

# 水稻 $S_5$ 区候选克隆 R2I19 的筛选及序列信息学分析<sup>\*</sup>

王宏斌, 刘 兵, 易厚富, 范 云, 刘良式, 王金发  
(基因工程教育部重点实验室 // 中山大学生命科学学院, 广东 广州 510275)

**摘 要:** 以水稻广亲和品种 Cpslo17 为材料, 构建了一个覆盖 9 倍核基因组的 cosmid 文库, 其平均插入片段约 40 kb。以与广亲和位点紧密连锁的单拷贝分子标记 BAC23D12 的 R 末端为探针, 从 cosmid 文库中筛选得到一个阳性克隆 R2I19 (~32 kb)。结合  $S_5$  位点的高密度连锁图和物理图谱, 初步确定为  $S_5$  区候选克隆。对该克隆的 2 个 TAC 亚克隆 TRW1510 (~15 kb) 及 TRW1517 (~15 kb) 进行了初步的生物信息学分析, 证实与已知的水稻基因组序列有很高的同源性, 并显示其中可能含有与水稻育性相关的基因。

**关键词:** 水稻 *Oryza sativa* L.; cosmid 文库; 广亲和基因; 生物信息学

**中图分类号:** Q751 **文献标识码:** A **文章编号:** 0529-6579 (2002) 06-0078-05

栽培稻籼粳亚种间杂交子代具有较强的杂种优势: 产量高、适应性广、抗逆能力强等。但是籼粳杂种普遍存在着不育或半不育现象。

Ikehashi 和 Araki 根据不同生态类型水稻品种间的杂交研究, 发现一些品种分别与籼、粳不同亚种杂交时, 都能产生正常可育的杂种  $F_1$  的遗传现象, 从而提出了广亲和品种 (Wide compatibility varieties, WCV) 的概念。他们把相应的控制广亲和性的基因——广亲和基因 (wide compatibility gene, WCG) 定位于第 6 染色体的  $S_5$  座位<sup>[1]</sup>。此后, 国内外许多研究者先后开展了广亲和基因的分子标记定位研究, 发现  $S_5$  基因座与 RG138、RG213、R2349 等分子标记紧密连锁, 并将其定位在分子标记 23D12 和 A30 之间的一个 1.0 cM 的区间内<sup>[2-4]</sup>。研究还发现, 广亲和基因和育性相关基因存在多个非等位基因座, 但第 6 染色体广亲和基因  $S_5$  的效应较大<sup>[5,6]</sup>。

为了克隆水稻第 6 染色体的广亲和基因, 我们采取图位克隆的策略, 通过构建广亲和品种 Cpslo17 的 cosmid 文库, 以广亲和基因连锁的分子标记 BAC23D12 的 R 末端为探针筛选到一个新克隆——R2I19。对该克隆的全序列进行了测定, 并进行了序列信息学分析, 初步推测与水稻育性相关。

## 1 材料与方法

### 1.1 材 料

#### 1.1.1 植物材料 水稻 *Oryza sativa* L. 广亲和品种

Cpslo17, 种籽由国家杂交水稻中心袁隆平院士惠赠。

1.1.2 菌株 DH10B 为大片段文库受体菌, 由华南农业大学刘耀光研究员惠赠; DH5 $\alpha$  及一般性宿主菌由本室保存种。

1.1.3 质粒 SuperCos1 载体购自 Stratagene 公司; pYLTAC17 由华南农业大学刘耀光研究员惠赠; pBluescript SK (一) 本室保存; BAC23D12 由华中农业大学农作物改良国家重点实验室张启发院士惠赠。

1.1.4 试剂 限制性内切酶、T4 连接酶均购自大连宝生物工程公司; 质粒提取试剂盒和 DNA 胶回收试剂盒购自上海华舜生物工程公司; 随机引物标记试剂盒购自 Amersham 及 GIBCO 公司。

### 1.2 cosmid 文库的构建及筛选

分离水稻广亲和品种 Cpslo17 的高分子量核 DNA, 经部分酶切后与 SuperCos1 载体连接, 构建 Cpslo17 核基因组的 cosmid 文库<sup>[7]</sup>。以与水稻广亲和基因座  $S_5$  区紧密连锁的分子标记 BAC23D12 的 R 末端 23D12R 为探针, 通过菌落原位杂交筛选 cosmid 文库。

### 1.3 R2I19 的染色体图位分析

用 Not I, BamH I 和 Hind III 3 种限制性内切核酸酶分别对 R2I19 和 BAC23D12 进行酶切, 酶切产物经琼脂糖凝胶电泳分析。

### 1.4 R2I19 的 TAC 亚克隆构建

R2I19 经 Not I 消化, 外源片段以 PEG 法回收,

<sup>\*</sup> 收稿日期: 2002-06-10

基金项目: 科技部植物转基因重大专项资助项目 (J00-A-009); 广东省自然科学基金资助项目 (990258)

作者简介: 王宏斌 (1977 年生) 男, 博士生; 通讯联系人: 王金发; E-mail: ls19@zsu.edu.cn

再与经 *Not* I 消化的 pYLTAC17 载体连接，连接产物用电激法转化大肠杆菌 DH10B。

1.5 TRW1510 及 TRW1517 的 pBlue 系列亚克隆构建及测序

将 R2I19 外源片段的 2 个 TAC 亚克隆 TRW1510、TRW1517 经 *Hind* III 酶切，分别得到 4 个外源 DNA 片段，用 DNA 回收试剂盒回收，与经 *Hind* II 酶切的 pBluescript 载体连接，连接产物分别转化大肠杆菌 DH5 $\alpha$ 。将分别筛选到的各重组子送上海 GeneCore 公司测序。

1.6 全长序列的拼接

根据各个亚克隆的末端序列设计拼接引物，以期通过 PCR 扩增得到相邻亚克隆间的衔接片段，根据接头片段的序列测定结果，确定各亚克隆间的相邻关系，从而进行全序列的拼接。

1.7 TRW1510 及 TRW1517 全序列的生物信息学分析

将 TRW1510、TRW1517 全序列分别用 BLAST 软件<sup>[8]</sup>与 GenBank 的 nr 库中所有已知的基因组序列进行同源性比较及初步的染色体定位分析；同时与华大基因中心公布的籼稻全基因组框架序列进行比较<sup>[9]</sup>，进一步验证 R2I19 与水稻基因组的关系。

2 结果与分析

2.1 cosmid 文库的构建、鉴定与筛选

以 SuperCosI 为载体，构建了一个水稻广亲和品种 Cps1017 的 cosmid 核基因组文库，其平均插入片段约为 40 kb，克隆总数 10 万以上，本文库的容量为普通水稻基因组 DNA 含量（1C）的 9 倍。经鉴定，文库具有较低的空载率（~5%），并且具有良好的稳定性与代表性<sup>[10]</sup>。以质粒拯救法分离得到的 BAC23D12 的 R 末端 23D12R（~1.5 kb）为探针，从该文库中筛选到一阳性克隆，插入片段大小约 32 kb，命名为 R2I19。

2.2 R2I19 染色体图位分析

由于 R2I19 是用 23D12R 为探针筛选到的克隆，因此需要正确分析 BAC23D12 与 R2I19 的重叠程度。为此，用 *Not* I，*Bam*HI 和 *Hind* III 等 3 种限制性内切核酸酶分别对 R2I19 和 BAC23D12 进行酶切。经琼脂糖凝胶电泳后进行酶切图谱分析，发现 R2I19 和 BAC23D12 的单酶切图谱中只有极少（约 2~3 kb）的片段相同（图 1）。由于 BAC23D12 的相对分子质量为 120 kb，因此推测 R2I19 和 BAC23D12 两克隆的重叠区域很小，并且 R2I19 向 *S<sub>5</sub>* 座位延伸。正常情况下水稻基因组的平均值 1 cm=250 kb，但在本实验室与刘克德分别进行的广

亲和基因的分子标记精细定位研究中，均发现此区域的遗传距离与物理距离存在一种异常的对应关系：本实验室刘文华推算 *S<sub>5</sub>* 区内 1 cm=70 kb（个人通讯），而刘克德推算 *S<sub>5</sub>* 区内 1 cm=35 kb<sup>[4]</sup>。由此可以推测所筛选到的 cosmid 克隆 R2I19 已经接近并有可能覆盖了部分 *S<sub>5</sub>* 座位。

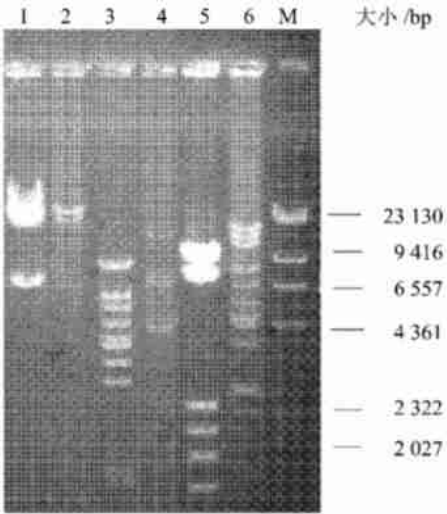


图1 R2I19 与 BAC23D12 限制酶谱的重叠关系  
Fig 1 The overlap relationship of R2I19 and BAC23D12

indicated by restrictive map  
1 R2I19/*Not* I ;  
2 BAC23D12/*Not* I ;  
3 R2I19/*Hind*III;  
4 BAC23D12/*Hind*III;  
5 R2I19/*Bam*HI ;  
6 BAC23D12/*Bam*HI ;  
M  $\lambda$  /*Hind*III核酸相对分子质量标准物

2.3 R2I19 的 TAC 亚克隆构建

R2I19 经 *Not* I 消化，得到两条约 16~17 kb 的外源片段，将外源片段回收后分别与经 *Not* I 消化的 TAC 载体连接，将连接产物经电激转化大肠杆菌 DH10B。筛选到两组亚克隆，相对分子质量大小分别为 ~15 kb 和 ~16 kb，经 *Hind* III 酶切鉴定它们是两种不同的重组子（图 2），命名为 TRW1510 和 TRW1517。

用限制性内切酶 *Hind* III 对 TAC 亚克隆 TRW1510 和 TRW1517 及 R2I19 进行完全酶切、电泳后，以 23D12R 为探针进行 Southern 杂交，发现 TRW1517、R2I19 与探针有明显的杂交信号，TRW1510 则没有（图 3（a）），因此，可认为 TRW1517 与 23D12R 部分重叠，而 TRW1510 远离 23D12R。鉴于 R2I19 与 BAC23D12 的重叠区域很少，推测 TRW1510 位于 BAC23D12 之外，并向 *S<sub>5</sub>*

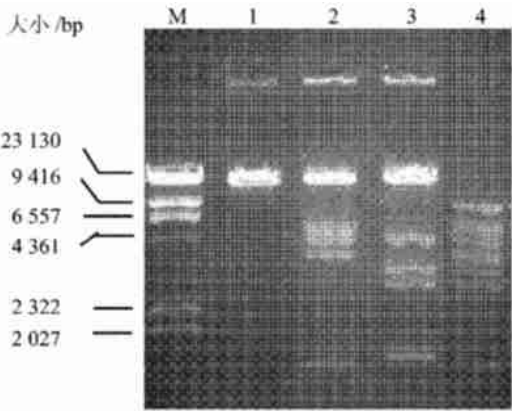


图 2 TAC 亚克隆的酶切鉴定

Fig. 2 Digestion with *Hind* III of TAC subclones  
M  $\lambda$ /*Hind* III 核酸相对分子质量标准物;  
1 pYLTAC17 vector/*Hind* III;  
2 TRW1510/*Hind* III;  
3 TRW1517/*Hind* III;  
4 cosmidR2119/*Hind* III

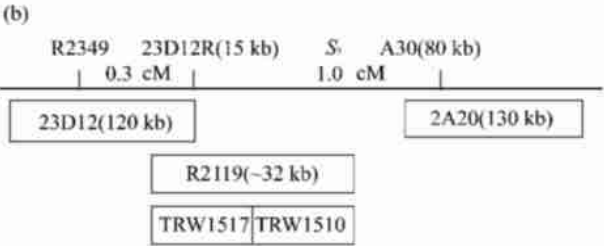
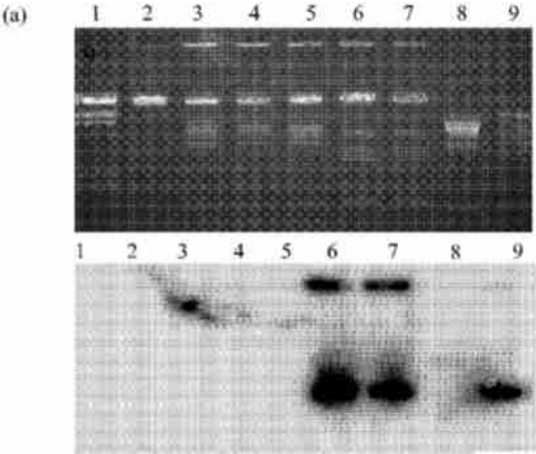


图 3 TAC 亚克隆的 Southern 分析 (a) 和  $S_5$  区的物理图谱 (b)

Fig. 3 Southern analysis of TAC subclones (a) and physical map of  $S_5$  locus  
(a) 1  $\lambda$ /*Hind* III 核酸相对分子质量标准物;  
2 TAC vector/*Hind* III; 3—5 TRW1510/*Hind* III;  
6—7 TRW1517/*Hind* III;  
8 cosmid21/*Hind* III (cosmid21 为文库中随机挑选的非阳性克隆);  
9 R2119/*Hind* III  
(b) Physical map of  $S_5$  region

方向延伸 (图 3 (b))。

2.4 TRW1510 及 TRW1517 的 pBlue 系列亚克隆构建及测序

为了便于进行测序, 将 TAC 载体上的外源片段构建成更小的亚克隆。TRW1510 经 *Hind* III 酶切后得到 4 个外源片段, 回收后分别与 pBluescript 载体连接, 连接产物转化 DH5 $\alpha$ , 筛选得到 pB1510-1、pB1510-2、pB1510-3、pB1510-4 共 4 个转化子, 将这 4 个亚克隆送上海 GeneCore 测序。相应的 1517 也获得 1517-1、1517-2、1517-3、1517-4 共 4 个亚克隆并测序。

2.5 TRW1510 及 TRW1517 全序列的拼接

依据测序结果, 针对每个亚克隆片段均设计两条末端引物, 通过组合 PCR 的方式扩增相邻两亚克隆片段间的接头片段。PCR 产物经克隆后测序, 根据各亚克隆序列的重叠性进行拼接。最终得到 TRW1510 全长 15607bp 的序列, 以及 TRW1517 全长 14668bp 的序列。由于某种原因, TRW1510 和 TRW1517 间尚有小段空缺。

2.6 TRW1510 及 TRW1517 全序列初步的生物信息学分析

2.6.1 TRW1517 信息学分析 将 TRW1517 全序列通过 BLAST 软件检索 GenBank 的 nr 库 (更新于 2002 年 4 月 24 日), 并没有发现显著同源的水稻基因组序列。尽管 Syngenta 公司已经完成了粳稻 *Oryza sativa* L. ssp. *japonica* 的全基因组框架序列<sup>[11]</sup>, 但由于其并未无条件向全世界公开, 因此无法作相应的全序列比较分析, 而 GenBank 中所提供的水稻基

因组序列为粳稻材料, 是由国际水稻基因组测序计划所负责, 目前还未完成<sup>[12]</sup>, 所以尚无法对 TRW1517 序列在粳稻染色体上的定位作一分析。但是, 与中国科学院基因组信息学中心及北京华大基因中心 (Beijing Genomics institute/Center of Genomic and Bioinformatics, Chinese Academy of Sciences) 公布的籼稻 *Oryza sativa* L. ssp. *indica* 全基因组框架序列进行比较<sup>[10]</sup>, 证实 TRW1517 与部分籼稻框架序列重叠 (结果未显示), 这提示该序列可能是籼稻与粳稻的共有序列。

2.6.2 TRW1510 信息学分析 将 TRW1510 全序列通过 BLAST 软件检索 GenBank 的 nr 库 (更新于 2002 年 4 月 24 日)。发现 TRW1510 中 8 541 ~ 15 438 bp 区间的 7 kb 序列与水稻粳稻第 1, 2, 3, 6, 5, 7, 9, 10 染色体部分序列有很高的同源性 (表 1), 其中在第 1, 6, 10 染色体上均检索到多条同源的序列, 尤其在第 1 染色体上包括了 53 个

BAC/PAC 克隆 (结果未显示)。通过将该段序列与 GenBank 的 EST 库进行检索, 并未发现任何与之同源的 EST 序列, 而在此区段的两侧却分别找到两个与开花期和成熟期有关的水稻 EST。

表 1 TRW1510 与 GenBank 水稻染色体序列的同源性比较  
Tab 1 Homology comparison between TRW1510 and sequence of rice chromosomes published on GenBank

| 染色体序号    | TRW1510 序列区间/bp | 同源区间长度/bp | 同源性/% |
|----------|-----------------|-----------|-------|
| 第 1 染色体  | 8 541~15 438    | 6 897     | 97    |
| 第 2 染色体  | 8 903~15 438    | 6 013     | 91    |
| 第 3 染色体  | 8 903~14 597    | 6 003     | 91    |
| 第 5 染色体  | 9 294~14 869    | 5 134     | 91    |
| 第 6 染色体  | 8 544~15 425    | 6 498     | 94    |
| 第 7 染色体  | 8 540~15 438    | 6 583     | 95    |
| 第 9 染色体  | 8 889~12 280    | 3 094     | 91    |
|          | 12 282~14 869   | 2 414     | 93    |
| 第 10 染色体 | 8 543~15 429    | 6 534     | 94    |

进一步将 TRW1510 全序列与中国科学院基因组信息学中心及北京华大基因中心公布的籼稻全基因组框架序列进行比较, 发现 TRW1510 序列基本与基因组框架序列重叠 (结果未显示)。

3 讨 论

随着水稻基因组大规模测序工作的完成, 寻找有价值的功能基因越来越具有重要意义。筛选含有特殊功能座位的大片段基因组克隆, 并搜索其中可能含有的表达序列是一条有效的途径。我们采取图位克隆的策略, 通过构建广亲和品种 Cps17 的 cosmid 文库, 以广亲和基因连锁的分子标记 (BAC23D12 的 R 末端) 为探针筛选到一个新克隆——R2I19。通过对所构建的 R2I19 两个 TAC 亚克隆 TRW1510 和 TRW1517 所进行的分段测序及拼接, 分别得到 TRW1510 及 TRW1517 的全序列。将这两段序列与 GenBank 中 nr 库的粳稻基因组序列以及中国科学院基因组信息学中心及北京华大基因中心公布的籼稻全基因组框架序列进行了比较, 并且在 GenBank 中的 EST 库中作了同源 EST 检索。

通过比较发现 TRW1510 中 8 541~15 438 bp 间这一 7 kb 区段与粳稻第 1—3, 5—7, 9, 10 染色体部分序列有很高的同源性, 表明该区段可能是一段重复序列。但该区段在 GenBank 的 EST 库中并未检索到较高同源性的 EST 序列, 提示此区段可能是一段非编码区。根据目前推测, 水稻基因组中每 15 kb 就有一个基因存在, 因此很难想像一段长达

7 kb 的非编码又无功能的“垃圾”序列却能在粳稻基因组的不同染色体上广泛存在。所以推测这一区段很可能含有某些调控元件以参与基因的表达调控。

另一方面, 由于迄今为止发现的广亲和基因不仅存在于水稻第 6 染色体上 (如 Dular 在第 5, 12 染色体上还存在有多个广亲和基因, 其中第 5 染色体上的广亲和基因也具有较大效应<sup>[5,9]</sup>), 并且对于亚种间杂种不育这一现象的遗传机理也没有一个统一的定论, 因此不排除多基因互作或基因表达调控影响的可能。所以, 对于 TRW1510 中 8 541~15 438 bp 的这一在水稻 1—3, 5—7, 9, 10 染色体上均具有高同源性区段的研究, 还可能有助于阐明不同染色体上广亲和基因及其它育性基因之间的关系及对于育性调控的作用机理。

有意思的是, 通过对该 7 kb 区段两侧的序列与 GenBank 的 EST 库进行检索, 分别找到了两个水稻开花期和成熟期粳稻花序的 EST, 该结果提示, 这两个与发育有关的基因的表达调控可能与这一区段有关。同时, 作为存在于水稻广亲和基因 *S<sub>5</sub>* 区候选克隆 R2I19 中的这 2 个 EST, 分别来自于与水稻育性有着直接关系的开花期和成熟期花序的 cDNA 文库, 它们有可能参与了育性的调控, 也可能与 *S<sub>5</sub>* 座位的广亲和基因有所关联。

另外, 在 TRW1517 序列中发现了两处水稻 EST 同源区, 其中一个 EST 可能与水稻酰胺水解酶有关, 而另一个 EST 则可能是一个新的膜结合蛋白, 相应全长 cDNA 的克隆及基因产物功能的研究正在进行中, 有关结果将另文发表。

参考文献:

[1] IKEHASHI H, ARAKI H. Genetics of *F<sub>1</sub>* sterility in remote crosses of rice // Rice genetics I . Proceeding of the international rice genetics symposium[ J] . IRRI, 1986, 119—130.  
[2] YANAGIHARA S, McCOUCH S R, ISHIKAWA K, et al. Molecular analysis of the inheritance of the *S-5* locus, conferring wide compatibility in Indica/Japonica hybrids of rice (*Oryza sativa* L. )[ J] . Theor Appl Genet, 1995, 90: 182—188.  
[3] LIU A, ZHANG Q, LIU H. Location of a gene for wide-compatibility in the RFLP linkage map[ J] . Rice Genet Newsl, 1992, 9: 134—136.  
[4] 刘克德. 水稻广亲和性遗传基础的全基因组分析及 *S<sub>5</sub>* 位点区段部分物理图谱的构建[ D] . 武汉: 华中农业大学, 1998.  
[5] WANG J, LIU K, XU G, et al. The high level of wide-compatibility of variety “Dular” has a complex genetic basis

[ J ] . Theor Appl Genet, 1998, 97: 407— 412.

[ 6 ] 严长杰, 梁国华, 朱立煌, 等. 秋稻品种 Dular 广亲和基因的 RFLP 分析[ J] . 遗传学报, 2000, 27(5): 409— 417.

[ 7 ] SAMBROOK J, FRITSCH E F, MANIATIS T. Molecular cloning-A laboratory manual[ M] . 2nd ed. Cold Spring harbor Laboratory Press, 1989.

[ 8 ] 胡德华, 方平. 基因库( Genbank) 的电子邮件检索[ J] . 遗传, 1999, 21(6): 43— 46.

[ 9 ] 于军, 胡松年, 王俊, 等. 籼稻( *Oryza sativa* ssp *indica*) 全基因组框架序列[ J] . 科学通报, 2001, 46(23): 1937— 1941.

[ 10] 易厚富, 刘兵, 范云, 等. 水稻广亲和品种 cosmid 文库的构建和鉴定[ J] . 热带亚热带植物学报, 2001, 9(3): 185— 189.

[ 11] GOFF S A, RICKE D, LAN T H, et al. A draft sequence of the rice genome( *Oryza sativa* L. ssp *japonica*) [ J] . Science, 2002, 296: 92— 100.

[ 12] IRGSP Sequencing Meeting 2002. [ C] February 5, 2002 (Tsukuba Japan). URL <http://demeter.bio.bnl.gov/Tsukuba02.html>

Identification and Sequence Analysis of a New Clone  
R2I19 in Rice ( *Oryza sativa* L.) *S*<sub>5</sub> Locus

WANG Hong bin, LIU Bing, YI Hou fu, FAN Yun, LIU Liang shi, WANG Jin fa

(The Key Laboratory of Gene Engineering of Ministry of Education //

School of Life Sciences, Sun Yat sen (Zhongshan) University, Guangzhou 510275, China)

**Abstract:** A genomic cosmid library of the wide compatibility variety rice ( *O. sativa* L. Cpslo17) had been generated using vector SuperCos1. The library represented more than 9 times haploid genomic equivalents with an average insert size of ~40 kb. A new clone R2I19 was isolated and determined using the *S*<sub>5</sub> tightly linked 23D12R as probe. Two sub-clones, TRW1510 and TRW1517 were constructed with pYLTAC17 as vector. Based on GenBank and published draft sequence of the rice genome ( *Oryza sativa* ssp *indica*), some sequence information of R2I19 have been analyzed.

**Key words:** rice ( *Oryza sativa* L. ); cosmid library; wide compatibility gene; bioinformatics