

## 专家论坛

## 拟南芥基因组研究进展\*

黄娟 李家洋\*\*

(中国科学院遗传研究所 北京 100101)

**摘要:** 拟南芥基因组全序列在2000年底已完全测定并公开发表,这是第一个经完全测序的开花植物。序列的获得为进行大规模高等植物基因的鉴定、基因结构与功能的分析、基因表达与调控的研究奠定了坚实物质基础,并将改进和发展一系列进行功能基因组研究的方法与技术。

**关键词:** 拟南芥, 基因组序列, 基因功能研究

**中图分类号:** Q94 **文献标识码:** A **文章编号:** 0253-2654 (2001) 03-0099-03

## ARABIDOPSIS GENOME SEQUENCING AND PLANT FUNCTIONAL GENOMICS

HUANG Juan, LI Jia-Yang

(Institute of Genetics, Chinese Academy of Sciences, Beijing 100101)

**Abstract:** The first flowering plant genome sequence of *Arabidopsis thaliana* was completed at the end of 2000. This will greatly promote the studies on systematic identification of functional genes and their expression profiles and the development of technology for functional genomics.

**Key Words:** *Arabidopsis thaliana*, Genome sequencing, Functional genomics

二十世纪八十年代后期,随着DNA自动测序技术的发展,具有远见卓识的科学家们认为可以通过对人类基因组DNA序列的破译来全面、深入、有效地研究人类这一奇妙而复杂的生物体的奥秘。1987年美国国立卫生研究院(NIH)和美国能源部(DOE)联合提出“人类基因组计划”,随后得到了美国政府的批准,并于1990年正式实施,此后成为包括我国在内的国际合作项目。其主要目标为测定人基因组 $3 \times 10^9$  bp的全部序列,阐明全部基因的位置、结构、功能、表达调控方式,以及重大疾病的致病机理。随着人类基因组计划的开展,许多模式生物的基因组计划也相继展开。自1995年以来,先后完成的基因组全序列测定的生物有:流感嗜血杆菌(*Haemophilus influenzae*)<sup>[1]</sup>、酿酒酵母(*Saccharomyces cerevisiae*)<sup>[2]</sup>、线虫(*Caenorhabditis elegans*)<sup>[3]</sup>、果蝇(*Drosophila melanogaster*)<sup>[4]</sup>、拟南芥(*Arabidopsis thaliana*)<sup>[5]</sup>。2001年2月,人类基因

\* 特约稿件

\*\* 中国科学院遗传研究所所长,中国遗传学会秘书长,研究员,博士生导师

收稿日期:2001-03-23

组序列的工作草图绘制完成并向大众公布<sup>[6]</sup>。作为植物分子遗传学研究的模式显花植物的拟南芥,其基因组全序列的测定与公开是生命科学研究与发展的一个里程碑。本文在简介拟南芥基因组计划的基础上论述在拟南芥序列测定完成后植物功能基因组学研究与技术方法的发展趋势。

## 1 拟南芥的特点及研究概况

传统上,对植物结构和功能的理解主要建立在对大范围的物种(特别是那些与农业有关的物种)方面的研究上。虽然通过这种方法能够获得大量的信息,但由于受到自然资源、人力及资金的限制,进展有限。玉米、番茄、豌豆、水稻、大麦、矮牵牛及金鱼草等几种植物都曾作为遗传研究的模式系统,但对于哪种植物最适于作为所有植物的模式,生物学家们并没有取得统一意见。因此,我们对植物生长和发育的基本方面如开花、根的生长、激素作用及对环境信号的反应等的了解仍处于较低水平。

二十年前,随着分子生物学的兴起,植物生物学家开始寻找一种适于用遗传及分子生物学方法做精细分析的模式植物。拟南芥是一种典型的开花植物,广泛分布于欧洲、亚洲和北美。它作为模式植物在基因组分析方面有很多优势<sup>[7]</sup>:(1)生长周期短。整个生长周期,从发芽、莲座叶的长成,到主花序的形成、第一粒种子的成熟可在6周内完成。(2)体形小,占地少。成熟植株一般15cm~20cm高,莲座叶长度不超过5cm。(3)后代多。每株拟南芥可产生上百个苹果,多达5000粒种子。(4)核基因组小。拟南芥细胞核共包含5对染色体,约120Mb(百万碱基)。这些优点使得拟南芥成为植物科学研究的模式植物。

早年间,遗传学家们通过经典的遗传分析绘制了包含约90个基因座位的拟南芥遗传图谱<sup>[8]</sup>,随着各种分子标记技术的发展与标记数目的增多,遗传图谱逐渐得到完善,物理图谱得以建立。1996年,拟南芥基因组全序列测定这一国际合作项目启动;至2000年底,全序列测定与分析基本完成<sup>[9~13]</sup>。拟南芥基因组的测序区段覆盖了全基因组125Mb中的115.4Mb,经分析共含有25498个基因,其编码的蛋白来自11000个家族<sup>[5]</sup>。

## 2 拟南芥基因组分析

与其它生物相比,开花植物有着自己独特的组织及生理特性。植物基因组序列的得到不仅提供了详细研究植物基因功能特征的基础,而且为理解植物与其它真核生物在遗传学基础上的不同提供了方法。

对拟南芥基因组全序列的分析表明,拟南芥的进化过程中包含了一个全基因组的复制,随后又发生了某些基因的缺失及重复复制,而且叶绿体和线粒体中的一部分基因转移至核基因组中也丰富了核基因组的内容。与线虫及果蝇相比,拟南芥基因组编码的11000个蛋白质家庭中虽然包含了许多新的家族,但也缺少了几种常见的蛋白质家族。这一结果表明在这3种多细胞真核生物中,一系列的普遍蛋白经历了不同的扩增及收缩过程。例如,根据序列分析,拟南芥中13%的基因与转录及信号转导有关<sup>[14]</sup>,其中只有8%~23%的蛋白质可在其它真核生物基因组中找到相关基因,这反映了许多植物转录因子的独特进化过程。

通过与已知功能的基因序列进行比较,可大致确定拟南芥中69%的基因功能。但

在事实上,了解某一类基因的普遍功能并不等于洞察了这一基因在特定有机体中所扮演的特殊角色。例如知道一个基因编码激酶或转录因子对了解这一基因如何控制生命过程没有任何帮助。而对拟南芥基因组研究的最终目的是为了认识所有基因并了解它们的功能<sup>[15]</sup>,因此,随着拟南芥基因组序列的完成,第二阶段的工作就是改进并发展基于序列的更为有效的基因功能的研究方法。

### 3 功能基因组的研究方法

研究基因功能的方法主要分为两大类:通过分析基因差异表达鉴定基因功能的研究方法及通过反向遗传学鉴定基因功能的研究方法。

**3.1 分析基因差异表达鉴定基因功能的研究方法** 基因的时空表达是植物生长、发育、分化、衰老及抗逆等生物学过程的分子基础,基因在不同环境与不同组织中的差异表达可为分析基因的功能提供重要的信息。经典的减法杂交、mRNA 差异显示等方法已被广泛应用,最近发展起来的基因表达连续分析技术、DNA 微阵列等方法也已经开始应用于鉴定基因表达的研究中。

基因表达连续分析技术(SAGE, Serial Analysis of Gene Expression)是系统分析基因表达差异的一种重要的方法<sup>[16]</sup>。它的理论依据是:来自 cDNA 3' 端的一段 9-11bp 的序列能够区分基因组中 95% 的基因,这一段序列被称为 SAGE 的标签。将这些 SAGE 标签通过接头连接起来后,对其进行序列测定,根据这些序列中标签的存在与否及存在频率可以判断此标签所代表的基因在特定组织或条件下是否表达及表达强度。

利用 SAGE 要求基因库中有足够的基因序列或 EST 序列,目前该方法已在人类及酵母中使用<sup>[17,18]</sup>。拟南芥基因组全序列的得到为这种方法在拟南芥中的实施提供了进一步发展的基础。

微阵列技术(Microarray Technology)是分析基因表达差异的又一种重要的实验方法<sup>[19,20]</sup>。理论上,代表某一有机体的所有基因的 DNA 序列都可固定在一个小型的固体支持物上,以此做为杂交底物可定量分析某一特定 mRNA 样品中各个基因的表达情况。使用此项技术可大致确定何种基因在何种程度上对病原体、昆虫、干旱、冷、盐、光周期及其他环境变化作出反应。同样,也能够有效地研究发育的某一过程如萌芽、开花等的分子机理,以及系统地鉴定哪种基因负责对植物激素、生长调节因子、安全剂、除草剂及相关农业化学药品作出反应。

拟南芥基因组测序的完成,使我们可在整个基因组的高度上理解这些差异<sup>[21]</sup>。当基因组被视为一个整体后,这些基因表达的差异就给我们提供了洞察植物对外界信号反应的途径,这也是了解基因组这一均衡系统的第一步。

**3.2 反向遗传学鉴定基因功能的研究方法** 除了对野生型植株进行各种处理,寻找基因在不同条件下的差异表达,以此判断基因的功能外,还可在已知基因序列的基础上,通过各种诱变技术制造定向突变,进而研究基因的功能。(未完待续)