence and structure of rice chromosome 1. *Nature*, 2002, 420: 312-316.

5 Feng Q, Zhang, YJ, Hao P et al. Sequence and analysis of rice chromosome 4. *Nature*, 2002, 420, 420: 316-320.

6 The Rice Chromosome 10 Sequencing Consortium. In-depth view of structure, activity, and evolution of rice chromosome 10. *Science*, 2003, 300: 1 566-1 569.

7 Tissier A F, Marillionnet S, Klimyuk V et al. Multiple independent defective Suppressor-mutator transposon insertions in *Arabidopsis*: A tool for functional genomics. *Plant Cell*, 1999, 11: 1 841-1 852.

8 Osato N, Itoh M, Konno H et al. A computer-based method of selecting clones for a full-length cDNA project: simultaneous collection of negligibly redundant and variant cDNAs. *Genome Res*, 2002, 12: 1 127-1 134.

9 Li X, Qian Q, Fu Z et al. Control of tillering in rice. *Nature*, 2003, 422: 618-621.

10 Han B, Xue Y B. Genome-wide intraspecific DNA-sequence variations in rice. *Current Opinion in Plant Biology*, 2003, 6 (2): 134-138.

Higher Plant Functional Genomics

Han Bin

(National Centre for Gene Research, SIBS, CAS, 200233 Shanghai)

Plant functional genomics is a scientific approach that seeks to identify and define the function of plant genes, and uncover when and how genes work together to produce traits. Rapid progress in rice genome sequencing has facilitated research in the rice functional genomics. A widely internal collaboration on rice functional genomics study is needed for accelerating rice gene discovery and application in China. It is no doubted that we will be one of leading countries in this study, if we can closely work together and try our best to make the efforts to rice functional genomics.

Keywords functional genomics, cultivated rice *Oryza sativa*, *Arabidopsis*

韩 斌 上海生命科学研究院国家基因研究中心主任,研究员,博士生导师;上海生命科学研究院植物生理和生态研究所副所长,科技部“十五”“863”计划基因操作主题专家,国际水稻基因组计划核心研究员。1963 年出生。1989—1992 年在英国 John Innes 中心 Sainsbury 实验室学习并获博士学位;1992—1998 年在英国剑桥大学植物科学系做博士后研究;1998 年至今在上海生命科学研究院国家基因研究中心工作,从事水稻基因组研究。