

## 水稻培矮 64S(PA64S) 基因组 BAC 文库的构建与分析

毛丹青, 罗美中

(华中农业大学作物遗传改良国家重点实验室, 武汉 430070)

**摘要:** 9311 和培矮 64S(PA64S) 分别是我国超级杂交稻两优培 9(LYP9) 的父母本, 9311 的基因组已由北京基因组研究所测序, 而 PA64S 至今仍无任何可供公众使用的基因组资源, 对 PA64S 基因组的研究将有助于对 9311 和 PA64S 这对亲本基因组的比较研究和其上优良基因的克隆, 为 LYP9 的基因组研究提供帮助。以水稻 PA64S 的幼嫩叶片作为材料, 为该品种构建了第一个高覆盖率、高质量的 BAC 文库。该文库一共有 37 632 个克隆, 平均插入片段为 138 kb, 覆盖水稻全基因组 12 倍左右, 空载率仅为 1.5%, 线粒体和叶绿体污染率分别为 0.35% 和 1.85%。该文库克隆被存放于 98 个 384 孔板中并储存在 -80℃ 冰箱中, 将作为公共基因组资源向研究者开放。

**关键词:** 水稻; PA64S; 细菌人工染色体; 基因组文库

doi:10.3969/j.issn.1008-0864.2010.03.18

中图分类号: Q785

文献标识码: A

文章编号: 1008-0864(2010)03-0103-05

## Construction and Analysis of a BAC Library of Rice PA64S Genome

MAO Dan-qing, LUO Mei-zhong

(National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China)

**Abstract:** 9311 and PA64S are the paternal and maternal cultivars of the super hybrid rice LYP9, respectively. The 9311 genome has been sequenced by Beijing Genomics Institute (BGI). However, for PA64S, no any public genomic resource is available to date. Research on PA64S genome will help in comparative studies between and cloning of agriculturally important genes from the two parental genomes 9311 and PA64S, facilitating the studies on LYP9 genome. In this paper, the first large insert and deep coverage BAC library for PA64S was constructed, taking the young tender leaf blade of PA64S as material. The library contains 37 632 clones with an average insert size of 138 kb, representing nearly 12-fold of rice genome. The empty-vector rate is only 1.5% and the mitochondrial and chloroplast DNA contamination rates are 0.35% and 1.85%, respectively. The clones have been deposited in 98 of 384-well plates and stored at -80℃ freezers, and are accessible to research communities as a public genome resource.

**Key words:** rice; PA64S; BAC; genomic DNA library

随着基因组时代的到来, 基因组文库在动植物基因组研究中发挥的作用已越来越重要。对高等动植物来说, 大片段基因组文库对于物理图谱构建、图位克隆、全基因组测序和完整基因的克隆都具有重要意义。细菌人工染色体 (bacterial artificial chromosome, BAC) 是应用较广的一种载体系统, 具有克隆片段大、外源片段易分离、嵌合体少、遗传特性稳定、转化效率高及易于操作等优

点, 在基因组文库构建和基因的快速克隆等方面已得到广泛应用<sup>[1,2]</sup>。至今, 国内外已构建了包括大多数重要农作物在内的大量动植物 BAC 文库<sup>[3-6]</sup>, 为基因组学的研究奠定了良好的基础。

超级杂交稻两优培 9(LYP9) 作为在我国广泛种植的杂交水稻, 其基因组及其两亲本的基因组研究已成为一个热点<sup>[7-10]</sup>。其中父本 9311 的全基因组序列已通过霰弹法 (whole genome shot-

收稿日期: 2010-01-26; 修回日期: 2010-03-29

基金项目: 华中农业大学人才引进启动基金和作物遗传改良国家重点实验室开放课题资助。

作者简介: 毛丹青, 硕士研究生, 主要从事植物 BAC 文库的构建研究。通讯作者: 罗美中, 教授, 博士生导师, 主要从事植物基因组和叶绿体功能研究。Tel: 027-87284213; E-mail: mzluo@mail.hzau.edu.cn

gun, WGS)测序完成<sup>[11,12]</sup>,对 LYP9 及两亲本的转录组方面的分析也已有报道<sup>[9,10]</sup>。然而,母本 PA64S 虽也进行了低覆盖度的 WGS 测序<sup>[11,12]</sup>,但其序列既没有发表,也没有对公众释放(于军,私人通信),至今还没有一个 BAC 文库。因此,构建一个高质量、高覆盖度的 PA64S 的 BAC 文库,将是研究超级杂交稻 LYP9 的杂种优势和克隆其亲本上优良功能基因的重要基因组资源。

## 1 材料与方法

### 1.1 材料

PA64S 种子由华中农业大学牟同敏教授惠赠,培养出苗后,取幼苗叶片进行实验。

### 1.2 方法

BAC 文库构建参照 Luo 等<sup>[13]</sup>的方法进行。

**1.2.1 载体的制备** 克隆载体 pHZAUBAC1 质粒是由本实验室从 pCUGIBAC1<sup>[5,6,13]</sup> 改建而来, pCUGIBAC1 由 pIndigoBAC536 和 pGEM-4Z 连接而成。本实验室将 pIndigoBAC536 作了改造,命名为 pIndigoBAC536-S, pCUGIBAC1 也相应改变,命名为 pHZAUBAC1(史雪等未发表结果)。pHZAUBAC1 质粒采用 Qiagen plasmid midi kit 提取并纯化,用 *Hind* III (NEB) 进行酶切,再用 CIAP (NEB) 酶去磷酸化,酚/氯仿/异戊醇抽提 1 次和氯仿/异戊醇抽提 2 次,无水乙醇沉淀 DNA,加水溶解 DNA 后进行自连反应,然后电泳分离目的 DNA 片段 pIndigoBAC536-S,用电洗脱法回收并估测浓度,根据 DNA 的浓度加入甘油 (Sigma) 调整浓度,分装后  $-80^{\circ}\text{C}$  保存。

**1.2.2 高分子量基因组 DNA 的提取** 取幼嫩的叶片材料,液氮研磨后,分离出细胞核,与等体积 1% 低熔点琼脂糖 (Sigma) 混匀后制成细胞核包埋块 plug。Plug 用蛋白酶 K (Merck) 处理 48 h 后,再用 PMSF (苯甲基磺酰氟) 去除蛋白酶 K,然后存放于  $4^{\circ}\text{C}$  的 TE 中备用。

**1.2.3 部分酶切和第一次筛选** 首先进行预先部分酶切反应,以找出最合适的酶用量和反应时间。每个反应体系为:切碎的 1/2 plug, 25  $\mu\text{L}$   $\text{H}_2\text{O}$ , 10  $\mu\text{L}$   $10\times$  buffer, 10  $\mu\text{L}$  40 mmol/L 亚精胺 (Sigma), 5  $\mu\text{L}$  *Hind* III (NEB) 稀释液。固定反应时间为 15 min,反应温度  $37^{\circ}\text{C}$ ,仅改变反应体系中的酶液浓度。反应结束后用脉冲场电泳 (Bio-

Rad) 检测酶切结果,电泳条件为:  $14^{\circ}\text{C}$ , 1 s ~ 50 s,  $0.5\times$  TBE,  $120^{\circ}$ , 6 V/cm, 19 h。根据电泳的结果决定最合适的酶浓度,然后以此反应条件进行数个相同的酶切反应,通过脉冲电泳进行大片 DNA 的分离,即第一次筛选,将 DNA 标记  $\lambda$  ladder PFG marker 点在样品两侧,电泳结束后切下两边含 DNA 标记的胶块进行溴化乙啶染色,使用直尺标记照像。

**1.2.4 第二次筛选和基因组 DNA 回收** 根据 DNA 标记和尺子刻度从未染色的凝胶中间部分切下 120 ~ 250 kb (a) 和 250 ~ 300 kb (b) 大小的胶块进行第二次电泳筛选,条件为:  $14^{\circ}\text{C}$ , 4 s ~ 4 s,  $0.5\times$  TBE,  $120^{\circ}$ , 6 V/cm, 18 h。结束后按照第一次筛选中的方法染色和照相,根据刻度切下未染色的含有 DNA 的胶块,将含有 120 ~ 250 kb 的胶横切成两半,分别命名为 a1 和 a2 (a1 < a2),含有 250 ~ 300 kb 的胶块则命名为 b,采用电洗脱法回收胶块中的 DNA,以已知浓度的  $\lambda$  DNA 来估测回收的大片段 DNA 的浓度。

**1.2.5 文库构建** 将载体与基因组 DNA 进行连接,反应体系为: 84  $\mu\text{L}$  基因组 DNA, 4  $\mu\text{L}$  载体, 10  $\mu\text{L}$   $10\times$  buffer, 2  $\mu\text{L}$  *T*<sub>4</sub> ligase (NEB)。  $16^{\circ}\text{C}$  反应过夜。连接产物在  $65^{\circ}\text{C}$  灭活后于冰上脱盐 1 h,取 2  $\mu\text{L}$  加入到 20  $\mu\text{L}$  感受态细胞 Electro-MAX DH10B T1 Phage-Resistant (Invitrogen) 中电转化。转化产物涂布在含有 12.5  $\mu\text{g}/\text{mL}$  氯霉素、80  $\mu\text{g}/\text{mL}$  X-gal 和 100  $\mu\text{g}/\text{mL}$  IPTG 的 LB 平板上培养,得到克隆后,挑取一定数量的白色克隆检测插入片段的大小及空载率,符合要求后,立即将剩余的连接产物全部转化,挑取白色克隆至 384 孔板中,复制一个拷贝后,  $-80^{\circ}\text{C}$  保存。

**1.2.6 文库的质量分析** 插入片段的检测:从每块 384 孔板中随机挑取 2 个克隆,提取质粒并用 *I-Sce* I 酶切,脉冲电泳检测插入片段,条件为  $14^{\circ}\text{C}$ , 5 s ~ 15 s,  $0.5\times$  TBE,  $120^{\circ}$ , 6 V/cm, 16 h。计算平均插入片段和空载率,由此估算全基因组覆盖度。

细胞器 DNA 污染的检测:用 Beckman2000 自动工作站按  $3\times 3$  点阵在 Hybond N<sup>+</sup> (GE) 膜上复制 BAC 克隆,制备全文库的高密度膜,每张膜上点有  $384\times 8$  共 3 072 个克隆。膜的处理参照 Zhang<sup>[14]</sup> 的方法。细胞器 DNA 探针制备和 Southern 杂交均按前人方法进行<sup>[5,6]</sup>。

## 2 结果与分析

### 2.1 载体制备

克隆载体 pIndigoBAC536-S 经酶切脱磷线性化后,根据 λ DNA 的浓度估算,其 DNA 的浓度大约为 32 ng/μL。由于 DNA 的浓度较高,因此以 1:1 的比例加入甘油对其进行了稀释。

为了检测载体的质量,将载体与用 *Hind* III 酶切后的 λ DNA 以 64 ng:50 ng 的质量比在 16℃ 连接 16h,脱盐后进行转化,同时设置一个载体自连反应作为对照。结果表明,自连反应中每涂布 100 μL 菌液蓝斑和白斑的数量均在 10 个以下,而外源片段连接反应中蓝斑极少,白斑则达到 700~800 个,证明本实验室制备的载体连接外源片段的效率非常高且自连得到有效控制。

### 2.2 基因组 DNA plug 的制备及部分酶切条件摸索

本研究采用在相同时间条件下改变酶的用量的方法,设计了酶量梯度,对相同量的 DNA 进行 *Hind* III 酶切,结果显示,plug 中基因组 DNA 完整性好,DNA 浓度很高,几乎没有降解,DNA 较纯(图 1,泳道 5);经切碎后,DNA 有稍许损伤或降解(图 1,泳道 1);加入 5U/半个 plug 酶切时,酶解不够(图 1,泳道 2);酶量加大到 10U/半个 plug,

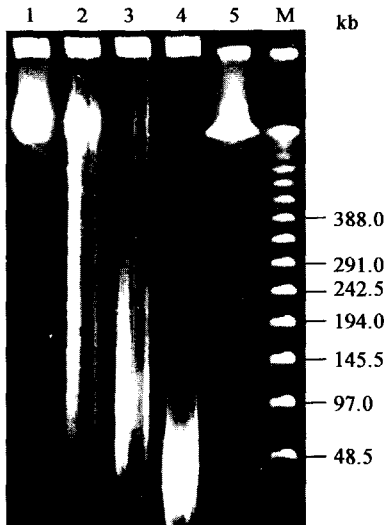


图 1 水稻 PA64S 基因组 DNA 预先部分酶切

Fig.1 Pilot partial digestion of rice PA64S genomic DNA. 泳道 1-4:分别为每半个 plug 加入 0.5 U、10 U、50 U *Hind* III;泳道 5:完整的半个 plug; M:λ ladder PFG 分子量标记  
Lanes 1-4:0.5, 10, 50U of *Hind* III per half plug, respectively; Lane 5: Intact half plug; M: λ ladder PFG marker

DNA 大部分都集中于 97 kb 左右(图 1,泳道 3);在加入 50U/半个 plug 的 *Hind* III 时,DNA 已酶解至小于 50 kb(图 1,泳道 4)。因此,需要在这 5U 和 10U 酶量之间再设置梯度,以找出最适合的酶切浓度获取范围较宽的 DNA 大片段。经过预先部分酶切反应,确定最佳酶切条件为 8U/半个 plug,37℃ 反应 15 min(图片未给出)。

### 2.3 基因组 DNA 的两次筛选和回收

以上述最佳酶切反应条件,进行了 16 个相同的酶切反应,通过脉冲电泳进行大片段 DNA 的分离。电泳结果如图 2-A 所示。从中切下 120~250 kb(a)和 250~300 kb(b)大小的胶块进行第二次电泳筛选,以去除其中夹杂的小片段。

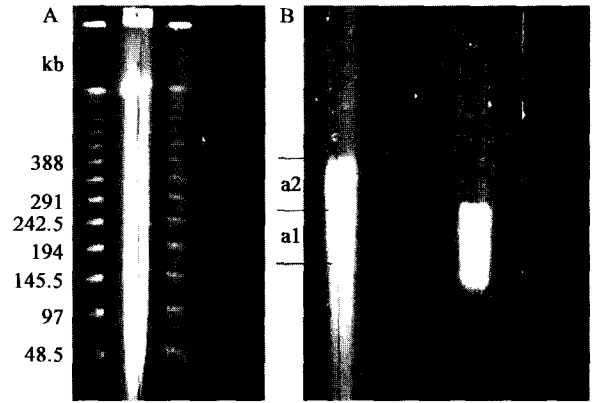


图 2 水稻 PA64S 基因组 DNA 部分酶切第一次(A)和第二次(B)片段筛选

Fig.2 The first(A) and second(B) size selections of the partial digestion products of rice PA64S genomic DNA.

M:λ ladder PFG 分子量标记  
M: λ ladder PFG marker

图 2-B 为第二次筛选的结果。图中的 a、b 即为第一次筛选时切下的胶块,a 胶块下方的 DNA 有大量拖尾,这说明第一次筛选后仍混杂有大量小于 120 kb 的 DNA 片段,这些小片段 DNA 通过第二次筛选后被分离出来。

经过两次电泳筛选后,将含有大片段 DNA 的胶块切下,由于 a 区段分离跨度比较大,被从上到下分成 a2 和 a1(DNA 的大小为 a2 > a1)。对胶块中的大片段 DNA 通过电洗脱法回收,并以已知浓度的 λ DNA 梯度估测浓度,表明 a1、a2 和 b 各区段的 DNA 浓度均在 1 ng/μL 左右。

### 2.4 文库的构建与质量检测

将大片段基因组 DNA 与载体连接,自每个连

接体系中取 2  $\mu\text{L}$  进行试转化,计数长出的白斑数目,发现其中 a1 的 2 个连接体系(a1 I、a1 II)得到的白斑数较多,故从这 2 个体系来源的白斑中各挑了 20 个检测插入片段,结果为 a1 I 体系的空载率较低,因此将 a1 I 剩余的体系全部作了转化,将所长出的白斑挑入灌注了冰冻培养基的 384 孔板中,共挑满 98 块 384 孔板计 37 632 个克隆,复制一份拷贝后存放于  $-80^{\circ}\text{C}$  冰箱中。

为了检测该文库插入片段的平均长度,从文库中随机挑取了 193 个克隆检测插入片段,图 3 为部分 BAC 克隆插入片段的检测结果。共发现 3 个空载克隆,空载率为 1.5%,其余克隆都含有插入片段,绝大部分克隆的插入片段集中在 120~159 kb 的范围之内,少量分布在 80~119 kb 和 160~179 kb 这两个区段,分布于 80 kb 以下和 180 kb 以上的克隆极少,说明整个文库的插入片段大小分布较为集中。平均插入片段为 138 kb,按水稻基因组 430 Mb 来推算,该文库覆盖全基因组达 12 倍。

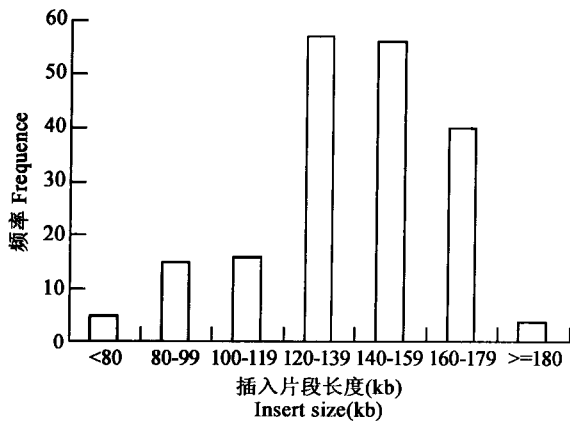


图 3 BAC 克隆插入片段检测

Fig. 3 Insert size test of BAC clones.

为了检测细胞器 DNA 的污染率,制备了全文库的高密度膜,并利用叶绿体和线粒体基因探针全文库进行了筛选,结果见表 1。全文库的线粒体污染率为 0.35%,叶绿体污染率为 1.85%,这说明细胞器的污染率很低。

为了验证文库的使用价值,以水稻品种日本晴中的单拷贝基因序列作探针,筛选了整个文库。这些探针序列分布于日本晴参考序列第 3 染色体的几个缺口附近,利用它们筛选到的 PA64S 阳性克隆有可能跨越日本晴参考序列第 3 染色体这几个缺口,从而填补这些缺口的序列。杂交结果见

表 2。一共使用了 7 个探针,每个探针得到 10~17 个阳性克隆不等,平均 12.7 个,这说明绝大多数探针都可以筛选得到克隆,并证明先前计算的该文库的 12 倍覆盖度是比较准确的。

表 1 PA64S BAC 文库中细胞器 DNA 污染检测结果

Table 1 Clone numbers of the organelle DNA in PA64S BAC library.

探针 Probe	阳性克隆个数 Number of positive clones	占总数 比例 Ratio
线粒体 Mitochondrial ( <i>atpA</i> , <i>atp9</i> , <i>cob</i> , <i>coxI</i> )	132	0.35%
叶绿体 Chloroplast ( <i>ndLA</i> , <i>rbcL</i> , <i>psbA</i> )	697	1.85%

表 2 PA64S BAC 文库中单拷贝基因探针杂交结果

Table 2 Screening results of the PA64S BAC library with single copy gene probes.

探针 Probe	引物序列 Primer sequences	阳性克隆数 Numbers of positive clones
1	f 5'-CAGAGGCAGTCAGGCAGTA-3' r 5'-ATATAAGACGGGCGTAACAT-3'	11
2	f 5'-CGCCGTAGACCCTAGGATA-3' r 5'-AATGCCTGTTTGGTGATGC-3'	10
3	f 5'-TCCGAAATGAACCTAACAG-3' r 5'-TGCTAACGGAAACCTAATC-3'	17
4	f 5'-CTTCTACTCCTGCCCGT-3' r 5'-ATGTCCAGTCTAATGCC-3'	17
5	f 5'-GCCAGGAAATGAAGGTGAG-3' r 5'-TGCTCTGCTGCAAGGGA-3'	11
6	f 5'-GACATTGTTTGGCTCTTGA-3' r 5'-TGTCAGGCTGTTTGTACTC-3'	12
7	f 5'-ATTGCCTACACTGACTGGG-3' r 5'-TTATGCCAACACCTGAAC-3'	11
平均 Average		12.7

### 3 讨论

水稻是世界上最重要的粮食作物之一。它在禾本科作物中基因组最小,遗传转化系统非常成熟,且拥有众多物理和遗传图谱信息,以及丰富的 EST 和 cDNA 数据库,因此早已成为禾本科作物基因组研究的模式植物<sup>[15-17]</sup>。

BAC 文库的构建是基因组学的基础,尤

其是在精细物理图谱的绘制及基因的定位克隆等方面。虽然经过几十年成熟的发展,但这项技术在具体实施时却仍是一项很大的挑战,在各个环节都必须非常小心,否则整个实验就将前功尽弃。具体主要包括以下几方面问题。

首先,高质量大片段 DNA 的提取尤为重要。对于植物基因组 DNA,理想的高质量 DNA 应为片段长和细胞器 DNA 污染少(动物线粒体 DNA 小,没有叶绿体,不存在此问题)的 DNA。因此在整个实验过程中都必须在冰上小心轻柔操作,防止造成 DNA 的机械性损伤和降解,同时,本实验采用水稻叶片为提取材料,在提取细胞核时必须加入 Triton X-100 以选择性地破坏线粒体和叶绿体<sup>[13]</sup>,防止造成线粒体、特别是叶绿体 DNA 的大量污染。其次,载体的质量也是 BAC 文库质量最大的制约因素之一。在文库构建之前,必须先对每批制备的载体进行测试,选取那些自连率低、连接效率高及空载率低的批次的载体。本实验中提取的基因组 DNA 和制备的载体质量都很高,为构建高质量的 BAC 文库打下了基础。

本实验选取的 PA64S 品种极具研究和利用价值,它与其他品种配制出的系列杂交水稻组合都具有明显的高产效应,PA64S 系列组合已作为两系法杂交水稻先锋组合大力推广,以 9311 为父本和 PA64S 为母本配制的超级杂交稻两优培 9(LYP9)既是我国主要推广杂交种之一,也是用来研究水稻杂种优势的主要杂交组合之一<sup>[9]</sup>,北京基因组研究所已对两优培 9 父本 9311 基因组进行了 WGS 测序<sup>[12,17]</sup>,而目前母本 PA64S 的基因组资源严重缺乏。北京基因组研究所虽也对 PA64S 基因组作了低覆盖度的 WGS 测序,但此序列既没有发表也没有对公众释放(于军,私人通信),如果没有 BAC 文库和 BAC 物理图谱帮助,这些低覆盖度 WGS 序列的组装会相当困难。PA64S 的遗传背景较为复杂,既有 *indica* 的遗传背景,又有一部分 *japonica* 和 *javanica* 的遗传背景<sup>[11,12]</sup>,因此研究它的基因组意义重大。本实验为 PA64S 构建了第一个高质量的 BAC 文库,该文库具有平均插入片段大、空载率低、覆盖度高、细胞器污染率低等优点,其可使用性已得到验证。该文库将作为公共基因组资源向研究者开放,为 PA64S 基因组研究、9311 和 PA64S 这对优良亲本间的比较分析和其上优良基因的克隆以及杂种优势分子机理的揭示等提供便利。

## 参 考 文 献

- [1] Osoegawa K, Woon Y P, Zhao B H, *et al.*. An improved approach for construction of bacterial artificial chromosome libraries[J]. *Genomics*, 1998, 52:1-8.
- [2] Shizuya H, Birren B, Kim U J, *et al.*. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector[J]. *Proc. Natl. Acad. Sci. USA*, 1992, 89:8794-8797.
- [3] Ammiraju S S J, Luo M Z, Goicoechea L J, *et al.*. 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*[J]. *Genome Res.*, 2006, 16:140-147.
- [4] Liu W, Zhao Y H, Liu Z L, *et al.*. Construction of a 7-fold BAC library and cytogenetic mapping of 10 genes in the giant panda (*Ailuropoda melanoleuca*) [J]. *BMC Genomics*, 2006, 7:294.
- [5] Luo M Z, Kim H, Kudrna D, *et al.*. Construction of a nurse shark (*Ginglymostoma cirratum*) bacterial artificial chromosome (BAC) library and a preliminary genome survey[J]. *BMC Genomics*, 2006, 7:106.
- [6] Luo M Z, Wang Y H, Frisch D, *et al.*. Melon bacterial artificial chromosome (BAC) library construction using improved methods and identification of clones linked to the locus conferring resistance to melon *Fusarium* wilt (*Fom-2*) [J]. *Genome*, 2001, 44:154-162.
- [7] 郭兆武. 高产杂交稻‘两优培九’的光合特性研究[D]. 长沙:湖南农业大学,博士学位论文,2007.
- [8] 钱立生. 两优培九及其亲本低温强光适应性特性的比较研究[D]. 合肥:安徽农业大学,硕士学位论文,2005.
- [9] Wei G, Tao Y, Liu G Z, *et al.*. A transcriptomic analysis of super hybrid rice LYP9 and its parents[J]. *Proc. Natl. Acad. Sci. USA*, 2009, 106:7695-7701.
- [10] Yu J, Wong K S G, Liu S Q, *et al.*. A comprehensive crop genome research project: the superhybrid rice genome project in China[J]. *Phil. Trans. R. Soc. B*, 2007, 362:1023-1034.
- [11] Yu J, Hu S N, Wang J, *et al.*. A draft sequence of the rice (*Oryza sativa* ssp. *indica*) genome[J]. *Chin. Sci. Bull.*, 2001, 46:1937-1942.
- [12] 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.
- [13] Luo M Z, Wing A R. An improved method for plant BAC library construction[J]. *Methods Mol. Bio.*, 2003, 236:3-20.
- [14] Zhang H B. Construction and manipulation of large-insert bacterial clone libraries-manual[M]. Texas, USA: Texas A&M University, 2000.
- [15] 张玉军. 水稻基因组序列分析及比较基因组研究[D]. 上海:中国科学院研究生院,博士学位论文,2003.
- [16] Yu J, Ni P X, Wong K S G. Comparing the whole-genome-shotgun and map-based sequences of the rice genome[J]. *Plant Sci.*, 2006, 11:387-391.
- [17] Yu J, Wang J, Lin W, *et al.*. The genomes of *Oryza sativa*: a history of duplications[J]. *PLoS Biol.*, 2005, 3:266-281.