

水稻基因的图位克隆技术

吴自明 (江西农业大学, 作物生理生态与遗传育种教育部重点实验室, 农业部双季稻生理生态与栽培重点实验室, 江西省作物生理生态与遗传育种重点实验室, 江西南昌 330045)

摘要 综述了水稻图位克隆技术的原理、技术环节及其在水稻基因克隆上的应用, 分析了当前存在的主要问题, 并对其应用前景作出了展望。

关键词 水稻; 图位克隆; 基因

中图分类号 S511 文献标识码 A 文章编号 0517 - 6611(2008) 34 - 14905 - 02

Map based Cloning Technique of Rice Genes

WU Zi-ming (Key Laboratory of Crop Physiology, Ecology and Genetic Breeding, Ministry of Education, Key Laboratory of Physiology, Ecology and Cultivation of Double Cropping Rice, Ministry of Agriculture, Key Laboratory of Crop Physiology, Ecology and Genetic Breeding of Jiangxi Province, Jiangxi Agricultural University, Nanchang, Jiangxi 330045)

Abstract The principle and technical links of map-based gene cloning technique in rice and its applications in the gene cloning of rice were summarized. And the main existing problems at present were analyzed. And its application foreground was predicted.

Key words Rice; Map-based cloning; Gene

近年来, 水稻基因组研究进展非常迅速, 高密度遗传图谱和物理图谱的构建, 全基因组序列的公布, 数以万计 EST、全长 cDNA 等序列及功能分析数据的释放以及新型水稻分子标记及其高效检测技术的发展等, 为基因的图位克隆带来了新的思路和方法。同时, 这些新的进展也能够使基因的精确定位和物理图谱构建等相关工作大大简化, 使基因的图位克隆朝着更加简便、快速的方向发展。

1 图位克隆技术原理

图位克隆 (Map-based Cloning) 又称定位克隆 (Positional Cloning), 1986 年首先由剑桥大学的 Alan Coulson 提出^[1]。用该方法分离基因是根据目的基因在染色体上的位置进行, 无需预先知道基因的 DNA 序列, 也无需预先知道其表达产物的有关信息, 而是通过分析突变位点与已知分子标记的连锁关系来确定突变表型的遗传基础。实现基因图位克隆的关键是筛选与目标基因连锁的分子标记。

近几年来, 水稻各种分子标记的日趋丰富和各种数据库的日趋完善, 在很大程度上得益于以粳稻日本晴和籼稻 9311 为代表的基因组测序的完成^[2-3]。目前已有几十种技术可用于分子标记筛选, 其中最常用的是简单序列长度多态性 (SSLPs)、单核苷酸多态性 (SNPs) 和插入缺失多态性 (Insertion/Deletion, InDel)^[4-7]。Shen 等利用日本晴和 9311 基因组序列构建了水稻基因组水平的 DNA 多态性数据库, 其中包括 1 703 176 个单核苷酸多态性 (Single Nucleotide Polymorphism, SNP) 和 479 406 个插入缺失多态性 (InDel)^[8]。Feltus 等通过对除去多重拷贝序列及低质量序列后的日本晴和 9311 基因组草图的比对分析, 得到 408 898 个 DNA 多态性, 包括 SNP 和单碱基 InDel^[9], 这些差别的核苷酸通常位于非编码区^[10]。

目前, 常把 SNP 多态性转化成基于 PCR 的剪切扩增多态性 (Cleaved Amplified Polymorphic Sequences, CAPS) 或 CAPS 衍生的 dCAPS 标记^[11-12]。CAPS 标记是 PCR 反应和酶切相结

合产生的一种分子标记。如果不同材料间在 PCR 扩增区域有 SNP 位点, 且该位点是限制性内切酶作用位点, 那么不同材料的 PCR 扩增产物经特定的酶切后, 再进行琼脂糖凝胶电泳就会表现多态性。当 SNP 恰好位于限制性酶切位点比较少时, 这种情况可以在 CAPS 标记的基础上通过在扩增引物中引入错配碱基, 则可以结合 SNP 位点引入新的限制性内切酶作用位点, 产生和 CAPS 标记类似的多态性, 这就是 dCAPS 的方法。用 CAPS 或 dCAPS 的方法则可以将几乎所有的 SNP 位点转化成以 PCR 为基础的分子标记^[12]。

2 图位克隆技术环节

2.1 构建遗传作图群体 对于基因图位克隆而言, 构建特殊的遗传作图群体是筛选与目标基因紧密连锁分子标记的关键环节。常用的作图群体有 F₂、近等基因系、重组自交系等群体, 水稻常用 F₂ 群体。创建 F₂ 群体应注意优先选择基因组已测序的日本晴 9311 和培矮 64S 等品种为亲本之一。

2.2 基因初定位 一般说来, 当标记为显性遗传时, 欲获得最大遗传信息量的 F₂ 群体, 需借助于进一步子代测验, 以分辨 F₂ 中的杂合体。为此, Michelmore 等发明分离群体分组分析法 (Bulked Segregant Analysis, BSA) 以筛选目标基因所在局部区域的分子标记^[13]。其原理是将目标基因的 F₂ (或 BC) 代分离群体各个体仅以目标基因所控制的性状按双亲表型分为 2 群, 每一群中各个体 DNA 等量混合, 形成 2 个 DNA 混合池 (如抗病和感病、不育和可育), 并且用目的基因附近的所有分子标记对混合 DNA 样品池进行分析, 根据所有池中包含有交换的 DNA 池的比例来确定与目的基因连锁最紧密的分子标记和目的基因附近所有分子标记的顺序。混合样品作图可以极大提高分子标记分析效率, 减小 DNA 提取工作量。

2.3 基因精细定位 一旦把目标基因初步定位在 2 个标记之间后, 就可以从国际水稻基因组测序计划 (IRGSP) 公布的序列中下载这 2 个标记区域的部分 PAC/BAC 克隆序列。利用软件 SSRIT 搜索克隆序列中的微卫星序列, 然后选择合适的微卫星序列进行特异 PCR 引物的设计。也可以直接借助于公共数据库的序列进行比较, 如寻找 RGP 基因组数据库 (粳稻) 与中国华大基因组数据库 (籼稻) 的单核苷酸多态性 (SNP) 序列差异, 设计 CAPS 或 dCAPS 标记和插入/缺失多态

基金项目 江西省教育厅项目 (GJ09477); 江西农业大学博士启动基金; 江西农业大学校自然科学基金。

作者简介 吴自明 (1974 -), 男, 江西鄱阳人, 副教授, 从事植物分子生物学研究。

收稿日期 2008-10-06

性标记(InDel)等。利用这些标记对大群体进行分析,可以迅速地将目标基因范围缩小到一个很小的基因组区域。

2.4 功能互补验证 完成基因精细定位后,可以根据已有的与目标基因紧密连锁的分子标记,找到目标基因在已公布的RGP物理图谱上的位置,通过序列拼接,实现目标基因区域遗传图谱和物理图谱的电子整合,得到该区域的基因组序列,从而减少文库构建、筛选等繁重的工作,加速图位克隆进程。从一系列候选基因中鉴定目标基因是定位克隆技术的最后一个关键环节。一种找到突变基因的有效方法是测序,通过设计PCR引物来扩增覆盖最小目标区域的多个重叠片段。将这些片段测序后拼接起来以得到整个目标区域的序列,然后将它与野生型的序列进行比对,就可以找到这个区域中的目标基因。现在最常用的方法是用含有目标基因的大片段克隆,如BAC克隆或YAC克隆去筛选cDNA文库,并查询生物数据信息库,待找出候选基因后,把这些候选基因进行下列分析以确定目标基因: 精确定位法检查cDNA是否与目标基因共分离; 检查cDNA时空表达特点是否与表型一致; 测定cDNA序列,查询数据库,以了解该基因的功能; 筛选突变体文库,找出DNA序列上的变化及与功能的关系; 进行功能互补试验,通过转化突变体观察突变体表型是否恢复正常或发生预期的表型变化。功能互补试验是最直接的最终鉴定基因的方法。利用新兴的RNA干扰技术(RNAi)也可有效地确定目的基因。

3 图位克隆技术的应用

水稻高精密遗传图谱和物理图谱的相继构建成功^[14],尤其是全基因组测序结果的公布^[2-3],为水稻基因的精细定位以及后续图位克隆创造了优越的条件。截至2006年11月,Oryabase网站(<http://www.shigen.nig.ac.jp>)公布了约2356个水稻突变体(含数量性状基因)。自1995年应用图位克隆技术在水稻中成功分离出Xa21基因以来,已克隆了50多个水稻突变体基因,其中包括我国克隆的MOC1、BC1、ALK、GH2、GS3和yg1等基因。

4 问题与展望

图位克隆过程中常会遇到许多问题,使工作难以进行。首先,在分析自然发生的变异时,最有可能遇到的复杂情况是一个给定的性状由不止一个基因位点控制,这样很难或者不能通过图位克隆技术来克隆基因。对这些基因中的任何一个位点精细定位都要求降低作图群体的遗传复杂性,通过高代回交创造只有一个位点保持多态性的重组近交系;其次,表观遗传是图位克隆工作中又一个可能遇到的复杂情况。表观遗传是描述一个基因在表达和功能上的可遗传改变,而不涉及DNA序列的改变^[15]。已有文献证明花发育基因SUPERMAN衍生的dark kant等位基因是可遗传的,在SUPERMAN基因的DNA序列中都具有相似的胞 甲基化现象,能减少SUPERMAN基因转录子的表达,但能被带有SUPERMAN基因的转基因所恢复^[16]。目前,对于这种表观遗传突变产生的原因以及出现的频率研究较少;最后,位于或接近着丝粒区域限制了图位克隆进程。通常染色体上位点的物理和遗传距离的比值变化较小,对作图影响较小^[17]。然而着丝粒附近较小的重组率严格限制了精细定位。在这个区域中重复DNA单元的广泛分布使研究

者很难辨认出散布的单拷贝序列,这些单拷贝序列能产生有疑问的遗传标记。一般在水稻染色体全序列1%重组的遗传距离相当于100~400 kb的物理距离,平均是250 kb,然而着丝粒区域1%重组的遗传距离相当于1000~2500 kb。最近对着丝粒区域的分析显示,这些区域通常包含重复DNA,几乎不含表达的基因^[18]。因此,位于或接近着丝粒区域的突变基因,不适合使用图位克隆技术进行克隆。另外,作图分析的株系DNA序列的重排,基因片段染色体之间的复制,以及线粒体基因组的DNA片段向染色体转移等,给图位克隆工作带来较大困难。

图位克隆有自身的局限性,整个过程涉及遗传图谱、物理图谱的构建,基因组文库或cDNA文库的建立与筛选,近等基因系等群体的构建以及DNA测序,基因表达分析和互补测验等,这些环节的工作比较复杂,所花费的人力、物力和时间较多。然而近些年来生物技术和生物信息学的迅猛发展,大大加快了水稻图位克隆工作的进行。随着对水稻基因组结构的进一步认识,也将有助于解决上述提到的复杂问题。

参考文献

- [1] COULSON A, SULTON J, SYDNEY BRENNER. Toward a physical map of the genome of the nematode *C. elegans* [J]. *Proc Natl Acad Sci USA*, 1986, 83: 7821 - 7825.
- [2] YU J, HUS N, WANG J, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*) [J]. *Science*, 2002, 296: 79 - 92.
- [3] GOFF S A, RICKE D, LANT H, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *Japonica*) [J]. *Science*, 2002, 296: 92.
- [4] LUKOWITZ W, GILMOR C S, SCHEIBLE W R. Positional cloning in Arabidopsis. Why it feels good to have a genome initiative working for you [J]. *Plant Physiology*, 2000, 123: 795 - 805.
- [5] CHOI S, SCHMITZ R J, FUJIOKA S, et al. Arabidopsis brassinosteroid insensitive dwarf12 mutants are semidominant and defective in a glycogen synthase kinase 3 beta-like kinase [J]. *Plant Physiology*, 2002, 130: 1506 - 1515.
- [6] GONZALEZ GUZMAN M, APOSTOLOVA N, BELLES J M, et al. The short-chain alcohol dehydrogenase ABA2 catalyzes the conversion of xanthoxin to abscisic aldehyde [J]. *Plant Cell*, 2002, 14: 1833 - 1846.
- [7] RAFALSKI A. Applications of single nucleotide polymorphisms in crop genetics [J]. *Curr Opin Plant Biol*, 2002, 5: 94 - 100.
- [8] SHEN Y J, JIANG H, JIN J P, et al. Development of genome-wide DNA polymorphism database for map-based cloning of rice genes [J]. *Plant Physiol*, 2004, 135: 1198 - 1205.
- [9] FELTUS F A, WAN J, SCHULZE S R, et al. An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments [J]. *Genome Res*, 2004, 14: 1812 - 1819.
- [10] PETERS J L, CNUDE F, GERAIS T. Forward genetics and map-based cloning approaches [J]. *Trends in Plant Science*, 2003, 8: 484 - 491.
- [11] NAM H G, GRAUDAT J, DEN BOER B, et al. Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana* [J]. *Plant Cell*, 1989, 1: 699 - 705.
- [12] MICHAELS S D, AMASINO R M. A robust method for detecting single nucleotide changes as polymorphic markers by PCR [J]. *Plant J*, 1998, 14: 381 - 385.
- [13] MICHELMORE R W, PARANI, KESSEI R V, et al. Identification of markers linked to disease resistance genes by bulked segregation analysis: A rapid method to detect markers in specific genomic regions by using segregating population [J]. *Proc Natl Acad Sci USA*, 1991, 88: 9828 - 9832.
- [14] CHEN M, PRESTING G, BARBAZUK W B, et al. An integrated physical and genetic map of the rice genome [J]. *Plant Cell*, 2002, 14: 537 - 545.
- [15] WOLFFE A P, MATZKE M A. Epigenetics: regulation through repression [J]. *Science*, 1999, 286: 481 - 486.
- [16] JACOBSONS E, MEYEROWITZ E M. Hypermethylated SUPERMAN epigenetic alleles in *Arabidopsis* [J]. *Science*, 1997, 277: 1100 - 1103.
- [17] COPENHAVER G P, BROWNE W E, PREUSS D. Assaying genome-wide recombination and centromere functions with *Arabidopsis* tetrads [J]. *Proc Natl Acad Sci USA*, 1998, 95: 247 - 252.
- [18] COPENHAVER G P, NICKEL K, KUROMORI T, et al. Genetic definition and sequence analysis of *Arabidopsis* centromeres [J]. *Science*, 1999, 286: 2468 - 2474.