

专论与综述 Reviews

水稻的功能基因组学

王江*

中国科学院上海生命科学研究院植物生理生态研究所, 上海 200032

Functional Genomics in Rice

WANG Jiang*

Institute of Plant Physiology & Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China

提要: 文章介绍近几年来水稻结构基因组学、基因表达的大规模分析平台、全基因组的突变群体库以及基因分离克隆方法上的研究进展。

关键词: 水稻; 基因组; 基因表达; 突变库; 基因克隆

水稻(*Oryza sativa* L.)是与小麦具有同等重要地位, 其种植面积占粮食作物总面积的30%左右, 而稻谷产量则占粮食总产的40%以上。我国是世界水稻的生产和消费大国, 种植面积占世界水稻面积的20%左右, 仅次于印度; 总产量占世界总产的33%左右, 居世界第一位(<http://www.knowledgebank.irri.org>)。由于工业化的发展, 耕地逐渐减少, 人口增加, 气候的变化(Peng等2004)等众多因素, 粮食安全问题日益成为人们关注的课题。

采用培育矮秆抗倒品种技术也称“绿色革命”、以及三系杂交稻育种技术, 我国水稻单产在1960年到2000年间增长了3倍(<http://www.knowledgebank.irri.org>), 但是要满足人口日益增长的需求, 在2025年之前产量还需要再提高50%(Khush 2001a, b)。近年来, 国际上将注意力集中在水稻功能基因组研究上, 希望利用迅猛发展的现代生物技术为上述问题的解决提供契机。

1 水稻结构基因组学

基因组学(genomics)研究主要是指建立高精度的遗传、物理图谱和转录图谱, 探索基因组的结构、基因的结构和功能以及物种的进化。水稻由于相对较小的基因组(约389 Mb)、比较完善的遗传转化系统(尤其是粳稻品种)(Hiei等1994)、高密度的遗传图与物理图资源(Chen等2002; Harushima等1998)、积累丰富的EST/cDNA数据库和良好的分子生物学基础(Kikuchi等2003; Liu等

2007; Wu等2002)、与其它禾本科作物基因组具有广泛的高度共线(Moore等1995)等原因, 它已成为单子叶禾本科作物基因组学研究的模式生物。

国际水稻基因组测序计划(International Rice Genome Sequencing Project, IRGSP)正式启动于1998年, 由日本启动, 随后有中国、美国、英国、法国、韩国、印度、泰国等10多个国家分别加入, 旨在制作粳稻品种‘日本晴’(*Oryza sativa* L. ssp. *Japonica* cv. Nipponbare)全基因组的高质量的精确序列图。我国承担第4号染色体的测序任务。2000年4月美国Monsanto公司率先完成了水稻基因组草图(Barry 2001)。2002年4月北京基因组研究所和美国Syngenta公司两个测序中心分别利用鸟枪法完成籼稻品种‘93-11’和粳稻品种‘日本晴’的全基因组序列草图(Goff等2002; Yu等2002)。2002年11月, 参与IRGSP的日本和中国分别在《Nature》杂志公布了‘日本晴’第1号和第4号染色体的基于物理图的精确测序图谱(Feng等2002; Sasaki等2002)。2005年8月, 国际水稻基因组测序计划宣布完成了‘日本晴’全基因组(约389 Mb)的95% (约370 Mb)高质量精确序列图谱, 还包含了2条染色体上完整的着丝粒序列(IRGSP 2005)。

收稿 2008-07-26 修定 2008-09-22

资助 国家“863”计划(2006AA10A102)。

* E-mail: jwang@sippe.ac.cn; Tel: 021-54924082

当前基于物理图精确测序的图谱研究表明(IRGSP 2005; Ouyang 等 2007; Tanaka 等 2008), 水稻‘日本晴’全基因组已获得372.1 Mb的高质量精确序列, 余下的5% 分布于12条染色体上的38个间隙(gaps)、10个着丝粒和10个端粒处; 水稻全基因组预测有56 278个基因位点, 因为6 498个基因位点编码10 432个转录本, 所以总转录本为66 710; 如果去除15 236个转座因子相关的蛋白编码基因后, 共有41 042个基因位点编码非转座因子相关的蛋白, 平均9.4 kb 含一个基因, 其中约29%的基因成簇出现, 约71%与拟南芥基因(*Arabidopsis*, 28 000~29 000个基因)享有同源性(反过来, 约90%的拟南芥基因与水稻基因享有同源性)。31 439个基因位点已经得到ESTs序列、全长cDNA序列、Tiling芯片检测、大规模平行测序(massively parallel signature sequencing, MPSS)检测的RNA转录水平上的确认, 8 226个基因位点的编码蛋白序列与功能已知的蛋白质序列相同或相似, 另有13 632个基因位点的编码蛋白含有已知的功能域。基因组中2 859个可能是水稻及禾本科作物所独有的, 推测它们与单双子叶分化相关。另外, 0.38%~0.43%核基因组中还有细胞器里的DNA片段。水稻的全基因组注释可以在以下两个网站查询: 日本的“Rice Annotation Project database”数据库(RAP-DB <http://rapdb.dna.affrc.go.jp>)和美国的“TIGR Rice Genome Annotation”数据库(<http://www.tigr.org/tdb/e2k1/osa1>), 其中RAP-DB数据库的数据已整合到NCBI (National Center for Biotechnology Information)、DDBJ (DNA Data Bank of Japan)和EMBL (European Molecular Biology Laboratory)三大数据库中。

栽培稻除了上面提到的亚洲栽培稻(*Oryza sativa*)——分为粳稻(*Japonica*)和籼稻(*Indica*)两亚种以外, 还有非洲栽培稻(*Oryza glaberrima*) (均具AA型染色体组, 单倍体都为12条染色体)。在稻属(*Oryza*)中还另有22个野生稻, 它们分别代表10个不同类型的染色体组(AA、BB、CC、DD、EE、FF、GG、HH、KK、JJ), 变异类型十分丰富, 含有许多未开发的重要农艺性状如抗病性状等, 为水稻遗传育种研究的宝贵资料, 并为阐明水稻起源和演化提供理论基础, 具有重要应用前景和科学价值。在2003年美国起始了一项“Oryza Map Alignment Project (OMAP)”计划旨在利用现有的

IRGSP‘日本晴’基因组序列装配各野生稻基因组全序列, 以便更好地挖掘野生稻资源及了解水稻的起源和演化(Ammiraju 等 2006; Wing 等 2005)。目前该计划已经构建了11个野生稻品种和1个非洲栽培稻品种(两者包含了稻属中全部10个不同类型的染色体组)的具高覆盖率的BAC文库, 并正在大规模地分析这12个BAC文库中各个克隆两末端序列以及克隆的指纹信息, 并根据IRGSP基因组全序列装配各野生稻基因组的BAC物理图谱, 优先装配第1、3和10号3条染色体。

2 水稻基因表达的大规模分析平台

芯片技术主要是有利于科学家在一次实验中检测了成千上万个基因表达水平的变化, 或者更广泛地识别染色体上转录活性区域及甲基化等特殊区域(Rensink 和 Buell 2005)。这个技术是现代生物信息并喷时代的一个重要的检测手段。通过比较两张不同芯片的结果, 就能在全基因组水平上调查各个基因的变化及其程度, 这样有助于我们发现突变体中不同处理之间或者处理前后以及不同发育阶段中那些特异基因发生变化, 从中寻找重要功能基因。

在水稻的商业基因芯片中有两个著名的国际产品, 一个是Affymetrix (昂飞)公司的GeneChip® Rice Genome Array, 另一个是Agilent(安捷伦)公司的Rice Oligo Microarray kit, 两者使用不同生产工艺在芯片上原位合成寡核苷酸探针(probes)。前者在一张芯片上设计了共610 665个合成的长度25 bp的寡核苷酸探针, 分为55 515组, 每组含11个探针, 在11个探针中要求至少7个探针能共同检测一个水稻转录本, 总共能检测大概48 564个粳稻转录本和1 260个籼稻转录本。而后者以前是一张22 k的芯片(NCBI里 Gene Expression Omnibus, 简称GEO, 号码为GPL892), Agilent公司新近开发了一张4×44 k芯片由4个44 k微阵列组成, 含40 901个合成的长度60 bp的寡核苷酸探针, 总共能检测大概44 000个水稻转录本(Shimono 等 2007)。另外, 在我国北京华大基因研究中心生物芯片平台(http://www.genomics.org.cn/bgi_new/platforms/ricechip.htm)提供的对外服务, 为北京基因组研究所设计的长度70 bp的寡核苷酸点样制造的芯片(Ma 等 2005)有2个版本, 第1个版本为BGI-RiceChip-60K单张一套, 第2个版本为BGI-RiceChip-30K两

张成一套,能检测大概41 k的非转座因子相关的水稻基因。美国加州大学戴维斯分校的NSF45K寡核苷酸点样制造的芯片,有43 312个寡核苷酸探针长度分布在50~70 bp之间,能检测44 974个水稻转录本(<http://www.ricearray.org/>)。

上述基因芯片的设计,都是依据已知的水稻表达序列(如EST、全长cDNA等数据库)以及IRGSP预测的水稻基因序列进行的。因此,这些基因芯片在检测未知的水稻表达序列上能力欠缺,由此产生了水稻全基因组或染色体的tiling(覆瓦式)芯片。tiling芯片是针对分析全基因组所有转录活性区域的DNA微阵列,根据各条染色体已知的序列,扩增一个个相互重叠的片段(Jiao等2005)或者每隔十或十几个碱基合成一段寡核苷酸探针,这样一步步覆盖整个水稻基因组或染色体(Li等2007, 2006, 2005)。一般tiling芯片由一个可操作的数目(如几十张芯片)组成,为了提高芯片的杂交质量还要求在设计的寡核苷酸探针中剔除一些高度重复区域内的探针。tiling芯片不但能全基因组验证已有的转录本,而且能发现新的转录活性区域。第一个利用一套34张tiling芯片分析得到的水稻全基因组转录活性图,验证了35 970个(81.9%)已知的或预测的水稻基因转录本,并发现5 464个新的转录活性区域,一些转录活性区域还存在正向和反向的转录本(Li等2006)。

一些发表的基因芯片表达信息沉积在2个公共生物信息网站NCBI的GEO和EBI的ArrayExpress(Edgar和Barrett 2006; Parkinson等2005),NCBI网站显示水稻基因芯片平台已接近100种,芯片杂交信息数据超过600张(包括样品的重复检测),这些数据包括了不同发育时期、不同组织器官、各种突变体、各种环境条件胁迫下的水稻样品,其中Affymetrix芯片183张,另一个是Agilent的22 k芯片164张。虽然在芯片质量控制(microarray quality control, MAQC)基础上各家设计的基因芯片之间检测基因表达的信息有较好的重复性(Canales等2006; Shi等2006),但是目前各基因芯片之间的实验结果还不能很好地进行直接比较,需要一些研究数据以及数据的校正和转换。另外,除了商用的芯片分析软件GeneSpring外,整合和开发各种开放的芯片分析软件仍然十分必要。

除了基因芯片采用杂交手段检测全基因组表

达水平变化外,水稻基因表达的大规模分析平台里还有两个以测序为基础的检测方法:大规模平行测序技术(massively parallel signature sequencing, MPSS)(Nobuta等2007)和基因表达的系统分析(serial analysis of gene expression, SAGE)(Gowda等2004; Matsumura等1999, 2003)。两者检测的理论基础都是一个长度20 bp左右的寡核苷酸序列足以作为特定基因的识别标签,这些识别标签在表达文库中出现的频率能很好地代表特定基因体内的表达丰度(Brenner等2000a, 2000b; Velculescu等1995)。最近一篇文章报道了利用MPSS技术建立的整个基因组的水稻表达图谱(Nobuta等2007),该研究从22个水稻mRNA文库中检测了46 971 553个短序列识别标签,验证和识别了大量已知的和未知的转录本,还从3个小RNA文库中检测了2 953 855个短序列识别标签,是目前最深入地挖掘作物中小RNA的研究工作。20~24 bp的小RNAs,包括micro-RNAs(简称miRNA)和short interfering RNAs(简称siRNAs),两者都能通过序列互补与靶基因的mRNA结合并降解靶mRNA,使靶基因沉默。siRNAs还能通过DNA的甲基化或组蛋白的修饰引起靶基因的转录沉默。小RNA的研究是当今生物领域的热点之一,被公认为广泛地控制植物的各生长发育,以及生理代谢途径包括环境胁迫响应途径。MPSS技术是高度自动化的价格昂贵的检测技术,目前只掌握在少数几家实验室。由于没有芯片杂交的背景噪音,不能采用信号与背景比值的阈值来剔除假阳性,MPSS技术分析中小部分短序列识别标签的真伪难辨,但是与tiling等基因芯片相比,大大提高了对极低表达丰度的转录活性位点的检出。

通过MPSS技术和tiling芯片的研究工作,发现水稻基因组的转录比我们目前想象的要复杂得多,成千上万个已知基因可能存在至今没有识别的各种转录本,这些特殊的转录本的表达可能只在高度特定的时期、组织或处理中才出现;以前认为基因间和重复序列区域是非转录区域,研究发现这些区域的转录活性大大出乎人们意外。这些工作加深人们对基因的认识水平。

信息时代的基因表达的大规模分析平台给我们带来极大的便利,通过上述各种实验方法系统地分析水稻各个组织、各个发育阶段、各个相关突

变体以及各种处理条件下全基因组的基因表达数据库, 然后加以合理地整合分析, 预计能大大加速水稻功能基因的信息阐述。

3 水稻全基因组的突变群体库

突变体是研究水稻功能基因组的基础和前提。目前国际上许多国家都竞相开展水稻随机插

入的突变体库的构建, 希望建立一个全基因组饱和突变体库。插入突变的基因功能研究方法就是将一个序列已知的DNA片段整合到植物基因组中, 而导致被插入植株的功能突变, 而已知序列的DNA片段可作为整合位点的标记, 故也称为基因标签技术(gene tagging)。表1列出了各种水稻插入突变

表1 水稻插入突变群体库

插入的类型	增加插入的方法	插入拷贝数	插入的热点	定位的插入旁邻 序列数目(总共约 177 000)	被插入的基因位 点数目(总共约 28 000)	参考文献
外源转座子 <i>Ac/Ds</i>	农杆菌转化, 田间 自交或杂交	1~7	有	~7 200	~2 500*	I
外源转座子 <i>Spm/dSpm</i>	农杆菌转化, 田间 自交或杂交	1~7	有	8 416	~2 500*	II
内源逆转座 子 <i>Tos17</i>	组织培养	5~10	有	~32 450	~5 000	III
T-DNA	农杆菌转化	1~2	可能无	~143 000 ^a	~21 000 ^a	IV
T-DNA 含有 增强子	农杆菌转化	1~2	可能无	~46 083	-	V
基因过量表 达系统	农杆菌转化	-	-	8 225	5 462	VI

* 表示 *Ac/Ds* 和 *Spm/dSpm* 相加的总数。a 表示数目包含了 T-DNA 含有增强子的这一插入类型的 T-DNA 旁邻序列。I: Chin 等 1999; Eamens 等 2004; Jiang 等 2007; Kohli 等 2001; Kolesnik 等 2004; Park 等 2007; van Enckevort 等 2005。II: Kumar 等 2005。III: Miyao 等 2007, 2003; Piffanelli 等 2007。IV: Hsing 等 2007; Jeon 等 2000, 2002, 2006; Ryu 等 2004。V: Hsing 等 2007; Jeong 等 2002, 2006。VI: Nakamura 等 2007。数据来源于 RiceGE 数据库(<http://signal.salk.edu/cgi-bin/RiceGE>, 2008年5月1日)及上述文献资料。

群体库类型的一些基本情况。

水稻插入突变体的一些实验设计方案, 包括了基因的敲除(gene knockout)、基因的增强(gene activation)、基因的过量表达(gene over-expression)系统、以及基因/启动子捕获(gene/promoter trap)和增强子捕获(enhancer trap), 并使用 T-DNA (transfer DNA)、外源转座子(如 *Ac/Ds* 系统、*En/Spm*)和内源逆转录转座子(如 *Tos17*)作为插入因子。T-DNA 是根癌农杆菌(*Agrobacterium tumefaciens*) Ti (tumor inducing)质粒或发根农杆菌(*Agrobacterium rhizogenes*) Ri (root inducing)质粒上的一段特殊DNA, 它可以从农杆菌转移出并稳定地整合入植物染色体中, 是目前植物中常用的一类介导外源基因的转化载体。当前采用 T-DNA 介导的水稻遗传转化系统(尤其是粳稻品种)已经相当完善了(Hiei 等 1994; Toki 等 2006), 加上 T-DNA 插入基本可随机发生在染色体的所有基因上, 因此是目前获取大量水稻插入突变植株的主要手段。外源转

座子由 T-DNA 介导整合到水稻基因组中后, 可在相应的转座酶作用下再次发生转座, 增加插入。因此, 对水稻全基因组饱和插入突变群体来说, 在一定程度上可减少数量巨大的 T-DNA 初始转基因水稻植株, 但其不足之处是目前发现的再转座现象常距原始插入位点的距离较近, 即具有再转座的热点, 故不易大规模产生新的插入突变植株。内源逆转录转座子(retrotransposon) *Tos17* 在水稻组织培养过程中比较活跃, 可被激活发生再次转座插入事件, 并且在水稻全基因组中存在相对较少的拷贝数, 因此在水稻中得到开发利用(Hirochika 2001; Hirochika 等 1996)。

利用 T-DNA, 外源转座子(如 *Ac/Ds* 系统、*En/Spm*)、内源逆转录转座子(如 *Tos17*)随机插入到水稻基因组, 一些插入影响到插入位点基因的表达造成基因功能丧失(loss-of-function), 即基因敲除, 其中一部分基因被敲除后产生可筛选的突变表型, 由于插入的DNA序列已知, 一旦遗传分析确定突变表

型与插入的标签连锁, 即可通过 TAIL-PCR 或反向 PCR 方法获得插入标签的旁邻序列, 了解突变基因的信息。另一方面, 建立突变库中各个水稻插入 DNA 序列的旁邻序列 (flanking sequence tags, FSTs) 数据库, 然后筛选感兴趣基因的插入突变株再进行基因功能的研究。目前, 水稻中已经产生了大约 177 000 个 FSTs, 这些序列表明大约 28 000 个基因位点 (包括起始密码子 ATG 上游 5' 非翻译区, 但不包括 3' 非翻译区和启动子区域) 里插有已知 DNA 序列, 占全基因组预测总基因位点的 48% 左右 (<http://signal.salk.edu/cgi-bin/RiceGE>, 2008 年 5 月 1 日)。

基因组中存在许多基因家族, 基因成员间在功能上有一定的重叠或互补, 因此单基因的插入突变所造成的基因敲除, 在大多数情况下没有产生明显的表型变化, 给基因功能分析带来难度; 而有些基因的表达明显具有组织器官的特异性, 发育阶段的特异性, 以及环境条件的特异性, 同样单纯的基因敲除方法很难发现这些特征。为了克服这些困难, 前者可以通过基因的增强系统 (gene activation) 和过量表达系统得到一定改善, 后者可以利用基因 / 启动子捕获和增强子捕获得到完善。基因的增强系统和过量表达系统两者都能造成基因的功能获得 (gain-of-function)。基因的增强系统 (Hsing 等 2007; Jeong 等 2002, 2006) 是通过在 T-DNA 一端引入多个强的增强子序列, 使增强子插入位点附近, 几千碱基范围甚至上万碱基范围内的基因表达都增强, 从而解决一些基因冗余的问题。由于增强子作用的范围大, 因此可以适当减少原始的转化插入株; 另一方面, 多个基因同时增强表达所造成的表型改变, 需要逐个基因的功能验证。基因的过量表达系统 (Nakamura 等 2007) 则是使用强的启动子序列 (如玉米的 Ubi 启动子) 引导水稻基因组各个全长 cDNA 的表达。由于这种表达是异位的增强表达, 因此有时会产生一些非基因自身功能的多余表型变化。

捕获系统是以报告基因 (如 GUS、荧光蛋白 GFP 和 RFP 等等) 的表达来挖掘特异表达的插入突变基因, 而不管插入突变体的表型。增强子捕获系统中, 报告基因与最小启动子融合, 典型的最小启动子仅含 TATA box 及转录起始位点, 其本身不能单独驱动报告基因的表达, 经附近增强子激活后, 才能导致报告基因表达。由于增强子捕获不受插入位置的限制, 且在相当远的距离也能起作用, 因此

带增强子捕获系统的 T-DNA 不仅可插入在转录单元内, 而且可位于转录单元外, 两者都能起捕获功能, 而其插入的方向性也不影响捕获。基因 / 启动子捕获系统中, 报告基因前则无启动子成分, 并且位于 T-DNA 或 Ds 的边界附近, 只有插入在一个转录单元内且方向正确时才能表达。为了使插入在内含子中的报告基因不被剪切剔除, 一般在报告基因之前带上一个或多个剪切受体位点 (splicing acceptor sites), 这样就可与染色体上基因的剪切供体位点 (splicing donor sites) 相对应, 剪切后留下的报告基因可与染色体基因融合, 蛋白序列正确融合的报告基因就能指示插入突变基因的表达特征, 而且一些融合适当的报告蛋白还可提供相关突变蛋白质的定位信息。

水稻突变群体库除了 T-DNA、转座子、逆转录转座子 *Tosl7* 的插入突变, 以及人为增强或过量表达基因外, 还包括天然自发突变和人工物理化学诱变。天然自发突变的频率极低, 理化诱变技术虽可快速获得大量突变群体, 但是射线诱变常引起 DNA 大片段的缺失或重排, 且变异位点不易确定。尽管如此该技术仍取得了一定的进展, 日本发现的一个 γ 射线 (gamma ray) 辐射引起的 DNA 大片段的缺失诱变, 正好是一处两个串联的高度同源基因 (*GluB5* 和 *GluB4*) 的位点, 因此射线诱变也许在染色体上串联的功能重叠的基因群分析中有一定优势 (Morita 等 2007)。国际水稻研究所 (IRRI) 利用快中子 (fast neutron) 和 γ 射线辐射诱变, 以及化学试剂双环氧丁烷 (diepoxybutane, DEB) 和甲磺酸乙酯 (ethylmethanesulfonate, EMS), 产生了约 60 000 个 IR64 的突变体, 并获得了 38 000 个 M4 系和大量的表型鉴定 (Wu 等 2005)。

上述提到的双环氧丁烷、甲磺酸乙酯以及叠氮钠 (sodium azide) 和 *N*-甲基亚硝基脲 (*N*-methyl-*N*-nitrosourea, MNU) 等是常用的化学诱变剂。这些化学诱变剂产生的基因突变为点突变, 而且一个基因内常会在不同碱基位点发生点突变, 经大规模筛选能找到一系列等位突变基因的植株。点突变的另一特点是, 突变表型与上述插入突变 (常引起基因功能的丧失) 相比, 常常表现比较温和, 适合于一些特殊基因 (如插入突变能导致早期致死的基因) 功能的深入全面的研究。TILLING (targeting induced local lesions in genomes) 技术 (McCallum 等

2000)的发展,使得大规模自动化筛选特定基因位点内发生化学点突变成为现实。在水稻作物中,美国和日本近年都发展了其水稻TILLING计划(Suzuki等2008;Till等2007)。在TILLING技术中,建立高密度突变的群体库、设计能扩增出基因特定区域的引物以及高灵敏的检测系统都是技术中的关键。另外,化学诱变虽然在被检测的DNA区域内是相当均匀的,但是这种均匀性在整个染色体上的情况仍然未见报道。

水稻突变群体库的创建工作当前正在蓬勃发展,上述提到的各种技术手段都有其优缺点,但随着研究的不断深入,研究手段将得到不断完善。

4 正向遗传学与反向遗传学

传统的遗传学即正向遗传学(forward genetics)分离基因的方法,是首先研究突变体中某一已知突变性状的表现以及其遗传行为(如突变性状在后代中的传递规律、核质基因的判断、控制突变性状的基因数目等等),然后对主效基因进行染色体上的定位及精细定位,接着对定位区域内的候选基因筛选并且验证其功能,最后分离出控制突变性状的基因。这种方法被称为图位克隆法。反向遗传学(reverse genetics)是相对正向遗传学而言,是在功能基因序列发生已知突变后的基础上,对这些突变基因进行分析,研究该基因突变后在生物体内的功能作用即表型变化。在具备了水稻T-DNA、外源转座子以及内源逆转录转座子Tos17插入突变群体和其它全基因组上已知序列突变的资源库后,水稻功能基因组学的研究正逐步由正向遗传学(突变—基因)向反向遗传学(基因—突变)发展。

图位克隆新基因的方法需要建立一个遗传背景差异大的水稻品种间或亚种间的杂交群体进行基因定位,精细定位群体通常要分析500~30 000株F₂代个体的性状及其基因型,需要大片的田地。而对于那些筛选分析工作难度大的性状,大群体分析的操作相当费时费力。水稻染色体上部分区域序列的高度同源性,都给图位方法克隆一些新基因带来一定局限。

在过去长期的传统育种工作中,人们已经挖掘出了大量自然突变发生的具有重要农艺性状的突变体,对这些重要基因的定位工作仍然得到高度重视,尤其在当今水稻全基因组序列的公布对图位克隆方法极大促进的基础下,这一点也可从目前报道已克

隆的水稻重要基因大多还是使用图位克隆方法的事实上得到反映。另一方面,最近已有几篇文章报道正对一些水稻插入突变库进行了大规模的性状观察或筛选(Chern等2007;Jiang等2007;Larmande等2008;Miyao等2007),但是从中筛选出有重要意义的大量突变体的工作仍然正在进行中,此外在水稻T-DNA或Tos17插入突变体库应用中,多个实验室报道只有较低频率观察到突变性状与插入标签存在连锁,部分原因是在构建插入突变群体的组织培养过程中,产生了大量的非T-DNA或Tos17等已知序列插入造成的突变体。这些现象使得先从突变体库筛选特定突变性状,再判断已知插入序列位点的基因与突变性状的连锁关系,最后分离出基因的传统正向遗传学方法困难不小。

反向遗传学研究水稻基因的功能,既可以通过对插入突变库构建插入序列的旁邻序列FSTs数据库,也可以对突变库分组后进行筛选特定区域的序列突变的单株(An等2005)。大规模的突变体的构建一方面对反向遗传学研究带来极大的便利,另一方面这些大量突变体种子的储存、繁殖和分发则是极大的不便。加上对转基因植株种植的政策管理,也给种子的繁殖和分发形成一定的阻碍。另外,水稻中部分功能基因的冗余在一定程度上也给反向遗传学研究方法造成了一些困难,虽然增强或过量表达基因的方法能够弥补一些不足。

除了上述两种基因克隆方法外,其它一些基因克隆方法也正在应用或逐步展开,如从突变体芯片里大规模基因的表达变化中筛选候选基因,或从一系列特定芯片中众多基因的相似表达模式上挖掘功能新基因,从蛋白质蛋白质或蛋白质核酸相互作用的筛选方法中寻找新基因,从传统的代谢途径中各步骤的相关蛋白酶家族里筛选功能基因,从系统进化树中寻找相关功能基因等等。加快水稻基因功能的识别仍然是当前备受关注的主题。

参考文献

- Ammiraju JS, Luo M, Goicoechea JL, Wang W, Kudrna D, Mueller C, Talag J, Kim H, Sisneros NB, Blackmon B et al (2006). The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*. *Genome Res*, 16: 140~147
- An G, Lee S, Kim SH, Kim SR (2005). Molecular genetics using T-DNA in rice. *Plant Cell Physiol*, 46: 14~22

- Barry GF (2001). The use of the Monsanto draft rice genome sequence in research. *Plant Physiol*, 125: 1164~1165
- Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M et al (2000a). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol*, 18: 630~634
- Brenner S, Williams SR, Vermaas EH, Storck T, Moon K, McCollum C, Mao JI, Luo S, Kirchner JJ, Eletr S et al (2000b). In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. *Proc Natl Acad Sci USA*, 97: 1665~1670
- Canales RD, Luo Y, Willey JC, Austermler B, Barbacioru CC, Boysen C, Hunkapiller K, Jensen RV, Knight CR, Lee KY et al (2006). Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat Biotechnol*, 24: 1115~1122
- Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang G, Kim H, Frisch D, Yu Y, Sun S et al (2002). An integrated physical and genetic map of the rice genome. *Plant Cell*, 14: 537~545
- Chern CG, Fan MJ, Yu SM, Hour AL, Lu PC, Lin YC, Wei FJ, Huang SC, Chen S, Lai MH et al (2007). A rice phenomics study—phenotype scoring and seed propagation of a T-DNA insertion-induced rice mutant population. *Plant Mol Biol*, 65: 427~438
- Chin HG, Choe MS, Lee SH, Park SH, Koo JC, Kim NY, Lee JJ, Oh BG, Yi GH, Kim SC et al (1999). Molecular analysis of rice plants harboring an *Ac/Ds* transposable element-mediated gene trapping system. *Plant J*, 19: 615~623
- Eamens AL, Blanchard CL, Dennis ES, Upadhyaya NM (2004). A bidirectional gene trap construct suitable for T-DNA and *Ds*-mediated insertional mutagenesis in rice (*Oryza sativa* L.). *Plant Biotechnol J*, 2: 367~380
- Edgar R, Barrett T (2006). NCBI GEO standards and services for microarray data. *Nat Biotechnol*, 24: 1471~1472
- Feng Q, Zhang Y, Hao P, Wang S, Fu G, Huang Y, Li Y, Zhu J, Liu Y, Hu X et al (2002). Sequence and analysis of rice chromosome 4. *Nature*, 420: 316~320
- Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H et al (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*, 296: 92~100
- Gowda M, Jantasuriyarat C, Dean RA, Wang GL (2004). Robust-LongSAGE (RL-SAGE): a substantially improved LongSAGE method for gene discovery and transcriptome analysis. *Plant Physiol*, 134: 890~897
- Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A et al (1998). A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics*, 148: 479~494
- Hiei Y, Ohta S, Komari T, Kumashiro T (1994). Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J*, 6: 271~282
- Hirochika H (2001). Contribution of the Tos17 retrotransposon to rice functional genomics. *Curr Opin Plant Biol*, 4: 118~122
- Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M (1996). Retrotransposons of rice involved in mutations induced by tissue culture. *Proc Natl Acad Sci USA*, 93: 7783~7788
- Hsing YI, Chern CG, Fan MJ, Lu PC, Chen KT, Lo SF, Sun PK, Ho SL, Lee KW, Wang YC et al (2007). A rice gene activation/knockout mutant resource for high throughput functional genomics. *Plant Mol Biol*, 63: 351~364
- IRGSP (2005). The map-based sequence of the rice genome. *Nature*, 436: 793~800
- Jeon JS, Lee S, Jung KH, Jun SH, Jeong DH, Lee J, Kim C, Jang S, Yang K, Nam J et al (2000). T-DNA insertional mutagenesis for functional genomics in rice. *Plant J*, 22: 561~570
- Jeong DH, An S, Kang HG, Moon S, Han JJ, Park S, Lee HS, An K, An G (2002). T-DNA insertional mutagenesis for activation tagging in rice. *Plant Physiol*, 130: 1636~1644
- Jeong DH, An S, Park S, Kang HG, Park GG, Kim SR, Sim J, Kim YO, Kim MK, Kim J et al (2006). Generation of a flanking sequence-tag database for activation-tagging lines in japonica rice. *Plant J*, 45: 123~132
- Jiang SY, Bachmann D, La H, Ma Z, Venkatesh PN, Ramamoorthy R, Ramachandran S (2007). *Ds* insertion mutagenesis as an efficient tool to produce diverse variations for rice breeding. *Plant Mol Biol*, 65: 385~402
- Jiao Y, Jia P, Wang X, Su N, Yu S, Zhang D, Ma L, Feng Q, Jin Z, Li L et al (2005). A tiling microarray expression analysis of rice chromosome 4 suggests a chromosome-level regulation of transcription. *Plant Cell*, 17: 1641~1657
- Khush GS (2001a). Challenges for meeting the global food and nutrient needs in the new millennium. *Proc Nutr Soc*, 60: 15~26
- Khush GS (2001b). Green revolution: the way forward. *Nat Rev Genet*, 2: 815~822
- Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H et al (2003). Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. *Science*, 301: 376~379
- Kohli A, Xiong J, Greco R, Christou P, Pereira A (2001). Tagged Transcriptome Display (TTD) in indica rice using *Ac* transposition. *Mol Genet Genomics*, 266: 1~11
- Kolesnik T, Szeverenyi I, Bachmann D, Kumar CS, Jiang S, Ramamoorthy R, Cai M, Ma ZG, Sundaresan V, Ramachandran S (2004). Establishing an efficient *Ac/Ds* tagging system in rice: large-scale analysis of *Ds* flanking sequences. *Plant J*, 37: 301~314
- Kumar CS, Wing RA, Sundaresan V (2005). Efficient insertional mutagenesis in rice using the maize *En/Spm* elements. *Plant J*, 44: 879~892
- Larmande P, Gay C, Lorieux M, Perin C, Bouniol M, Droc G, Sallaud C, Perez P, Barnola I, Biderre-Petit C et al (2008). *Oryza Tag Line*, a phenotypic mutant database for the Génoplante rice insertion line library. *Nucleic Acids Res*, 36: D1022~1027
- Li L, Wang X, Sasidharan R, Stolc V, Deng W, He H, Korbel J, Chen X, Tongprasit W, Ronald P et al (2007). Global identifica-

- tion and characterization of transcriptionally active regions in the rice genome. *PLoS ONE*, 2: e294
- Li L, Wang X, Stolz V, Li X, Zhang D, Su N, Tongprasit W, Li S, Cheng Z, Wang J et al (2006). Genome-wide transcription analyses in rice using tiling microarrays. *Nat Genet*, 38: 124~129
- Li L, Wang X, Xia M, Stolz V, Su N, Peng Z, Li S, Wang J, Deng XW (2005). Tiling microarray analysis of rice chromosome 10 to identify the transcriptome and relate its expression to chromosomal architecture. *Genome Biol*, 6: R52
- Liu X, Lu T, Yu S, Li Y, Huang Y, Huang T, Zhang L, Zhu J, Zhao Q, Fan D et al (2007). A collection of 10,096 indica rice full-length cDNAs reveals highly expressed sequence divergence between *Oryza sativa indica* and *japonica* subspecies. *Plant Mol Biol*, 65: 403~415
- Ma L, Chen C, Liu X, Jiao Y, Su N, Li L, Wang X, Cao M, Sun N, Zhang X et al (2005). A microarray analysis of the rice transcriptome and its comparison to *Arabidopsis*. *Genome Res*, 15: 1274~1283
- Matsumura H, Nirasawa S, Terauchi R (1999). Technical advance: transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). *Plant J*, 20: 719~726
- Matsumura H, Reich S, Ito A, Saitoh H, Kamoun S, Winter P, Kahl G, Reuter M, Kruger DH, Terauchi R (2003). Gene expression analysis of plant host-pathogen interactions by SuperSAGE. *Proc Natl Acad Sci USA*, 100: 15718~15723
- McCallum CM, Comai L, Greene EA, Henikoff S (2000). Targeted screening for induced mutations. *Nat Biotechnol*, 18: 455~457
- Miyao A, Iwasaki Y, Kitano H, Itoh J, Maekawa M, Murata K, Yatou O, Nagato Y, Hirochika H (2007). A large-scale collection of phenotypic data describing an insertional mutant population to facilitate functional analysis of rice genes. *Plant Mol Biol*, 63: 625~635
- Miyao A, Tanaka K, Murata K, Sawaki H, Takeda S, Abe K, Shinozuka Y, Onosato K, Hirochika H (2003). Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell*, 15: 1771~1780
- Moore G, Devos KM, Wang Z, Gale MD (1995). Cereal genome evolution. Grasses, line up and form a circle. *Curr Biol*, 5: 737~739
- Morita R, Kusaba M, Iida S, Nishio T, Nishimura M (2007). Knockout of glutelin genes which form a tandem array with a high level of homology in rice by gamma irradiation. *Genes Genet Syst*, 82: 321~327
- Nakamura H, Hakata M, Amano K, Miyao A, Toki N, Kajikawa M, Pang J, Higashi N, Ando S, Toki S et al (2007). A genome-wide gain-of function analysis of rice genes using the FOX-hunting system. *Plant Mol Biol*, 65: 357~371
- Nobuta K, Venu RC, Lu C, Belo A, Vemaraju K, Kulkarni K, Wang W, Pillay M, Green PJ, Wang GL et al (2007). An expression atlas of rice mRNAs and small RNAs. *Nat Biotechnol*, 25: 473~477
- Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L et al (2007). The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res*, 35: D883~887
- Park SH, Jun NS, Kim CM, Oh TY, Huang J, Xuan YH, Park SJ, Je BI, Piao HL, Cha YS et al (2007). Analysis of gene-trap *Ds* rice populations in Korea. *Plant Mol Biol*, 65: 373~384
- Parkinson H, Sarkans U, Shojatalab M, Abeygunawardena N, Contrino S, Coulson R, Farne A, Lara GG, Holloway E, Kapushesky M et al (2005). ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res*, 33: D553~555
- Peng S, Huang J, Sheehy JE, Laza RC, Visperas RM, Zhong X, Centeno GS, Khush GS, Cassman KG (2004). Rice yields decline with higher night temperature from global warming. *Proc Natl Acad Sci USA*, 101: 9971~9975
- Piffanelli P, Droc G, Mieulet D, Lanau N, Bes M, Bourgeois E, Rouviere C, Gavory F, Cruaud C, Ghesquiere A et al (2007). Large-scale characterization of *Tos17* insertion sites in a rice T-DNA mutant library. *Plant Mol Biol*, 65: 587~601
- Rensink WA, Buell CR (2005). Microarray expression profiling resources for plant genomics. *Trends Plant Sci*, 10: 603~609
- Ryu CH, You JH, Kang HG, Hur J, Kim YH, Han MJ, An K, Chung BC, Lee CH, An G (2004). Generation of T-DNA tagging lines with a bidirectional gene trap vector and the establishment of an insertion-site database. *Plant Mol Biol*, 54: 489~502
- Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T, Katayose Y, Wu J, Niimura Y, Cheng Z, Nagamura Y et al (2002). The genome sequence and structure of rice chromosome 1. *Nature*, 420: 312~316
- Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, Collins PJ, de Longueville F, Kawasaki ES, Lee KY et al (2006). The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol*, 24: 1151~1161
- Shimono M, Sugano S, Nakayama A, Jiang CJ, Ono K, Toki S, Takatsuji H (2007). Rice WRKY45 plays a crucial role in benzothiadiazole-inducible blast resistance. *Plant Cell*, 19: 2064~2076
- Suzuki T, Eiguchi M, Kumamaru T, Satoh H, Matsusaka H, Moriguchi K, Nagato Y, Kurata N (2008). MNU-induced mutant pools and high performance TILLING enable finding of any gene mutation in rice. *Mol Genet Genomics*, 279: 213~223
- Tanaka T, Antonio BA, Kikuchi S, Matsumoto T, Nagamura Y, Numa H, Sakai H, Wu J, Itoh T, Sasaki T et al (2008). The Rice Annotation Project Database (RAP-DB): 2008 update. *Nucleic Acids Res*, 36: D1028~1033
- Till BJ, Cooper J, Tai TH, Colowit P, Greene EA, Henikoff S, Comai L (2007). Discovery of chemically induced mutations in rice by TILLING. *BMC Plant Biol*, 7: 19
- Toki S, Hara N, Ono K, Onodera H, Tagiri A, Oka S, Tanaka H (2006). Early infection of scutellum tissue with

- Agrobacterium* allows high-speed transformation of rice. *Plant J*, 47: 969~976
- van Enckevort LJ, Droc G, Piffanelli P, Greco R, Gagneur C, Weber C, Gonzalez VM, Cabot P, Fornara F, Berri S et al (2005). EU-OSTID: a collection of transposon insertional mutants for functional genomics in rice. *Plant Mol Biol*, 59: 99~110
- Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995). Serial analysis of gene expression. *Science*, 270: 484~487
- Wing RA, Ammiraju JS, Luo M, Kim H, Yu Y, Kudrna D, Goicoechea JL, Wang W, Nelson W, Rao K et al (2005). The *Oryza* map alignment project: the golden path to unlocking the genetic potential of wild rice species. *Plant Mol Biol*, 59: 53~62
- Wu J, Maehara T, Shimokawa T, Yamamoto S, Harada C, Takazaki Y, Ono N, Mukai Y, Koike K, Yazaki J et al (2002). A comprehensive rice transcript map containing 6591 expressed sequence tag sites. *Plant Cell*, 14: 525~535
- Wu JL, Wu C, Lei C, Baraoidan M, Bordeos A, Madamba MR, Ramos-Pamplona M, Mauleon R, Portugal A, Ulat VJ et al (2005). Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. *Plant Mol Biol*, 59: 85~97
- Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X et al (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, 296: 79~92