生物化学与生物物理学报
ISSN 0582-9879 ACTA BIOCHIMICA et BIOPHYSICA SINICA 2002, **34**(3): 358 - 364 CN 31-1300/Q

Short Communication

# A High Efficient Approach Used for BAC-contig Extension of *Oryza sativa* with PCR Screening the BAC Clone Pools

HU Hao, LI Tao, MU Jie, HAN Bin *, HONG Guo-Fan *

(*National Center for Gene Research, Shanghai Institutes for Biological Sciences, the Chinese Academy of Sciences, Shanghai* 200233, *China*)

**Abstract** To extend 8 BAC contigs, which were previously located in the 56.1 —68 cM region of the chromosome 4 of the *Oryza sativa indica GuangLuAi*4, 14 pairs of primers were designed according to the terminal sequences of the existing seed BACs and were deliberately divided into 3 groups. With the 3 groups of primer mixtures, 233 pools of BAC DNA that represent 22 368 BAC clones from *O. sativa indica GuangLuAi*4 genomic library were screened. 65 positive clones corresponding to the 8 contigs were isolated and 29 clones of them were confirmed to be extended to the seed BACs by end-sequencing and fingerprinting. The protocol greatly enhanced the efficiency of the contig extension and was also superior for its specificity, sensitivity and reusability to the colony *in situ* hybridization which is a conventional method employed in contig extension and physical map construction.

**Key words** *Oryza sativa*; BAC; PCR; contig

Clone by clone strategy, which is adopted by our laboratory in sequencing chromosome 4 of *Oryza sativa indica GuangLuAi*4, is based on an extensive physical map[1 —5]. Three steps were taken to construct physical map in our laboratory[6 —10].

Firstly, anchoring by genetic markers: The 22 368 BAC clones representing the whole library were blotted on nylon membranes and colony *in situ* hybridization using genetic markers as the probes were carried out. Thus the first batch of seed BACs were anchored to the specific loci in the chromosome. Secondly, chromosome walking: colony *in situ* hybridization was further carried out using short DNA fragments in the terminal sequences of seed BACs as the probes, thus BAC contigs were extended. Thirdly, end-sequencing, fingerprinting and Southern blotting were carried out to corroborate the overlapping relationships. Extending sizes as well as the overlapping size were also determined.

Colony *in situ* hybridization is a conventional method employed in contig extension and physical map construction. But it has two major disadvantages. Firstly, it was very time-consuming. Secondly, the sensitivity of hybridization is not high enough so that some weak signals, which probably come from the low-content templates, may be ignored. Thirdly, since repeat sequences take up about 50 percent of the rice genome, the non-specific hybridization rate could be very high.

To improve the process of the contig extension, an alternative approach based on PCR screening[11 —15] was employed to extend BAC contigs in the region from 56.1 cM to 68 cM on the long arm of chromosome 4 of *Oryza sativa indica GuangLuAi*4 in the study. The results were corroborated by several methods such as fingerprinting, end-sequencing and especially the finished sequencing results.

## 1 Materials

The *Hin*dIII BAC library of the *O. sativa indica GuangLuAi*4 genome was constructed by TAO Quan-Zhou and HONG Guo-Fan *et al.* in 1994; ECL hybridization kits were from Amersham Pharmacia Biotech Inc.; r*Taq* enzyme kits and DL 2000 DNA marker were from TaKaRa Biotechnology (Dalian) Co., Ltd; primers were synthesized by BioAsia Biotechnology Co., Ltd, Shanghai; restriction enzymes were from Boehringer Mannheim; chemicals were from Sigma Chemical Co., Ltd.

## 2 Methods

Received: November 20, 2001 Accepted: January 9, 2002

This work was involved in the International Rice Genome Sequencing Project (IRGSP) and was supported by the Chinese Academy of Sciences, The Ministry of Science and Technology and Shanghai Municipal Commission of Science and Technology

Both authors contributed equally to this work

* Corresponding authors: HAN Bin: Tel, 86-21-64845260; Fax, 86-21-64825775; e-mail, bhan@newnetra.ncgr.ac.cn; HONG Guo-Fan: Tel, 86-21-64516371; Fax, 86-21-64825775; e-mail, gfhong@newnetra.ncgr.ac.cn

© 1995-2005 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.

Templates of PCR were prepared as following: 22368 BAC clones in the *Hin*dIII library are preserved on 233 wells of 96-well plates. Firstly, each BAC clone was inoculated in deep 96-well plates filled with 1.4 ml LB media with 25 mg/L chloromycetin and incubated at 37 ℃ overnight, vigorously agitated. Then the plasmids were extracted by alkaline lysis method. The product in each well was dissolved in 30 μl TE buffer. Meanwhile, 233 tubes of DNA mixtures called template pools were obtained in the similar way, each of which was equivalent to the mixture of the plasmids from a specific plate. Thus the 233 plates and the same number of tubes represented the 22 368 BACs, namely the *Hin*dIII library. They are "single templates" and "template pools" of PCR, respectively. Furthermore, before use, the 233 tubes of plasmids mixture were dispatched on three 96-well plates, one well corresponding to one tube.

Primers were designed according to the following six criteria with the aid of software GAPv4.4: (1) Primers were based on the sequences near the ends of seed BACs anchored by genetic markers. (2) The sequences of the primers should not be repetitive sequences or have highly similar sequences in the rice genome, which was affirmed by BLAST. (3) The lengths of PCR products were dispersed in a range from 200—500 bp thus length information obtained from agarose electrophoresis image could help reduce the amount of labor, yet the range was not further wider for the sake of PCR condition uniformity which directly affected the PCR efficiency. (4) As to the two primers that are in a pair, there should not be continuous base pairing that is 4 bp or longer both in the same primer and between the two primers. (5) The lengths of primers range from 18—22 bp. (6) The annealing temperature is about 52 ℃. Thus, altogether 14 pairs of primers were designed and synthesized.

The sequences of the primers are listed in Table 1.

The 14 pairs of primers were divided into 3 groups each including 3, 5, 6 pairs of specific primers. The groups were divided following 2 criteria: (1) No base pairing that is 4 bp or longer exists in the same primer group lest the false result happen or the PCR efficiency drops. (2) The lengths of PCR products in one group should be different enough to be distinguished on the agarose electrophoresis image, which help find the specific pair of primers

**Table 1 The sequences of the primers**

| Group name | Primers name | Sequences | Length of products (bp) |
|---|---|---|---|
| Group 1 | H815c01HP1F/R | 5′CTG AAT AGT GAA AAC TGA TG 3′<br>TAG TGA TAA ATC AAA CCA CG | 296 |
| | H622f05SP6F/R | 5′CTC TAT GGA AGT CAT TAA C 3′<br>CAC TCT TTA TCT CAA TAG G | 313 |
| | H123b08HP1F/R | 5′CTT ATA GCA GAG ATT TGT G 3′<br>CTA CAA TCT TGA TTT TTG TCC | 213 |
| Group 2 | B621d02SP6F/R | 5′AAC TAT TAC CAA CAA ACC C 3′<br>ACC ACA ATT CCA CAA ATA C | 367 |
| | H717b12HP1F/R | 5′CCT TTG AGA CCT AAC TAC 3′<br>CTT ATT ATC CCACAC ATA TC | 303 |
| | H622f05HP1F/R | 5′CTT AAA TTC ACT CTC TCA C 3′<br>GAA AAT GGC TCT ACT CGG | 331 |
| | H502b11SP6F/R | 5′TGG CAA TTT AGA AGG ATT C 3′<br>TGG TTA ATG TGT ATT GAT G | 247 |
| | H502b11HP1F/R | 5′TAA TCG AAG AAG TAT CCT G 3′<br>TGA TCT TCT TGG AAT CTT G | 220 |
| Group 3 | B621d02T7F/R | 5′CAA GAA CAT ATA CAC GAA G 3′<br>CGA CTG TAC TTC TTG ATT G | 343 |
| | H525c06SP6F/R | 5′GGT TAT GGC TAA AGA ATA ATG 3′<br>GTC AAG AGA GGT TAA TTA TC | 341 |
| | H525c06HP1F/R | 5′TTG TAG ACA TTG TAC TAG G 3′<br>GTT AGT ATG GTT AGT TCA G | 314 |
| | B222c05SP6F/R | 5′CAC ACA AAA CTA TTG CTA TC 3′<br>GAC AAA TCA AAC ACA CAA TAC | 320 |
| | B222c05HP1F/R | 5′GAG AGA AAG TGA ACA AAA G 3′<br>CAC CAC TAT CTC TAA TTT ATC | 339 |
| | H123b08SP6F/R | 5′AGA TTG TAG ACT GTT GAA G 3′<br>GTG ACA TAT TAT AGA ATC CAA G | 292 |

© 1995-2005 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.

corresponding to the PCR products. 2 % agarose was used for best resolution.

The PCR procedure was as following: Denature: (1)95 ℃, 20 s; First round: (2)95 ℃, 20 s; (3)54 ℃, 30 s; (4)72 ℃, 30 s; (5)go to (2)for 5 cycles; Second round: (6)95 ℃, 20 s; (7)52 ℃, 30 s; (8)72 ℃, 30 s; (9)go to (6)for 25 cycles; Store: (10)4 ℃ store. The annealing temperature of first round was 2 ℃ higher than that of second round so that false positive rate could be reduced. The PCR system was premixed as indicated in Table 2. The most important in the Table 2 is the concentration of the templates. Too high concentration would lead to complex false positive results.

PCR reactions were carried out with a 3-step strategy. PCR step 1: To determine the relation between the primer groups and the template pools. In this step, primer group were reacted with template pools. For example, the primer group 1, which contained 3 pairs of primers, reacted with the three 96-well plates representing the whole 233 template

**Table 2 The components of PCR reaction**

| Templates (approx. 10 mg/L) | 1 μl |
|---|---|
| Primers (10 μmol/L each) | 0.5 μl |
| dNTPs (2.5 mmol/L each) | 4 μl |
| $MgCl_2$(25 mmol/L) | 3 μl |
| r *Taq* | 2 u |
| 10 ×buffer | 5 μl |
| Sterilized double distilled water | add to 50 μl |

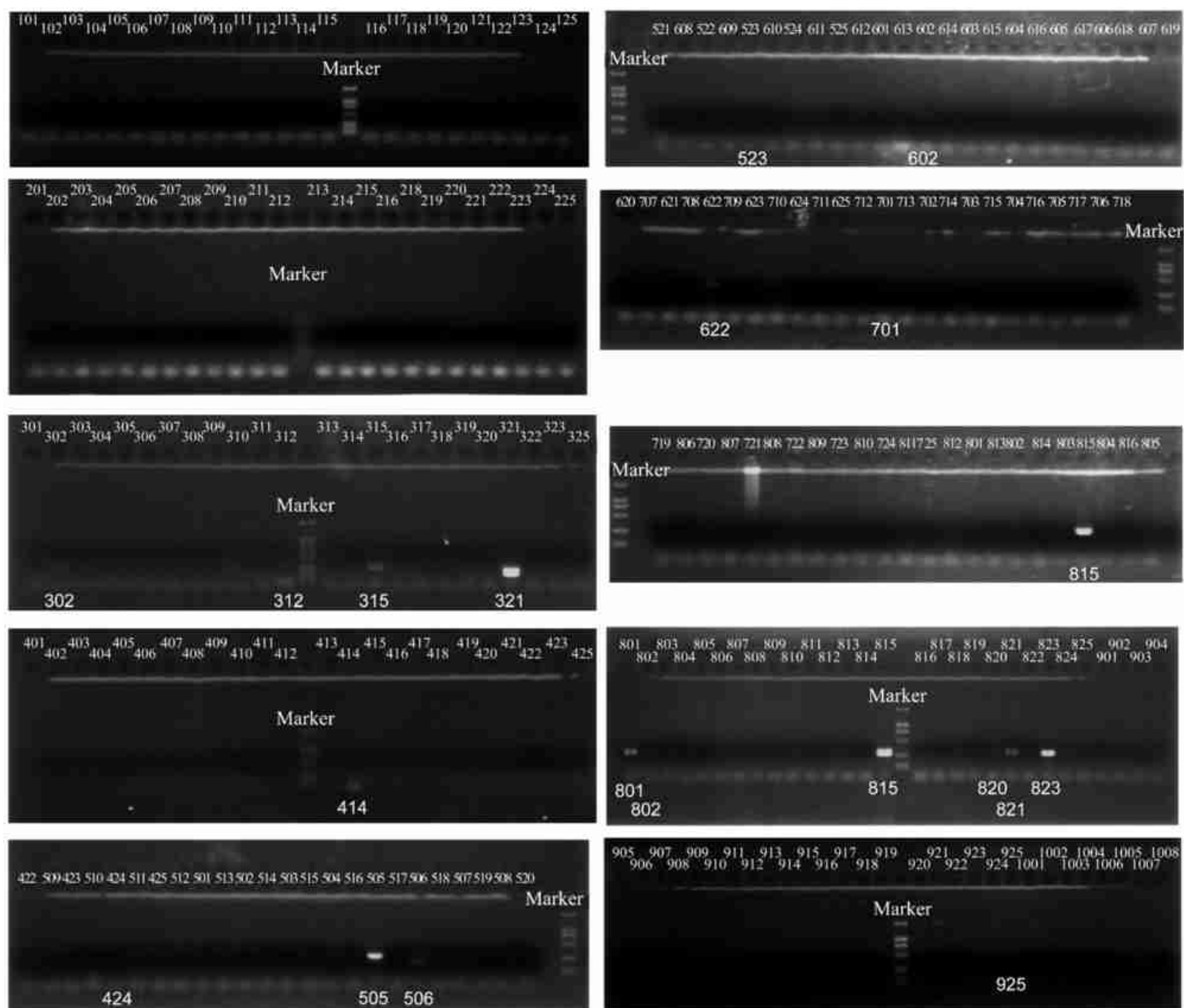


**Fig. 1 Gel electrophoresis of the pool-PCR results using Group1 primers**

There are three pairs of primers(H815c01HP1F/R, H622f05SP6F/R, H123b08HP1F/R) in primer Group 1. The product of H123b08HP1F/R was shorter than 250 bp(213 bp) while the products of the other two pair were longer than 250 bp(296 bp and 331 bp). Since one band of the DL2000 marker represents 250 bp, each positive result was attributed to H123b08HP1F/R(when <250 bp) or the other two primer pairs(when >250 bp). Such an attribution reduced the amount of labor in step 2. The positive results belonging to H123b08HP1F/R: H321, H506; The positive results belonging to either H815c01HP1F/R or H622f05SP6F/R : H302, H505, H602, H801, H821, H823; The false results: H312, H315, H321, H414, H424, H506, H523, H622, H701, H815, H802, H820, H925.

© 1995-2005 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.

pools. Only with one 4-block thermal cycler and in 80 minutes, information was obtained about which template pools had positive clones corresponding to the certain primer group (Fig. 1).

Moreover, since the length of products had distinguishable differences, positive results could be attributed to fewer, such as 2 specific pairs of primers within the primer group 1. PCR step 2: To determine the relation between the specific pair of primers and the template pool(s). In this step, every positive template pool was attributed to a certain pair of primers in the primer group (Fig. 2) by PCR.

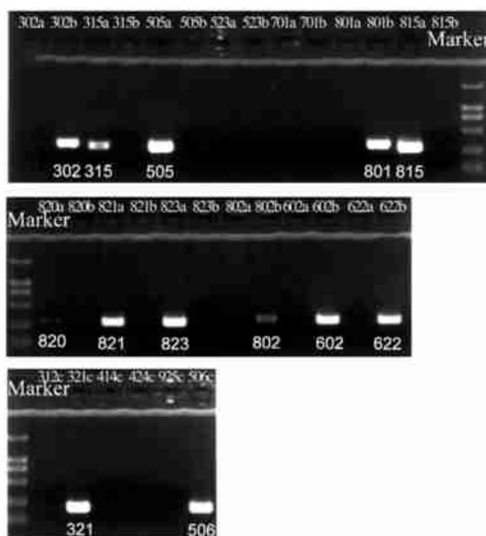


**Fig. 2 Part of the results from PCR step 2**

Each positive result in Fig. 1 was picked out and reacted with one of the pair(s) of primers to which it was attributed in PCR step 1. Thus each positive result (representing 96 wells on a certain plate) was attributed to a single pair of primers.

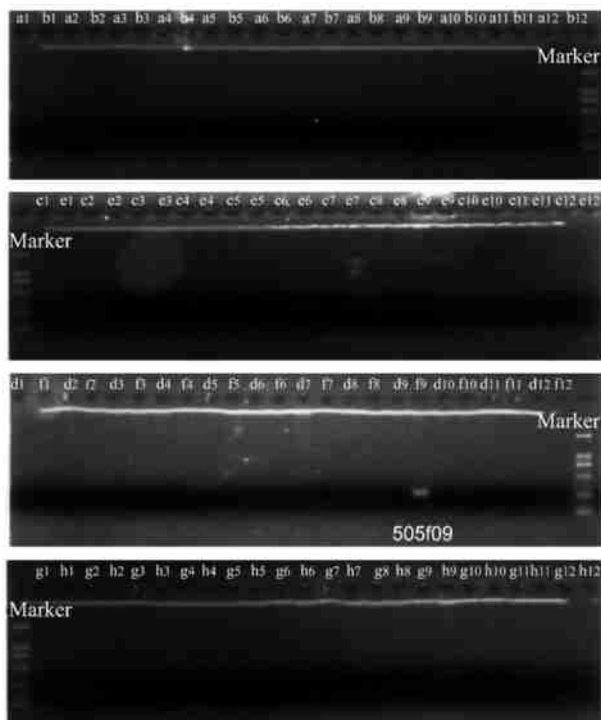


**Fig. 3 Part of the results from step 3**

After PCR step 1 and step 2, the relationship between single plate and single pair of primers was established. In step 3 each one of the 96 BAC templates that were on a positive plate was reacted with the pair of primers to which the positive plate was attributed. A positive BAC clone H505f09 was picked out.

PCR step 3: To determine the positive BAC(s) for the specific pair of primers. In this step, the specific pair of primers reacted with the certain 96-well plate templates corresponding to the positive template pool, to determine the positive BAC template (Fig. 3).

Positive results obtained from PCR were subjected to mature techniques in our lab for confirmation, such as end-sequencing, finger-printing, Southern blotting (see Fig. 4).

Overlapping size between the seed BAC and the positive BAC was estimated based on fingerprinting and Southern blotting results and accurately corroborated by end sequencing. Clones that had overlapping size less than 25 % but more than 3 kb were selected for sequencing.

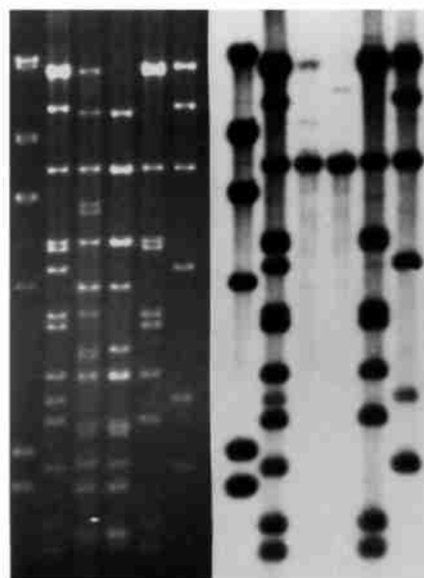

**Fig. 4 One of the results of fingerprinting (A) and hybridization [(B), seed BAC as the probe]**

1, λ/*Hin*dIII marker; 2, H502b11 (the seed BAC); 3, H123f08; 4, H204e02; 5, H615d10; 6, H712d03.

## 3 Results

65 BAC clones were identified by using pool-PCR screening to be located in the region from 56.1 – 68 cM on the chromosome 4 of *O. sativa indica GuangLuAi*4. 29 BAC clones of them were further identified to be located on the 8 contigs by Southern hybridization and BAC end-sequencing analysis. The relationship between the identified BACs and the seed BACs are shown in the Table 3.

© 1995-2005 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.

**Table 3 The relation between the seed BACs and the positive BACs in the 56.1–68 cM region of chromosome 4 of** O. sativa indica GuangLuAi**4**

| Seed BACs (length and location) | Positive BACs | Overlapping length (kb) (appromix.) | Extension length (kb) (appromix.) | Direction of seed BACs |
|---|---|---|---|---|
| H123b08 (99 kb, 56.1 cM) | H506a08 | 70 | 20 | HP1 |
| | H321a05 | 30 | 80 | HP1 |
| | H314f01 | 60 | 9 | SP6 |
| H815c01 (109 kb, 67.5 cM) | H505f09 | 30 | 70 | HP1 |
| | H821d12 | 109 | <2 | HP1 |
| | H823e10 | 109 | 5 | HP1 |
| | H823c12 | 109 | 5 | HP1 |
| H622f05 (71 kb, 58.9 cM) | H602e02 | 45 | 3 | SP6 |
| | H302b04 | 45 | 10 | SP6 |
| | H801e06 | 70 | 3 | SP6 |
| | H414h05 | 35 | 30 | HP1 |
| | H617h08 | 50 | 10 | HP1 |
| H717b12 (121 kb, 57.5 cM) | H414e10 | >90 | <2 | HP1 |
| | H417b06 | >60 | <2 | HP1 |
| B621d02 (51 kb, 66.4 cM) | H903e04 | 42 | 30 | SP6 |
| | H218a01 | >45 | 10 | SP6 |
| | H213c02 | >40 | 10 | T7 |
| | H216a02 | 15 | >50 | T7 |
| | H218a01 | >45 | 10 | T7 |
| | H919c01 | 8 | <2 | T7 |
| H502b11 (114 kb, 56.1 cM) | H123f08 | 0.4 | 115 | SP6 |
| | H204e02 | 0.5 | 82 | SP6 |
| | H615d10 | 53 | <2 | SP6 |
| | H712d03 | 45 | <2 | HP1 |
| H525c06 (99 kb, 65.0 cM) | H118b03 | 60 | 15 | HP1 |
| | H225d01 | 55 | 10 | HP1 |
| B222c05 (40 kb, 62.1 cM) | H206a12 | 40 | 30 | SP6 |
| | H125f04 | 20 | >65 | HP1 |
| | H206a12 | 40 | 30 | HP1 |

Note: HP1/T7 and SP6 are not equal to the 5′and 3′ends of the DNA, but the specific sequences in the BAC vector which determine the direction of the inserts.

In total, the 8 contigs have been extended 510 kb from both directions. One of the identified BAC clones H321a05 was selected and sequenced and the sequencing results supported the conclusion in this work (Fig. 5).

## 4 Discussion

Pool-PCR has 4 advantages over traditional method: (1) High efficiency Pooled templates and primer groups reduced the time of contig extension greatly. To finish the elongation from 16 BAC ends, a proficient technician need only 2–3 weeks using pool-PCR while several months will be needed by traditional methods. (2) High sensitivity Following the protocol in this article, all seed BAC clones in the physical map that has been identified by traditional methods have been picked out without exception. Furthermore, PCR consumed by far less DNA templates than end-sequencing, fingerprinting and Southern blotting. Through a series of tests we have found the least yet still efficient template dosage as little as 1 ng. Too much template could cause false positive results. (3) Reusability Once the BAC template pools has been prepared, they are enough for at least 60 times of chromosome walking from 14 BAC ends. In most situations, it is well enough for the whole chromosome walking. (4) Specificity Some information from the electrophoresis images could help us to distinguish the false positive results from the real ones thus the specificity and readability were greatly enhanced compared to the colony *in situ* hybridization. Firstly, the lengths of the false positive products were different from what be supposed in most cases (see Fig. 1). Secondly, false positive results due to cross-contamination of templates or homologous sequences always gave out weaker signals. Thirdly, since the cross-contamination often happened in the wells near the real positive ones. If on the electrophoresis image a strong signal was followed by a series of weaker signals especially in a declining pattern, conclusion could be made with confidence that the strongest

© 1995-2005 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.

signal represented the real positive result.

**Fig. 5 The BAC contig map of the chromosome 4 of** O. sativa indica GuangLuAi**4 (56.1—68 cM)**

To make the protocol more efficient, we also tried to use the bacteria in the BAC library directly as the templates for PCR. In 8 times of test in which all reaction conditions were the same, only 2—3 times had we got positive results. It seemed that the sensitivity and repeatability were greatly reduced compared with that DNA itself acting as the templates. Further experiments will be carried out to optimize the reaction conditions for bacterial PCR.

However, pool-PCR also has its disadvantages. From a lot of experimental results, we can see PCR is a very sensitive and complex reaction conforming to Chaos theory, somewhat like explosion process. Small disturbance can grow more and more violent until at last a negative clone could give out a seemly positive result. Such disturbance comes from 2 resources: (1) "Agglomeration" among primers We have reduced such disturbance to a tolerable extent through computer-aided primer segregation (See Methods). (2) Cross-contamination of templates PCR is a very sensitive reaction that even a single copy of template can be identified. We've developed a set of strict rules to minimize cross-contamination, from plasmid extraction to preservation, and to PCR reaction. Such a set of rules can reduce false positive results greatly. In most cases, false positive results can be distinguished from real positive results for an experienced researcher.

## References


1 Hong GF. Rice *Genome Project*, Shanghai: Shanghai Scientific and Technical Publishers, 1999

2 Shimamoto K. The molecular biology of rce. *Science*, 1995, **270** (5243): 1772—1773

3 Tao QZ, Zhao HY, Qiu LF, Hong GF. Construction of a full bacterial artificial chromosome (BAC) library of *Oryza sativa* genome. *Cell Res*, 1994, **4**: 127—133

4 Tao QZ, Qian YM, Zhao HY, Yu SL, Qiu LF, Wu BQ, Zhu J *et al*. Construction of *Oryza sativa* genome contigs by fingerprint strategy. *Cell Res*, 1995, **5**: 263—271

5 Hong GF, Qian YM, Yu SL, Hu X, Zhu J, Tao WH, Li W *et al*. A 120 kilobase resolution contig map of the rice genome. *DNA Seq*, 1997, **7**(6): 319—335

6 McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A, Wallis J, Sekhon M *et al*. A physical map of the human genome. *Nature*, 2001, **409**(6822): 934—941

7 Zhang HB, Wing RA. Physical mapping of the rice genome with BACs. *Plant Mol Biol*, 1997, **35**(1-2): 115—127

8 Hong GF. A rapid and accurate strategy for rice contig map construction by combination of fingerprinting and hybridization. *Plant Mol Biol*, 1997, **35**(1-2): 129—133

9 Qin S, Nowak NJ, Zhang J, Sait SN, Mayers PG, Higgins MJ, Cheng Y *et al*. A high-resolution physical map of human chromosome 11. *Proc Natl Acad Sci USA*, 1996, **93**(7): 3149—3154

10 Chen ZH, Zhou B, Han B, Qian YM, Hong GF. Structural analysis of a gene cluster encoding two cationic and three anionic peroxidases from rice chromosome 4. *Acta Biochim Biophys Sin*, 2001, **33**(2): 163—172
(引自: 生物化学与生物物理学报)

11 Aburatani H, Stanton VP Jr, Housman DE. High-resolution


© 1995-2005 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.

physical mapping by combined Alu-hybridization / PCR screening: Construction of a yeast artificial chromosome map covering 31 centimorgan in 3p21-p14. *Proc Natl Acad Sci USA*, 1996, **93**(9): 4474–4479

12 Munroe DJ, Loebbert R, Bric E, Whitton T, Prawitt D, Vu D, Buckler A *et al*. Systematic screening of an arrayed cDNA library by PCR. *Proc Natl Acad Sci USA*, 1995, **92**(6): 2209–2213

13 D'Esposito M, Mazzarella R, Pengue G, Jones C, D'Urso M, Schlessinger D. PCR-based immortalization and screening of hierarchical pools of cDNAs. *Nucleic Acids Res*, 1994, **22**(22): 4806–4809

14 Asakawa S, Abe I, Kudoh Y, Kishi N, Wang Y, Kubota R, Kudoh J *et al*. Human BAC library: Construction and rapid screening. *Gene*, 1997, **191**(1): 69–79

15 Yoshida K, Strathmann MP, Mayeda CA, Martin CH, Palazzolo MJ. A simple and efficient method for constructing high resolution physical maps. *Nucleic Acids Res*, 1993, **21**(15): 3553–3562

# 一种使用混合 PCR 筛选技术高效延伸水稻 BAC-重叠群的方法

胡 昊 李 滔 穆 洁 韩 斌 * 洪国藩 *

（中国科学院上海生命科学研究院国家基因研究中心，上海 200233）

**摘要** 使用“克隆连克隆(clone by clone)”战略进行水稻基因组测序需要依赖于构建好的基因组物理图。工作着眼于水稻籼稻广陆矮 4 号(*Oryza sativa indica GuangLuAi4*)第四号染色体长臂上 56.1～68 cM 的区域，采用 PCR 方法筛选 BAC 全库来延伸重叠群，构建物理图。通过参照特异遗传探针定位的 BAC 克隆(seed BAC)末端序列设计了 14 对引物，按特定规则分成 3 组，分别以代表水稻 BAC 库(共 22 368 个 BAC)的 233 个 BAC pool 为模板进行 PCR 反应，一共获得了 65 个阳性 BAC 克隆，通过末端测序、酶切杂交等方法确定了其中 29 个 BAC 克隆作为有效延伸的克隆，延伸了 8 个重叠群。通过酶切杂交、末端测序等方法还获知阳性 BAC 的延伸方向、延伸长度以及与 seed BAC 之间的重叠长度。8 个重叠群总的延伸长度达到 510 kb。与实验室原用于作物理图的其他方法如指纹图、点杂交等相比，该方法有高效率、高灵敏度、专一性好、可重复使用等优点。创新之处在于通过引物的合理分组和 PCR 实验条件的改进降低了假阳性和假阴性率。

**关键词** 水稻；细菌人工染色体(BAC)；PCR；重叠群

收稿日期：2001-11-20 接受日期：2002-01-09

水稻基因组测序国际合作项目(IRGSP)参与项目，中国科学院、国家科技部和上海市科委资助项目

为并列第一作者

*联系人：韩斌：Tel，021-64845260；Fax，021-64825775；e-mail，bhan@newnetra.ncgr.ac.cn；洪国藩，Tel，021-64516371；Fax，021-64825775；e-mail：gfhong@newnetra.ncgr.ac.cn

© 1995-2005 Tsinghua Tongfang Optical Disc Co., Ltd. All rights reserved.