

基因捕捉及其在植物基因分离和功能基因组学上的应用

江树业¹ 陈启锋¹ 葛学军²

(1. 福建农林大学作物遗传育种研究所, 福建 福州 350002; 2. 中国科学院华南植物研究所, 广东 广州 510650)

摘要: 基因捕捉是一种报告基因的随机整合技术。基因捕捉系统已成为分离基因、鉴定基因功能的重要手段。基因捕捉(gene traps)包括增强子捕捉(enhancer trap)、启动子捕捉(promoter trap)和基因捕捉(gene trap), 通称为基因捕捉(gene traps)。在增强子捕捉中, 报告基因与一个基本启动子融合, 这个启动子不能使报告基因表达, 但可被临近的增强子激活。在启动子捕捉和基因捕捉中, 报告基因的启动子被去除, 融合基因只有以正确的方向插入到转录单元内才能表达。对基因捕捉系统的结构特征、构建方法、应用范围、研究现状和应用前景等作了系统论述, 并对有关问题进行了讨论。

关键词: 基因捕捉; 基因分离; 功能基因组学

中图分类号: Q943.2

文献标识码: A

文章编号: 1005-3395 (2003) 02-0181-09

Gene Traps and Their Applications on Plant Gene Isolation and Functional Genomics

JIANG Shu-ye¹ CHEN Qi-feng¹ GE Xue-jun²

(1. Institute of Genetics and Crop Breeding, Fujian Agriculture and Forestry University, Fuzhou 350002, China;

2. South China Institute of Botany, the Chinese Academy of Sciences, Guangzhou 510650, China)

Abstract: Gene trap is a system that allows gene activity to be monitored by creating gene fusions with a reporter gene. Gene traps provide a powerful tool for isolating genes and determining gene functions. Reporter genes can be used to construct three basic types of gene traps: enhancer trap, promoter trap and gene trap. In an enhancer trap, the reporter gene is fused to a minimal promoter that is unable to drive reporter gene expression alone but can be activated by neighboring enhancer elements. Promoter traps and gene traps contain a promoterless report gene, so that reporter gene expression can occur only when the reporter gene inserts within a transcribed chromosomal gene, creating a transcriptional fusion. In this review, we describe how gene traps were developed, how to construct gene trap system, how gene traps were adopted to isolate genes and determine gene functions, what is the present status about applications of gene traps on plant molecular biology and why gene traps are likely to play an increasing important role in the future.

Key words: Gene traps; Isolating genes; Functional genomics

植物基因分离的方法有多种^[1, 2], 这些方法主要依赖于经典遗传学上可识别表型的基因突变或基于基因的表达形态。

依赖于可识别表型的突变基因的分离要求基因突变的表现型应与基因的功能一致。然而, 许多基因有多个功能, 且这些功能别的基因也有, 但它

们的序列不一定相关, 这种突变不易产生易识别的表型, 因为有一个或多个基因家族具有相同的功能。例如, 基因敲除(knockout)分析表明, 尽管曾在广泛生长条件下进行测试, 仍有相当比例的酵母基因被敲除时没有明显的表型效应^[3]。另外, 许多基因的功能在多个发育阶段表现, 这些基因的突变可能导致早期致死或可能一因多效而与特定代谢路径上的某一基因相混淆而难以分离。

分离基于基因的表达形态的基因,已发展了若干新技术,如 mRNA 差别显示技术(mRNA differential display)^[6]、微阵列(microarray)^[5]和基因表达的系列分析(Serial Analysis of Gene Expression, SAGE)^[6]技术等。这些方法可用来分离某一特定空间、时间或特定条件下表达的基因。然而,这些方法最终还是要受到用于分离 RNA 探针的组织的来源所限制,瞬间表达的、低丰度的和只在少数细胞中表达的基因可能不能分离。

在植物基因组学研究方面,拟南芥的基因组全序列的测定工作已于 2000 年完成^[7],并在 GenBank 数据库中登记注册。在水稻方面,1997 以来,以日本为主的国际水稻基因组协作组已发表了长度为 174.4 Mb 的 BAC/PAC 基因组序列[http://www.tigr.org/tdb/e3k1/osa1/BACmapping/description.shtml];美国 Monsanto 公司于 2000 年 4 月 6 日宣布已完成粳稻全基因组工作框架图[http://www.rice-research.org],但未给公共数据库提供序列;国际 Syngenta 公司也于 2001 年 1 月 26 日宣布已完成粳稻全基因组工作框架图[http://www.syngenta.com],草图数据可在其网站有条件登记使用。我国学者于 2001 年独立完成了籼稻全基因组工作框架图^[8],其草图数据已在 GenBank 中登记注册,并于 2002 年与国际 Syngenta 公司在《科学》上发表了有关水稻全基因组工作框架图的构建工作^[9,10]。还有报道指出,45 000 个水稻 EST 序列已测序并于 2001 年释放到公用互联网数据库,但只有其中的 25%可找到已知基因的同源序列,大部分基因的功能还有待鉴定^[11];到 2002 年 6 月 28 日止,释放到 NCBI 数据库的水稻 EST 序列已达 104 973 (http://www.ncbi.nlm.nih.gov/dbEST_summary.html)。

可见,在植物基因的分离克隆方面,现有的方法仍然存在缺陷;同时随着拟南芥、水稻等植物基因组全序列的测定、大量 EST 序列的测定释放,如何鉴定未知基因的功能已成为功能基因组学研究的重点课题。近几年发展起来的基因捕捉系统(Gene trap),可望成为植物基因分离和功能基因组学研究的强有力工具。

1 基因捕捉系统的结构体系

基因捕捉,又称基因陷阱,是一种报告基因的随机整合技术。20 多年前,在细菌遗传学中就曾采用报告基因与目的基因的融合来检测目的基因的活性^[12]。根据融合基因在不同时间的表达形态或在

各种条件下突变表型缺失或任何序列信息,就可分离和鉴定相应的基因。因为被插入的 DNA 的序列已知,插入序列可用做“标签”很容易地分离染色体基因。

植物中建立第一代基因捕捉系统是为了判断 T-DNA 插入整合到基因组中的频率。Andre 等和 Teeri 等用无启动子的抗生素抗性基因与 T-DNA 的一端相连后转化烟草原生质体,来检测是否发生基因融合和融合基因的表达状况^[13,14],从而筛选出已发生基因融合使抗生素抗性基因得以表达的转化体。后来,Koncz 等和 Herman 等在 T-DNA 中引进第二个选择标记 β -葡萄糖苷酸酶报告基因(*gusA* 或 *uidA*),以便转化植株再生后可对无启动子的抗生素抗性基因的表达状况作进一步筛选^[15,16]。这就是植物中最早的基因捕捉系统。目前,所采用的报告基因仍然以 *gusA* 基因为主,也有采用 *gfp* 基因和 *lc* 基因等。细菌的 *gusA* (*uidA*) 基因是目前最常用的植物报告基因。GUS 蛋白的表达相当稳定^[17],其活性可用商业化的各种底物进行组织化学染色来检测。GUS 活性的检测相当敏感^[17,18],甚至能检测单细胞中的 GUS 活性。但大部分的底物相当昂贵,且组织化学染色是破坏性的,GUS 分析不能在活体组织中进行。水母的绿色荧光蛋白(GFP)基因也已被用于植物的报告基因。因为 GFP 可通过荧光来检测,因此检测只要有适当光源即可,相对廉价;GFP 的检测是非破坏性的,可在活细胞中检测,且可多次检测。*lc* 是类 -Myc 转录因子的玉米 *r* 基因家族的一个成员,其功能是调控花青素的合成,是一个很有前景的报告基因;*lc* 在其它植物中的表达会导致花青素的积累^[19,20]。这个报告基因的优点是检测不需要昂贵的底物,且是非破坏性的。

根据所构建的报告基因的结构,目前已发展了 3 种基因捕捉的基本类型:增强子捕捉、启动子捕捉和基因捕捉^[21](图 1)。通常的术语“基因捕捉”泛指各个捕捉系统,而不管特异的报告基因的结构如何。

在增强子捕捉中(图 1A),报告基因与一个基本启动子融合,这个基本启动子通常包含一个 TATA 盒和转录起始位点,不能使报告基因表达,但可被临近的增强子激活。在启动子捕捉和基因捕捉中,报告基因的启动子被去除,融合基因只有以正确的方向插入到转录单元内才能表达(图 1B 和 1C)。启动子捕捉的报告基因只有插入到一个外显子,导致转录融合时才能表达(图 1B)。相对地,基

因捕捉结构中, 报告基因前包含一到多个剪接序列, 当其插入到内含子时也可表达(图 1C)。从染色体基因的剪接供体位点到报告基因的剪接受体位点的剪接导致上游外显子与报告基因的融合。从它们的结构差异可以看出, 增强子捕捉只有以正确的方向插入到一个基因内的报告基因才能表达, 优

点是其插入可导致报告基因的高频表达, 而启动子捕捉和基因捕捉则更易导致基因断裂。由于增强子捕捉能使与报告基因相距较远的基因得到表达, 在增强子捕捉中控制报告基因表达的基因比启动子捕捉或基因捕捉中的更容易被检测到。

可见, 判断和鉴定未知基因的功能只是依据报

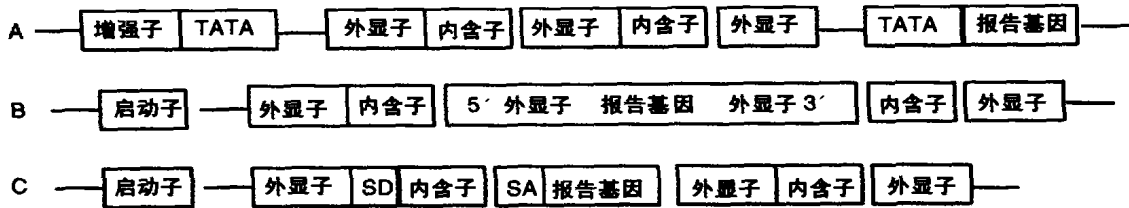


图 1 增强子(A)、启动子(B)和基因捕捉(C)的结构示意图

Fig. 1 Structure of enhancer (A), promotor (B) and gene trap (C) elements

SD 为染色体基因的剪接供体位点 The chromosomal splice donor site;

SA 为报告基因的剪接受体位点 Splice acceptor sequences from the report gene.

告基因的表达状况, 可不依赖突变表型, 并可分离鉴定杂合子中的致死基因; 报告基因检测的敏感性使瞬间表达的、低丰度的和只在少数细胞中表达的基因也可分离, 为植物基因分离和功能基因组学的研究注入了新的活力。

除了以上 3 种捕捉系统外, 人们还构建了“信号-外显子捕捉”(Signal-exon trap, SET)系统^[22]。这一系统可在 DNA 水平上侦察信号分子, 通过筛选被捕捉的外显子编码的多肽, 鉴定信号多肽的功能。

2 基因捕捉突变库的构建

为了使用基因捕捉, 必须构建包含报告基因随机插入整个基因组的个体突变库。主要采用两种方法: T-DNA 和转座子。

T-DNA 介导的转化是产生转基因植株的常用方法。在转化效率较高的植物中, T-DNA 方法是快速构建插入序列突变库的主要方法。优点是基因组 T-DNA 的插入通常是稳定的, 且 T-DNA 的插入不具有位点特异性, 可以建立饱和的 T-DNA 插入突变库^[23]。如在拟南芥中已获得了大量的 T-DNA 插入突变库^[24-28]。然而, T-DNA 方法只适用于这些易被 T-DNA 转化的植物。且在转化植株中, 通常可以检测到多个 T-DNA 插入到一个转化植株中, 一个位点包含多个 T-DNA 插入或多个位点包含多个插入^[18, 29]; 或产生 T-DNA 的正向和反向重复、与比邻的染色体 DNA 发生重排^[30-32]等, 导致随后分子分析的复杂性, 增加了驱动报告基因表达的染色体基因

的分离难度。

转座因子也已广泛用于插入突变库的构建。其优点是, 在转座酶不具活性的情况下, 整合的转座因子是稳定的; 当转座酶具有活性时, 可产生回复突变, 从而验证突变性状是否由转座子插入引起的。但一些转座子(如玉米的 *Ac/Ds*) 具有偏好向邻近连锁位置转座的特性^[33], 大大增加了构建饱和突变库的工作量。目前一些改进措施可有效增加向非邻近位置转座的频率^[34]。

到目前为止, 只有 *Ac/Ds* 系统可用于增强子或基因捕捉, 因为 *Ac/Ds* 具有低的拷贝数^[35], 高拷贝数的转座子, 如 *En/Spm*^[36] 往往会导致报告基因表达的复杂化。

Ac/Ds 转座系统由自主元件 *Ac* 和非自主元件 *Ds* 组成。 *Ac* 元素编码的转座酶与 *Ac* 和 *Ds* 末端倒位重复序列结合, 催化向基因组新位置转座。 *Ds* 是 *Ac* 失去转座酶活性而保留末端倒位重复序列的衍生物。 *Ac* 转座酶基因产生的转座酶能够识别 *Ds* 的末端序列, 从而催化向染色体的新位置转座。这种二元系统可生产稳定的转座插入, 因为自主元件可从转座插入中分离。此外, 可通过有性杂交将转座酶导入转座植株从而诱导再次转座, 引起回复突变, 验证所分离基因的功能。

这里介绍一个转座子基因捕捉系统的构建方法^[34] (<http://genetrapp.cshl.org>)。这个系统采用 *Ac/Ds* 转座元素和正向 / 反向选择标记基因进行转座的选择(图 2)。在这个系统中, *Ac* 元素由于去掉了具有转

座能力的两臂,只能合成转座酶,由花椰菜花叶病毒 CaMv 35S 启动子驱动,将通用表达的 2' 启动子驱动的反向选择标记基因 *iaah* 也被构建到 T-DNA 上。*Ds* 元素由玉米 *Ac* 元素发展而来,含有由通用表达的 1' 启动子驱动的 *NPTII* 基因和 *GUS* 报告基因。在增强子捕捉元素 (DsE) 中,*GUS* 报告基因与来自 CaMv 35S 的弱启动子融合,这个启动子区域检测不到活性,除非其附近具有染色体的增强子元素^[37]。在基因捕捉元素 (DsG) 中,*GUS* 报告基因缺乏启动子,但在三个阅读框上均包含一个 SA 位点,与上游的起始密码子相融合。这种结构使 *GUS* 报告基因在 DsG 元素插入内含子时通过转录和翻译融合而得以表达。另外,位于 *Ds* 元素 3' 末端的供体位点通过自然剪接^[38,39],使 *GUS* 报告基因在 DsG 元素插入外显子时得以表达。在每一个 T-DNA 中,均含有由通用表达的 2' 启动子驱动的反向选择标记基因 *iaah*,使转座后 *Ac* 和 *Ds* 元素均能反向选择。它们均被亚克隆到二元 T-DNA 载体上,通过农杆菌介导法导入拟南芥基因组。

Ac 同源亲本与 *Ds* 同源亲本杂交后,*Ds* 在 *Ac* 转座酶的作用下开始转座。 F_1 植株自交后,收获 F_2 种子。 F_2 种子在包含有萘乙酰胺 (Naphthalene acetamide, NAM) 和卡那霉素 (Kanamycin, Kan) 的培养基上发芽。对 NAM 和卡那霉素均表现出抗性的植株表明其基因组的某个位置上插入了发生转

座的 *Ds* 元素,且所在位点与 *Ac* 元素和供体 T-DNA 均表现出独立分离。这一选择富集了不与原始 *Ds* 位点连锁的转座事件。双抗 F_2 植株再转移到土壤中生长,自交结实并收获 F_3 种子。 F_3 植株再进行 *GUS* 表达分析。

为了鉴定基因捕捉系统中控制报告基因表达的目标基因,首先可采用 PCR 技术扩增位于 *Ds* 插入序列两端的基因组序列^[40]。热敏式不对称交错 PCR (Thermal asymmetric interlaced PCR) 在扩增这些序列时特别有效^[41,42]。一旦两端序列被测定一小段,就可以进行基因库数据检索,确定其在基因组中的位置及其候选基因。因为拟南芥基因组序列已测定完成,即使一小段序列就有相当高的比率确定其相应的基因组序列,从而确定相应的 *Ds* 插入的图谱位置。因为基因组序列已被系统描述,使得插入区内的基因可以得到初步确定,从而初步确定驱动报告基因表达的候选基因。如果基因捕捉插入导致转录融合,就可以用 5'-RACE-PCR 技术 (随机扩增 cDNA 末端技术) 扩增插入上游的外显子序列^[43,44]。这种方法对于鉴定相对高丰度表达的基因相当有效。因为 RACE-PCR 技术扩增的是外显子序列,可以直接用作探针筛选 cDNA 文库,从而获得目标基因。

目前,美国冷泉港实验室已建立了一个有关拟南芥基因捕捉的专门网站,网址是:<http://genetrapp>。

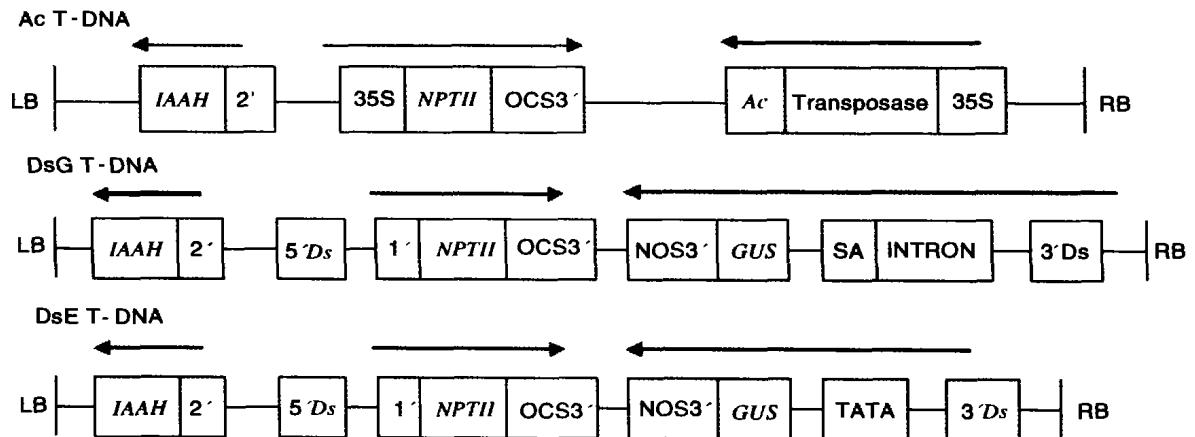


图 2 用于转化 *Ac* 转座酶以及基因捕捉或增强子捕捉 *Ds* 元件的 T-DNA 构建体

Fig. 2 T-DNA constructs used to introduce *Ac* transposase and gene trap or enhancer trap *Ds* elements
IAAH: 吲哚乙酰胺水解酶基因 (对萘乙酰胺 NAM 敏感); *NPTII*: 潮霉素磷酸转移酶基因 (对卡那霉素敏感); *GUS*: β -葡萄糖苷酸酶报告基因; SA: 剪接受体; LB, RB: T-DNA 的左右臂; 1', 2': 启动子; OCS3', NOS3': 转录终止子; INTRON: 内含子; DsG: 基因捕捉转座子; DsE: 增强子捕捉转座子。
IAAH: indole acetic acid hydrolase (sensitive to naphthalene acetamide, NAM); *NPTII*: neomycin phosphotransferase (sensitive to Kanamycin, Kan); *GUS*: β -glucuronidase; SA: splice acceptor; LB and RB: left and right arms of T-DNA; 1', 2': promoters; OCS3', NOS3': transcriptional terminator; INTRON: non-coding region between exons; DsG: gene trap *Ds* element; DsE: enhancer trap *Ds* element

cshl.org。这个实验室已获得 15 000 个拟南芥基因捕捉突变体,到 2002 年底预计可达 40 000 个,其中约 6 000 突变系已进行 *GUS* 分析,并获得相应的 Flanking 序列。包含这些测序信息的数据库也已建成,并在不断扩充。这种数据库将最终包含这些序列的表达资料及其表现型信息。有了这些序列信息就可进行反向遗传学筛选,确定先前克隆基因及基因家族的功能。所有这些信息将对全面弄清拟南芥基因组全部基因的功能起到极其重要的作用。

此外,新加坡国立大学分子农业生物学研究院也正在进行拟南芥和水稻基因捕捉突变库的构建工作,具体情况可参见网站(<http://www.til.org.sg/research/plant/sri/>)及相关文献^[45-49]。

3 基因捕捉系统在植物基因分离和功能基因组学上的应用

基因捕捉系统应用于植物基因分离和功能基因组学研究的一个重要特点是,不但可以分离基因,还能研究其功能。

基因捕捉系统可有效分离鉴定可差异调控的具特异表达形态的基因。具有细胞、组织、特异性表达、暂时性表达或有条件表达的基因,通过报告基因的表达就可得以分离。如已分离出在侧根和正在发育的胚胎中特异表达的基因^[50-52],以及在一定条件下表达的基因,如被线虫类感染而表达的基因^[53-54]等。一个典型例子是基因捕捉系统在有关衰老基因的分离鉴定上的应用:总共筛选了 1 300 个拟南芥增强子捕捉系,发现其中 147 个系的 *GUS* 活性只在衰老叶片中表达;分析了 6 种可能与衰老有关的因子:乙烯、茉莉酮酸、脱落酸、brassinosteroids、脱水和黑暗对 *GUS* 表达调控的影响,从而建立了调节植物衰老的网络路径;对其中的 Sel139 系作了进一步的分析鉴定,分离出一个与衰老有关的基因 *SAG101*,它只在衰老叶片中表达,其编码的蛋白具有酰基水解酶活性,反义 RNA 的转基因植株分析和化学诱导分析表明,此基因在叶片衰老中具有重要作用;还分离了一个 *OPR1* 基因,它与叶片衰老有关,同时受茉莉酮酸的调控^[55-58]。

在拟南芥中具有许多胚胎致死突变,表现出多种类型,使得胚胎致死原因难以判断^[59]。基因捕捉系统的应用可以克服这一问题,因为在分离基因之前,根据 *GUS* 表达形态就可预测其功能。如在基因

捕捉分析中,*DsG* 插入分化细胞中表现出 *GUS* 活性,据此分离到 *PROLIFERA (PRL)* 基因。*PRL* 基因是真核生物中 MCM 基因家族的一个成员,为启动 DNA 复制所必须;*DsG* 的插入打断了 *PRL* 基因,从而导致大量配子体和胚胎致死^[44,60]。

AGL8 基因是通过鉴定增强子捕捉中 *DsE* 元素的插入突变而分离的。*DsE* 元素在 *AGL8* 基因 5' 非转录区的插入导致受精后长角果不能伸长,使长角果变短且挤满了种子,因此 *AGL8* 基因被重新命名为全果 *FRUITFUL (FUL)* 基因。*FUL* 突变引起的长角果伸长受阻与此基因在心皮中的突变表型相一致;然而,*FUL* 基因也在茎尖分生组织中表达,诱导茎尖分化至开花,被认为在营养生长转向生殖生长中起作用;当 *MADS* 盒基因 *APETALA1 (API)* 和 *CAULIFLOWER (CAL)* 发生突变时,*FUL* 突变会导致无法开花^[61,62]。这样,*FUL* 基因表现出多效性,与 *API* 和 *CAL* 一起共同控制是否开花。

越来越多的文献报道表明,基因捕捉系统在分离基因、鉴定基因功能上的应用潜力是巨大的。*Kumaran* 等通过筛选拟南芥基因捕捉系而分离了一个与拟南芥花发育有关的新基因 *afo*^[45]。*Dubreucq* 等通过筛选 10 000 基因捕捉转基因系,获得一个与种子发芽有关的基因 *AtEPR1*,此基因在种子发芽过程中的表达明显受 *GA* 的正向调控,但不受 *ABA* 的负向调节^[63]。*Swaminathan* 等在筛选拟南芥增强子捕捉系中分离到一个与细胞周期有关的 D 类 Cyclin 基因 *CYCD3,2*,*GUS* 分析和 Northern 杂交分析表明,突变体中这个基因的功能已丧失,但没有明显的突变表型,互补实验进一步验证了此基因的功能^[64]。*Hiwatashi* 等构建了包含 235 个基因捕捉系和 1 037 个增强子捕捉系的苔藓 (*Physcomitrella patens*) 突变库,并分离、鉴定出一个酸性 α -葡萄糖苷酸酶基因 *PpGL*^[65]。*Weijers* 等在拟南芥的启动子捕捉系中首次分离和鉴定出一个不完全显性基因 *RPS5*,此基因在杂合状态下细胞分裂延迟或被破坏,但在纯合状态下会降低配子体的生存能力和胚胎致死^[66]。*Yoshioka* 等从拟南芥基因捕捉突变库中分离鉴定出一个编码单功能天门冬氨酸激酶的基因 *AK-lys3*,在叶的木质部、下胚轴和根维管组织中部等高水平表达^[67]。*Hall* 等利用拟南芥增强子基因捕捉系分离和鉴定了一个富含羟脯氨酸的糖蛋白基因 *RSH*,它在决定胚胎发育中细胞板的正确位置上是必要的,此基因功能的缺失导致根、苗和胚轴

的严重畸型,形成玻璃质幼苗,最高可长至 1-3 mm,存活期不到 3 周^[69]。Lechner 等利用基因捕捉系 SGT6304 分离和鉴定了拟南芥中 finger 蛋白基因家族中的一个 *RHA2b* 基因,尽管此基因已被打断,也没有导致明显的生长和发育缺陷,但通过 *GUS* 分析也能分离相应的基因^[69]。Tanaka 等从拟南芥启动子捕捉系中筛选鉴定出与大豆基因 *GH3* 同源的基因 *AtGH3a*,受光敏色素 B 调控,通过改变激素水平来调节其表达^[70]。

除了用于分离基因外,基因捕捉也可用于分离驱动特异表达的启动子。分离的特异表达的启动子又可用于驱动基因的异位表达而检测此基因的表达形态。如 Tsugeki 等利用增强子捕捉系统分离了一个根冠特异表达启动子,并用于启动 *DT-A* (diphtheria toxin) 基因来研究拟南芥中根冠细胞切除后的发育状况^[71]。

基因捕捉的另一个潜在应用是分离调节基因下游的目标基因。这个方法在果蝇中应用得相当成功,如已用来分离受 *Antennapedia* 调控的基因^[72]。其基本原理是,经筛选获得的候选基因捕捉系中,报告基因的表达受感兴趣的调节基因的调控。这样的基因捕捉系与失去调节基因功能的突变株或调控基因得以异位表达的转基因植株杂交,杂交后代中报告基因的表达就会发生改变,这种改变是受调节基因控制的。

4 值得探讨的问题

利用基因捕捉系统分离基因,并研究其功能的前提是报告基因必须能真正反映染色体基因的表达状况。然而实际情况是不是这样呢?在果蝇和老鼠的研究中,大部分增强子和基因捕捉的报告基因是能够真正反映内源基因的表达的^[73]。在植物中,初步结果也表明,报告基因的表达的确能够反映植物基因的表达^[60,61,74-76]。事实上,在基因捕捉系统中,报告基因比通常用于表达分析的启动子-报告基因更能反映其表达,因为基因捕捉元素插入到植物基因组后,基因组可提供所有的调控元素(基因捕捉元素插入到调控元素位置的情况例外);且这些元件是以自然状态保存于植物染色体上,不存在可能导致转基因植株不能正确表达的位置效应问题。而启动子-报告基因不可能包含从内含子或 3' 端到转录区的调控序列,能否真正反映报告基因的表达已引起人们的质疑,并认为单纯用启动子-报告基因

模式进行表达分析是不够的^[77]。

然而,也有报道报告基因与其染色体基因的表达不一致的例子,这可能与基因捕捉元素刚好插入到调控元素位置上有关。例如,一个增强子捕捉插入导致根冠特异表达形态,然而,克隆出的相应基因具有不同的表达形态,可能是插入打断了 *GUS* 报告基因表达所需的启动子^[71]。又如, T-DNA 插入烟草后,导致 *GUS* 报告基因在种子外壳特异表达,但在插入区却找不到可转录表达的基因,这是因为 T-DNA 的插入激活了烟草中本不具活性的启动子^[78]。然而,也可能这个特殊的启动子驱动了非编码 RNA 或小多肽的表达,但无法被所采用的检测方法检测到。

另外,一个值得提及的问题是,个体突变库的群体到底需要多大才可以形成一个饱和的突变库,即一个基因至少可以找到一个相应的突变体。普遍认为这个问题与所研究的植物种的基因总数及其在基因组内的分布情况有关。就水稻而言,基因组大小约 430 Mb,若要构建以平均每 15 kb 范围内就可找到一个插入的突变库,则需约 30 000 个插入突变体。在禾谷类植物中,基因在基因组中的分布是不均匀的,存在基因的集中区^[79]。初步研究表明,水稻中的基因集中区大约占水稻基因组的 25%,而大部分的 T-DNA 具有插入到水稻基因组基因集中区的偏好^[80]。这样,突变库的群体似乎可以比用基因组大小和基因密度预测的小。但通过拟南芥突变库的构建认为,所构建的突变库的群体还不够大,且群体越大,发现被打断基因的个体的机会就越高^[81]。

另一个值得注意的问题是,只有单拷贝插入的基因捕捉突变体为分析相应基因的功能提供了方便,但由于多因一效或基因家族的存在,以及高等植物对各种生理压力、遗传变异的高适应性,单个基因的打断并不会导致其相应功能的丧失或功能太弱而无法检测,因而增加了对功能基因鉴定的难度,但在分析多因一效或基因家族上仍然是有效的。此外,由于基因间的相互作用,对单个基因打断的分析有时可能会扭曲对相应基因功能的解释。可见,尽管基因捕捉系统具有许多优点,但对许多基因功能的分析还应结合其它的分析方法。

总的来说,基因捕捉系统已成为分离基因、鉴定基因功能的有效工具;随着拟南芥、水稻基因组全序列的测定,在应用于鉴定基因组功能方面将起到越来越重要的作用。

参考文献

- [1] Gibson S, Somerville C. Isolating plant genes [J]. Trends Biotechnol, 1993, 11:306-313.
- [2] Jiang S Y (江树业), Chen Q F (陈启锋), Fang X J (方宜钧), et al. Reviews and comments on isolating plant genes [J]. J Fujian Agri For Univ (福建农林大学学报), 2000, 29:261-268. (in Chinese)
- [3] Ross-MacDonald P, Coelho P S R, Roemer T, et al. Large-scale analysis of the yeast genome by transposon tagging and gene disruption [J]. Nature, 1999, 402:413-418.
- [4] Liang P, Pardee A B. Differential display of eukaryotic message RNA by means of the polymerase chain-reaction [J]. Science, 1992, 257:967-971.
- [5] Schena M, Shalon D, Davis R W, et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray [J]. Science, 1995, 270:467-470.
- [6] Velculescu V E, Zhang L, Vogelstein B, et al. Serial analysis of gene expression [J]. Science, 1995, 270:484-487.
- [7] The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana* [J]. Nature, 2000, 408:796-815.
- [8] Yu J (于军), Hu S N (胡松年), Wang J (王俊), et al. A draft sequence of the rice (*Oryza sativa* L. ssp. *indica*) genome [J]. Chin Sci Bull (科学通报), 2001, 46:1937-1942. (in Chinese)
- [9] Yu J, Hu S N, Wang J, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*) [J]. Science, 2002, 296:79-92
- [10] Goff S A, Picke D, Lan T H, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*) [J]. Science, 2002, 296: 92-100.
- [11] Jeon J S, An G. Gene tagging in rice: a high throughput system for functional genomics [J]. Plant Sci, 2001, 161:211-219.
- [12] Casadaban M J, Cohen S N. Lactose genes fused to exogenous promoters in one step using a Mu-*lac* bacteriophage: *In vivo* probe for transcriptional control sequences [J]. Proc Natl Acad Sci USA, 1979, 76:4530-4533.
- [13] André D, Colau D, Schell J, et al. Gene tagging in plants by a T-DNA insertion mutagen that generates APH(39)II-plant gene fusions [J]. Mol Gen Genet, 1986, 204:512-518.
- [14] Teeri T H, Herrera-Estrella L, Depicker A, et al. Identification of plant promoters *in situ* by T-DNA-mediated transcriptional fusions to the *npt-II* gene [J]. EMBO J, 1986, 5:1755-1760.
- [15] Koncz C, Martini N, Mayerhofer R, et al. High-frequency T-DNA-mediated gene tagging in plants [J]. Proc Natl Acad Sci USA, 1989, 86:8467-8471.
- [16] Herman L, Jacobs A, Van Montagu M, et al. Plant chromosome/ marker gene fusion assay for study of normal and truncated T-DNA integration events [J]. Mol Gen Genet, 1990, 224:248-256.
- [17] Jefferson R A, Kavanagh T A, Bevan M W. GUS fusions: β -Glucuronidase as a sensitive and versatile gene fusion marker in higher plants [J]. EMBO J, 1987, 6:3901-3907.
- [18] Lindsey K, Wei W, Clarke M C, et al. Tagging genomic sequences that direct transgene expression by activation of a promoter trap in plants [J]. Transgenic Res, 1993, 2:33-47.
- [19] Lloyd A M, Walbot V, Davis R W. *Arabidopsis* and *Nicotiana* anthocyanin production activated by maize regulators *R* and *C1* [J]. Science, 1992, 258:1773-1775.
- [20] Goldsbrough A P, Tong Y, Yoder J I. *Lc* as a nondestructive visual reporter and transposition excision marker gene for tomato [J]. Plant J, 1996, 9:927-933.
- [21] Springer P S. Gene traps: tools for plant development and genomics [J]. Plant Cell, 2000, 12:1007-1020.
- [22] Peterfy M, Gyuris T, Takacs L. Signal-exon trap: a novel method for the identification of signal sequences from genomics DNA [J]. Nucl Acids Res, 2000, 28:E26
- [23] Azpiroz-Leehan R, Feldmann K A. T-DNA insertion mutagenesis in *Arabidopsis*: going back and forth [J]. Trends Genet, 1997, 13: 152-156.
- [24] Feldmann K A, Marks M D. *Agrobacterium*-mediated transformation of germinating seeds of *Arabidopsis thaliana*: a non-tissue culture approach [J]. Mol Gen Genet, 1987, 208:1-9.
- [25] Bouchez D, Camilleri C, Caboche M. A binary vector based on Basta resistance for *in planta* transformation of *Arabidopsis thaliana* [J]. C R Acad Sci Ser III Sci Vie, 1993, 316:1188-1193.
- [26] Campisi L, Yang Y, Yi Y, et al. Generation of enhancer trap lines in *Arabidopsis* and characterization of expression patterns in the inflorescence [J]. Plant J, 1999, 17:699-707.
- [27] Krysan P J, Young J C, Sussman M R. T-DNA as an insertional mutagen in *Arabidopsis* [J]. Plant Cell, 1999, 11:2283-2290.
- [28] Weigel D, Ahn J H, Blazquez M A, et al. Activation tagging in *Arabidopsis* [J]. Plant Physiol, 2000, 122:1003-1014.
- [29] Bechtold N, Ellis J, Pelletier G. *In planta Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants [J]. C R Acad Sci Ser III Sci Vie, 1993, 316:1194-1199.
- [30] Ohba T, Yoshioka Y, Machida C, et al. DNA rearrangement associated with the integration of T-DNA in tobacco: an example for multiple duplications of DNA around the integration target [J]. Plant J, 1995, 7:157-164.
- [31] Nacry P, Camilleri C, Courtial B, et al. Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis* [J]. Genetics, 1998, 149:641-650.
- [32] Laufs P, Autran D, Traas J. A chromosomal paracentric inversion associated with T-DNA integration in *Arabidopsis* [J]. Plant J, 1999, 18:131-139.
- [33] Greenblatt I M. A chromosome replication pattern deduced from pericarp phenotypes resulting from movements of the transposable element *Modulator* in maize [J]. Genetics, 1984, 108:471-485.
- [34] Sundaresan V, Springer P, Volpe T, et al. Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements [J]. Genes Dev, 1995, 9:1797-1810.
- [35] Bancroft I, Bhatt A M, Sjodin C, et al. Development of an efficient two-element transposon tagging system in *Arabidopsis thaliana* [J]. Mol Gen Genet, 1992, 233:449-461.
- [36] Aarts M G M, Corzaan P, Stiekema W J, et al. A two-element

- Enhancer-Inhibitor* transposon system in *Arabidopsis thaliana* [J]. *Mol Gen Genet*, 1995, 247:555–564.
- [37] Benfey P N, Ren L, Chua N H. The CaMV 35S enhancer contains at least two domains which can confer different developmental and tissue-specific expression patterns [J]. *EMBO J*, 1989, 8:2195–2202.
- [38] Wessler S R, Baran G, Varagona M. The maize transposable element *Ds* is spliced from RNA [J]. *Science*, 1987, 237:916–918.
- [39] Nussaume L, Harrison K, Klimyuk V, et al. Analysis of splice donor and acceptor site function in a transposable gene trap derived from the maize element *Activator* [J]. *Mol Gen Genet*, 1995, 249:91–101.
- [40] Martienssen R A. Functional genomics: probing plant gene function and expression with transposons [J]. *Proc Natl Acad Sci USA*, 1998, 95:2021–2026.
- [41] Liu Y G, Mitsukawa N, Oosumi T, et al. Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR [J]. *Plant J*, 1995, 8:457–463.
- [42] Tsugeki R, Kochieva E Z, Fedoroff N V. A transposon insertion in the *Arabidopsis* *SSR16* gene causes an embryodeficient lethal mutation [J]. *Plant J*, 1996, 10:479–489.
- [43] Skarnes W C, Auerbach B A, Joyner A L. A gene trap approach in mouse embryonic stem cells: the *lacZ* reporter is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice [J]. *Genes Dev*, 1992, 6: 903–918.
- [44] Springer P S, McCombie W R, Sundaresan V, et al. Gene trap tagging of *PROLIFERA*, an essential *MCM2-3-5*-like gene in *Arabidopsis* [J]. *Science*, 1995, 268:877–880.
- [45] Kumaran M K, Ye D, Yang W C, et al. Molecular cloning of *ABNORMAL FLORAL ORGANS*: a gene required for flower development in *Arabidopsis* [J]. *Sex Plant Reprod*, 1999, 12:118–122.
- [46] Yang W C, Ye D, Xu J, et al. The *SPOROCTELESS* gene of *Arabidopsis* is required for initiation of sporogenesis and encodes a novel nuclear protein [J]. *Genes Dev*, 1999, 13:2108–2117.
- [47] Parinov S, Sevugan M, Ye D, et al. Analysis of flanking sequences from *Dissociation* insertion lines: a database for reverse genetics in *Arabidopsis* [J]. *Plant Cell*, 1999, 11:2263–2270.
- [48] Parinov S, Sundaresan V. Functional genomics in *Arabidopsis*: large-scale insertional mutagenesis complements the genome sequencing project [J]. *Curr Opin Biotech*, 2000, 11: 157–161.
- [49] Ramachandran S, Sundaresan V. Transposons as tools for functional genomics [J]. *Plant Physiol Biochem*, 2001, 39:243–252.
- [50] Malamy J E, Benfey P N. Organization and cell differentiation in lateral roots of *Arabidopsis thaliana* [J]. *Development*, 1997, 124: 33–44.
- [51] Topping J F, Agyeman F, Henricot B, et al. Identification of molecular markers of embryogenesis in *Arabidopsis thaliana* by promoter trapping [J]. *Plant J*, 1994, 5:895–903.
- [52] Topping J F, Lindsey K. Promoter trap markers differentiate structural and positional components of polar development in *Arabidopsis* [J]. *Plant Cell*, 1997, 9:1713–1725.
- [53] Barthels N, Van der Lee F M, Klap J, et al. Regulatory sequences of *Arabidopsis* drive reporter gene expression in nematode feeding structures [J]. *Plant Cell*, 1997, 9:2119–2134.
- [54] Favery B, Lecomte P, Gil N, et al. A plant gene involved in early developmental steps of nematode feeding cells [J]. *EMBO J*, 1998, 17:6799–6811.
- [55] He Y H, Gan S S. Identical promoter elements are involved in regulation of the *OPRI* gene by senescence and jasmonic acid in *Arabidopsis* [J]. *Plant Mol Biol*, 2001, 47:595–605.
- [56] He Y H, Tang W N, Swain J D, et al. Networking senescence-regulating pathways by using *Arabidopsis* enhancer trap lines [J]. *Plant Physiol*, 2001, 126:707–716.
- [57] He Y H, Gan S S. A gene encoding an acyl hydrolase is involved in leaf senescence in *Arabidopsis* [J]. *Plant Cell*, 2002, 14:805–815.
- [58] He Y H, Fukushige H, Hildebrand D F, et al. Evidence supporting a role of jasmonic acid in *Arabidopsis* leaf senescence [J]. *Plant Physiol*, 2002, 128:876–884.
- [59] Meinke D W. Embryonic mutants in *Arabidopsis thaliana* [J]. *Dev Genet*, 1991, 12:382–392.
- [60] Springer P S, Holding D R, Groover A, et al. The essential *Mcm7* protein *PROLIFERA* is localized to the nucleus of dividing cells during the G1 phase and is required maternally for early *Arabidopsis* development [J]. *Development*, 2000, 127:1815–1822.
- [61] Gu Q, Ferrándiz C, Yanofsky M F, et al. The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development [J]. *Development*, 1998, 125:1509–1517.
- [62] Ferrándiz C, Gu Q, Martienssen R, et al. Redundant regulation of meristem identity and plant architecture by *FRUITFULL*, *APETALA1* and *CAULIFLOWER* [J]. *Development*, 2000, 127: 725–734.
- [63] Dubreucq B, Berger N, Vincent E, et al. The *Arabidopsis* *AtEPRI* extensin-like gene is specifically expressed in endosperm during seed germination [J]. *Plant J*, 2000, 23:643–652.
- [64] Swaminathan K, Yang Y Z, Grotz C, et al. An enhancer trap line associated with a D-class cyclin gene in *Arabidopsis* [J]. *Plant Physiol*, 2000, 124:1658–1667.
- [65] Hiwatashi Y, Nishiyama T, Fujita T, et al. Establishment of gene-trap and enhancer-trap systems in the moss *Physcomitrella patens* [J]. *Plant J*, 2001, 28:105–116.
- [66] Weijers D, Dijk M F V, Vencken R J, et al. An *Arabidopsis* minute-like phenotype caused by a semi-dominant mutation in a *RIBOSOMAL PROTEIN S5* gene [J]. *Development*, 2001, 128: 4289–4299.
- [67] Yoshioka Y, Kurei S, Machida Y. Identification of monofunctional aspartate kinase gene of *Arabidopsis thaliana* with spatially and temporally regulated expression [J]. *Genes Genet Syst*, 2001, 76: 189–198.
- [68] Hall Q, Cannon M C. The cell wall hydroxyproline-rich glycoprotein RSH is essential for normal embryo development in

- Arabidopsis* [J]. *Plant Cell*, 2002, 14:1161-1172.
- [69] Lechner E, Goloubinoff P, Genschik P, et al. A gene trap *Dissociation* insertion line, associated with a *RING-H2* finger gene, shows tissue specific and developmental regulated expression of the gene in *Arabidopsis* [J]. *Gene*, 2002, 290:63-71.
- [70] Tanaka S, Mochizuki N, Nagatani. Expression of the *AtGH3a* gene, an *Arabidopsis* homologue of the soybean *GH3* gene, is regulated by phytochrome B [J]. *Plant Cell Physiol*, 2002, 43: 281-289.
- [71] Tsugeki R, Fedoroff N V. Genetic ablation of root cap cells in *Arabidopsis* [J]. *Proc Natl Acad Sci USA*, 1999, 96:12941-12946.
- [72] Wagner-Bernholz J T, Wilson C, Gibson G, et al. Identification of target genes of the homeotic gene *Antennapedia* by enhancer detection [J]. *Genes Dev*, 1991, 5:2467-2480.
- [73] Bellen H J. Ten years of enhancer detection: lessons from the fly [J]. *Plant Cell*, 1999, 11:2271-2281.
- [74] Smith D L, Fedoroff N V. *LRPI*, a gene expressed in lateral and adventitious root primordia of *Arabidopsis* [J]. *Plant Cell*, 1995, 7: 735-745.
- [75] Di Laurenzio L, Wysocka-Diller J, Malamy J E, et al. The *SCARECROW* gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root [J]. *Cell*, 1996, 86:423-433.
- [76] Vielle-Calzada J P, Baskar R, Grossniklaus U. Delayed activation of the paternal genome during seed development [J]. *Nature*, 2000, 404:91-94.
- [77] Taylor C B. Promoter fusion analysis: an insufficient measure of gene expression [J]. *Plant Cell*, 1997, 9:273-275.
- [78] Fobert PR, Labbé H, Cosmopoulos J, et al. T-DNA tagging of a seed coat-specific cryptic promoter in tobacco [J]. *Plant J*, 1994, 6: 567-577.
- [79] Barakat A, Carels N, Bernardi G. The distribution of genes in the genomes of Graminae [J]. *Proc Natl Acad Sci USA*, 1997, 94: 6857-6861.
- [80] Barakat A, Gallois P, Reynal M, et al. The distribution of T-DNA in the genomes of transgenic *Arabidopsis* and rice [J]. *FEBS Lett*, 2000, 471:161-164.
- [81] Delseay M, Salses J, Cooke R, et al. Rice genomics: present and future [J]. *Plant Physiol Biochem*, 2001, 39:323-334.