Cell Research(1995), **5,** 263–271

Construction of *Oryza sativa* genome contigs by fingerprint strategy

TAO QUANZHOU[#*], YUEMIN QIAN[#*], HAIYING ZHAO[*], SHULIANG YU[*], LONGFANG QIU[*], BOQIAN WU[*], JIA ZHU[*], DI YU[*], XIAOHUI LIU[*], GUOFAN HONG[**1]

[*] National Center for Gene Research,Chinese Academy of Sciences, Shanghai 200233, China.

[**] Shanghai Institute of Biochemistry,Chinese Academy of Sciences, Shanghai 200031, China.

ABSTRACT

We described the construction of BAC contigs of the genome of a indica variety of *Oryza sativa,* Guang Lu Ai 4. An entire representative (sixfold coverage of rice chromosomes) and genetically stable BAC library of rice genome constructed in this lab has been systematically analysed by restriction enzyme fragmentation and polyacrylamide gel electrophoresis. And all the images thus obtained were subject to image-processing, which consisted of preliminary location of bands, cooperative tracking of lanes by correlation of adjacent bands, a precise densitometric pass, alignment at the marker bands with the standard, optional interactive editing, and normalization of the accepted bands. The contigs were generated based on the Computer Software specially designed for genome mapping. The number of contigs with 600 kb in length on average was 464; of contigs with 1000 kb in length on average was 107; of contigs with 1500 kb in length on average was

1. Corresponding author

#. First authors

Abbreviations: BAC = bacterial artificial chromosome; Solution I = 50 m*M* glucose, 10 m*M* EDTA, 25 m*M* Tris-HCl, pH 8.0; Solution II = 0.2 N NaOH, 1.0 % SDS; Solution III = 60 m*M* 5*M* potassium acetate, 28.5 ml glacial acetic acid, 11.5 ml water, pH 4.8; TE = 10 m*M* Tris-HCl, 1 m*M* EDTA, pH 8.0; Multi-Core buffer(10 ×) = 250 mM Tris acetate, pH 7.8 (at 25 ℃), 1 M potassium acetate, 100 m*M* magnesium acetate, 10 m*M* DTT.

23. Therefore, all the contigs we have obtained amounted up to 420 megabases in length. Considering the size of rice genome (430 megabased), the contigs generated in this lab have covered nearly 98 % of the rice genome. We are now in the process of mapping the contigs to chromosomes.

Key words: *Contig, genome, indica, oryza sativa, BAC.*

INTRODUCTION

Rice is one of the most important food crops in the world. It serves as staple food for nearly half of the world population, who are largely living in developing countries. Rice has now also become a model plant among the cereals for molecular genetic studies. It is a diploid with n=12 chromosomes, and has the genome of approx, 4.3×10^8 base pairs (bp) in length, the smallest genome of any monocots known[1]. Rice has a large germplasm collection of more than 120,000 accessions worldwide, can be regenerated from protoplasts[2] and has a high degree of transformation efficiency relative to other cereal species[3].

In recent years rice genome is intensively studied, and a striking progress has been made. Using the restriction fragment length polymorphism (RFLP), McCouch et al[4] successfully constructed the RFLP map in rice, more recently, a high resolution genetic linkage map of rice was constructed[5] , where 1383 mixed DNA markers including genes were mapping on 12 linkage groups with the markers at average interval of 300bp. Genetic maps are very useful, especially for identification and evaluation of genes of useful traits.

We have been focusing on the construction of a contig map for the rice genome. The contig map is of tremendous importance for the following two reasons: (1) in fact, a contig map is an ordered library of cloned DNA fragments that covers all of the genome. Therefore, when adequate biological information has been made available, directly from the contig map could be obtained genes and /or DNA sequences of interest, which could then be further studied at the desired levels for the purpose of either improving the quality and yield of rice, or of our better understanding of the living phenomena in the plant kingdom; (2) The ultimate goal of a genome program is to determine the entire nucleotide sequence of the genome to unveil the genetic mysteries of an organism at the molecular level. Sequencing one by one all the ordered cloned DNA fragments of the contig map fulfils the task.

The purpose of this report is to describe the construction of contigs for the rice genome, based on the DNA restriction fingerprinting technique developed by Sulston et al[6, 7] with modifications in enzymatic reactions designed for fingerprinting analysis. A representative[8] and genetically stable BAC library of the genome of rice (*Oryza sativa)* Guang Lu Ai 4 was used throughout this work

MATERIALS AND METHODS

Materials

A representative[8] and genetically stable (data not published) BAC library of the genome of rice (Oryza sativa) Guang Lu Ai 4 was constructed in this Center. Individual clones were kept in stock medium in 96-well microtiter plates at -70 ℃. Enzymes HandIII, HaeIII, Sau3AI, AMV reverse transcriptase and RNase were purchased from Sigma. Tryptone and Yeast extract were the Oxide Products. Tris base, acrylamide, bis-acrylamide and TEMED were from Sigma. DNA sequencing apparatus (Sequi-GenII) was from BioRad. The computer software for genome mapping by fingerprinting was a kind gift from DR. Alan Coulson of the Sanger Center of Cambridge, UK. Patterns analysis was performed on UNIX Computer Working Station imported from the USA.

BAC DNA preparation

5 μl of each individual BAC stock was taken from the microtiter plates, and was inoculated into 5 ml of LB medium containing 12.5 μg/ml of chloramphenicol. The BACs were incubated at 37 ℃ overnight with rotating at speed of 200 rpm. the culture was centrifuged at 600 rpm for 10 min. The cell pellet was resuspended in 0.2 ml Solution I, to which 0.4 ml of solution II was added. After well mixing and leaving on ice for 5 min, 0.3 ml of Solution III was added and gently vortexed. The mixture was freezed at -70 ℃ for 30 min , and then was left at room temperature for slow thrawing. The solution was centrifuged for 15 min in a microfuge (12,000 g), 0.75 ml of the supernatant was carefully removed and transferred into a clean microfuge tube, to which 0.45 ml of isopropanol was added. After thoroughly mixing, the solution was placed at -70 ℃ for 30 min, then was left to warm to room temperature. The DNA was pelleted by centrifugation in a microfuge for 5 min. DNA thus obtained was rinsed with lml of cold 70 % ethanol, and dried, and dissolved in 40 μl of TE, from which 3 μl was taken for fingerprint analysis.

Mapping gel

The mapping strategy adopted in this work was based on that developed by Sulston et al[6, 7, 9] with modifications. The original multistep enzymatic reactions for generating ^{32}P labeled DNA fragmemts for fingerprint analysis were simplified by this lab into a single one, thus greatly shortening the time required for each cycle of data analysis, and enhancing the overall efficiency of the strategy for contig map construction. 3 μl (50-100ng) of DNA was placed in a 0.5 ml microfuge tube on ice, to which was added the equal volume of the mixture, that consists of 128 μl of water, 39 μl of Multi-core buffer, 4 μl of HindIII (50 U /ml), 4 μl of HaeIII (50 U/ μl), 5 μl of ddGTP (0.5 m*M*), 5 μl of AMV reverse transcriptase (10 U/μl) and 4 μl of α -^{32}P-dATP (800 Ci /m*M*). The reaction mixture was incubated at 37 ℃ for 1.5 h, to stop reaction, 3 μl of the dye, containing Xylane cyanol FF (0.1 % w /v) bromophenol blue (0.1 % w /v) and EDTA (0.3 % w /v) was added. All the components of the dye were dissolved in de-ionized formamide. Fractionation of the resulting fragment was performed on a 8 *M* urea /4.0 % denaturing sequencing gel. The markers DNA, which ran alongside with analyzed DNA fragments, was the Sau3AI digest products of lambda DNA, that were end labeled by α -^{32}P- dATP. The gel was run for 110 min at 85 W until the bromophenol blue is about 4 cm from the bottom of the gel. The gel was then dried at 90 ℃ in a vacuum gel drier, and was autoradiographed for 48 h without intensifying screen. On average, there were 30 discrete bands generated per BAC clone for fingerprint analysis.

Image analysis

The software in ANSI-C programming language is currently run on a Sun sparcstation 10 with SunOS4.1.3 operating system and SGI indigo 2 with IRIX 5.2 System. The data is entered by

Scanning the autorad film and interpreted by an image-processing system. Graphics workstations are used for viewing, editing and analyzing the processed data. Scanning on the Sharp JX-610 transparent. Scanner is controlled by the program PhotoStyler. The image data is transfered to a SGI indigo 2 workstation. The program IMAGE loads the scanned mapping gel into memory, analyses the image, finds lanes, extracts bands and corrects for gel distortion. Accepted bands are automatically digitized, normalized and written to the database. The program Mapsub reads and matches every clone of a new subset against clones of the second subset. Matching involves counting within the preset tolerance the number of bands which span over the overlapping regions of clones. The program CONTIG is used for placing clones in contig. We depend on human judgement to decide the extents of overlap when contigs are being extended or joined. The verification of exact of overlaps was performed either by examination of the MAPSUB's output file MAP. OUT or by visual inspection of the films.

```
20510-y8801(24b, 0)


        15 matches80500-q1062      (20b, 391)  6.1e-12  18  9d
        13 matches70280-w15442     (18b, 391)  3.6e-10  19  2
        15 matches60090-l3845      (25b, 391)  6.1e-10  19  5
        13 matches10360-t14497     (20b, 391)  2.7e-09  22  3
        13 matches10581-t6884      (20b, 391)  2.7e-09  22  4

20510-y8803 (43b, 0)


       23 matches 70040-w18870  (32b, 372)  7.8e-12  27  20b
       25 matches 10320-t17519  (37b, 372)  8.5e-12  14  4
       18 matches 50350-b17583  (27b, 372)  1.0e-08  16  5
       18 matches 50350-b17585  (31b, 372)  23e-07   15  6
       16 matches 50342-b17582  (27b, 372)  7.4e-07  0   3
       11 matches 70420-w2578   (18b,0)     5.7e-05  0   6
       15 matches 60200-l1297   (33b,817)   1.1e-04  18  5
```

Fig 1. Showing an example of one of the output forms.
From left: incoming clones, matching clones, probability of match
Clearly, y8801 was internal in contig 391, and y8803 in contig 372.

RESULTS

We have analyzed the entire BAC library consisting 19660 BAC clones with average insert DNA of 120 kb in length, using the fingerprinting strategy developed by Sulston et al[6, 7, 9] with modifications in enzymatic reactions. Of 19660 BAC clones 603 were found to be the repeats. Contig with lengths ranged from 600-1500 kb were obtained. The proportion of contigs with differing lengths can be seen in Fig 3.

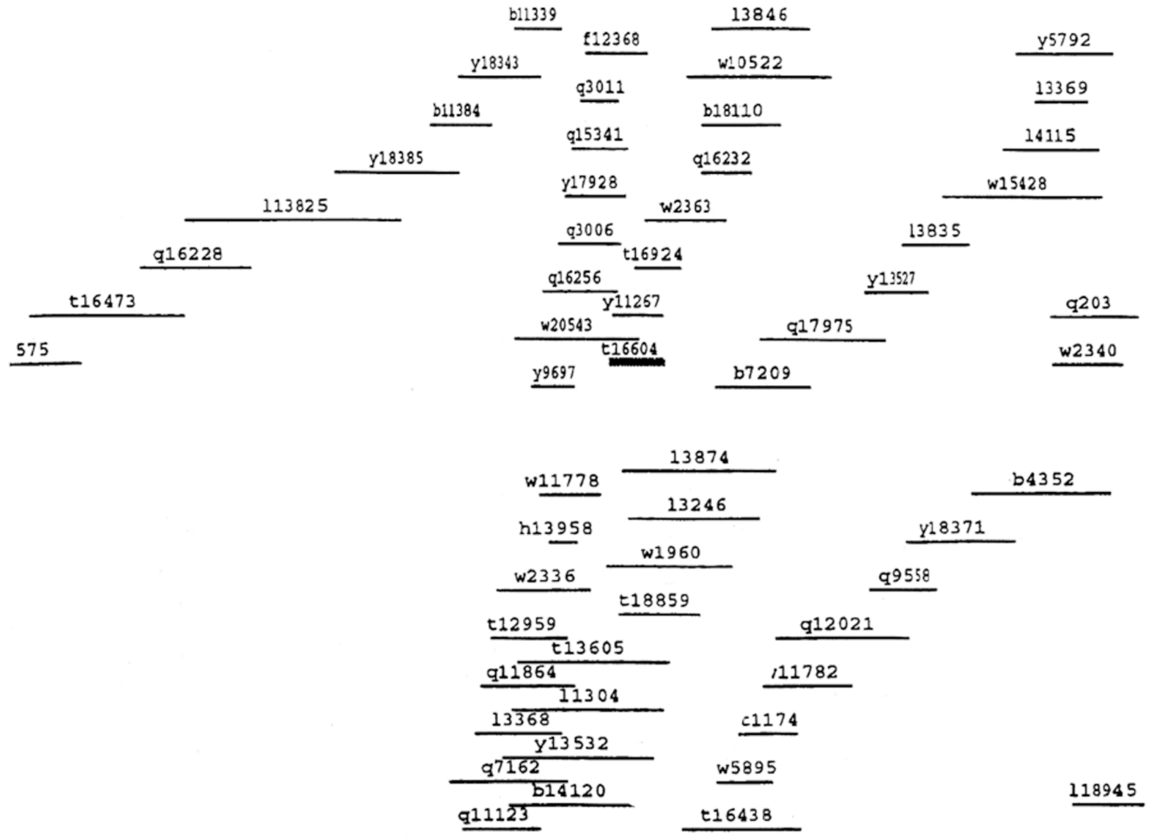


Fig 2. Showing an example of one of the Contigs, where the array of rice DNA fragments, the degree of redundency of clones, and the extent, to which DNAs overlap, can be clearly seen. Lines represent rice DNA fragments, figures and letters above lines were names for BAC clones.

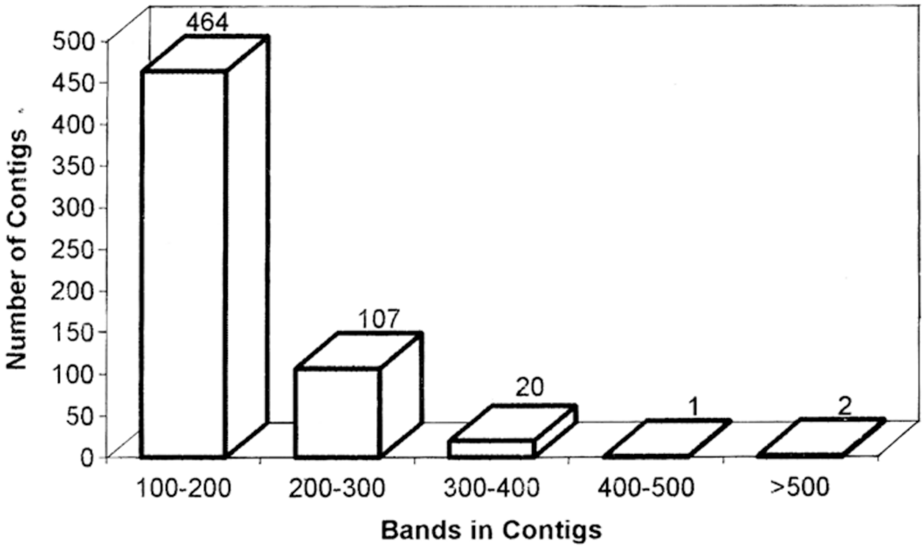


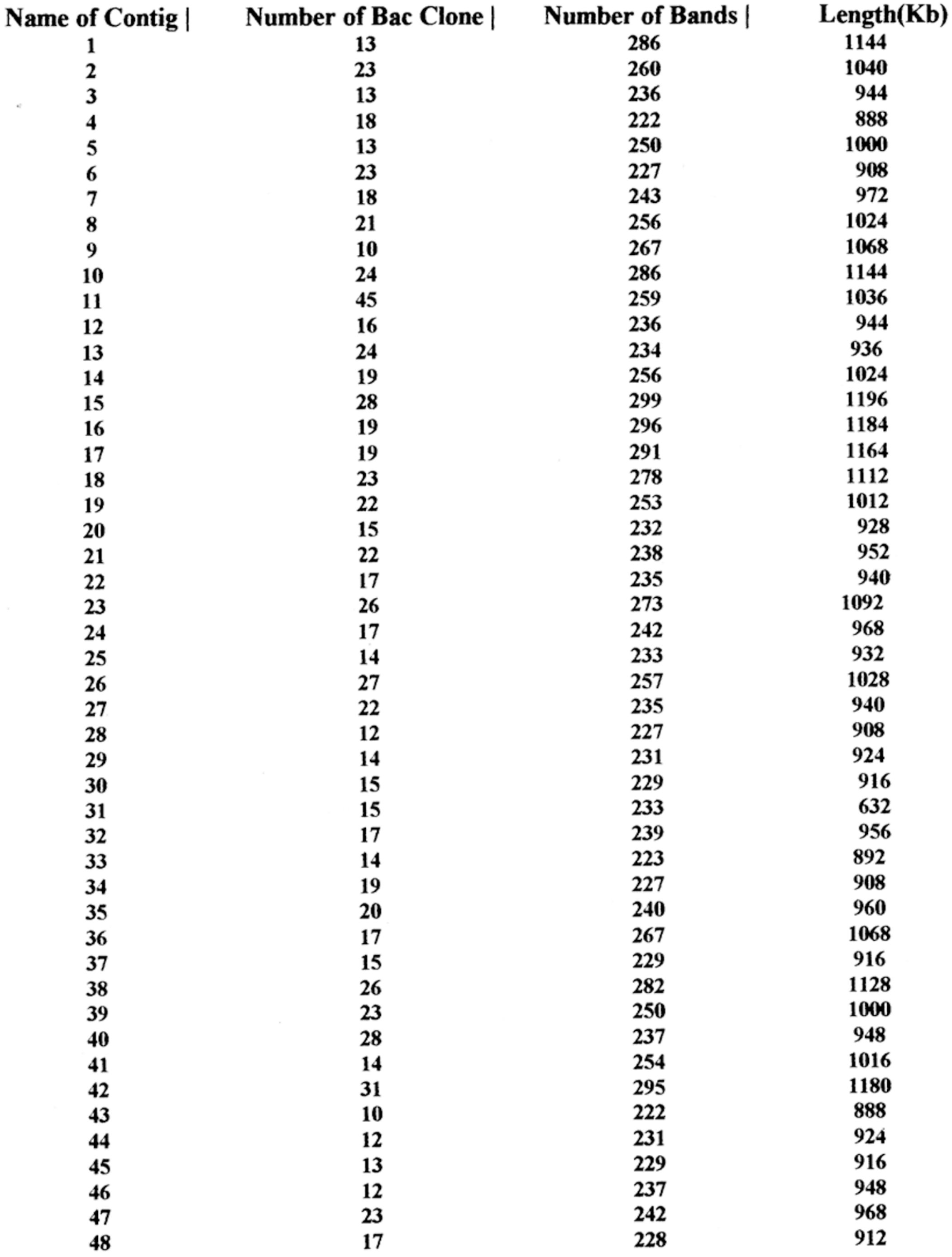
Fig 3. The proportion of contigs with differing lengths in our experiments, the average number of bands generated from single BAC clones was 30. Therefore, the length of a contig in Kb could be calculated by the equation, length of a contig=number of bands/30 × 120 kb.

From the equation in Fig 3, the total length of contigs in kb could be calculated. We have obtained 464 clones with average number of bands of 150, 107 clones with bands of 250 and 23 clones with bands of 400, i.e. they were 600 kb × 464=278400 kb, 1000kb × 107=107000 kb, 1500 kb × 23=34500 kb respectively. The entire length of all contigs we have obtained was therefore 420 megabase, which covers nearly 98 % of the rice genome (430 megabases). Tab 1 showed the details of same major contigs.

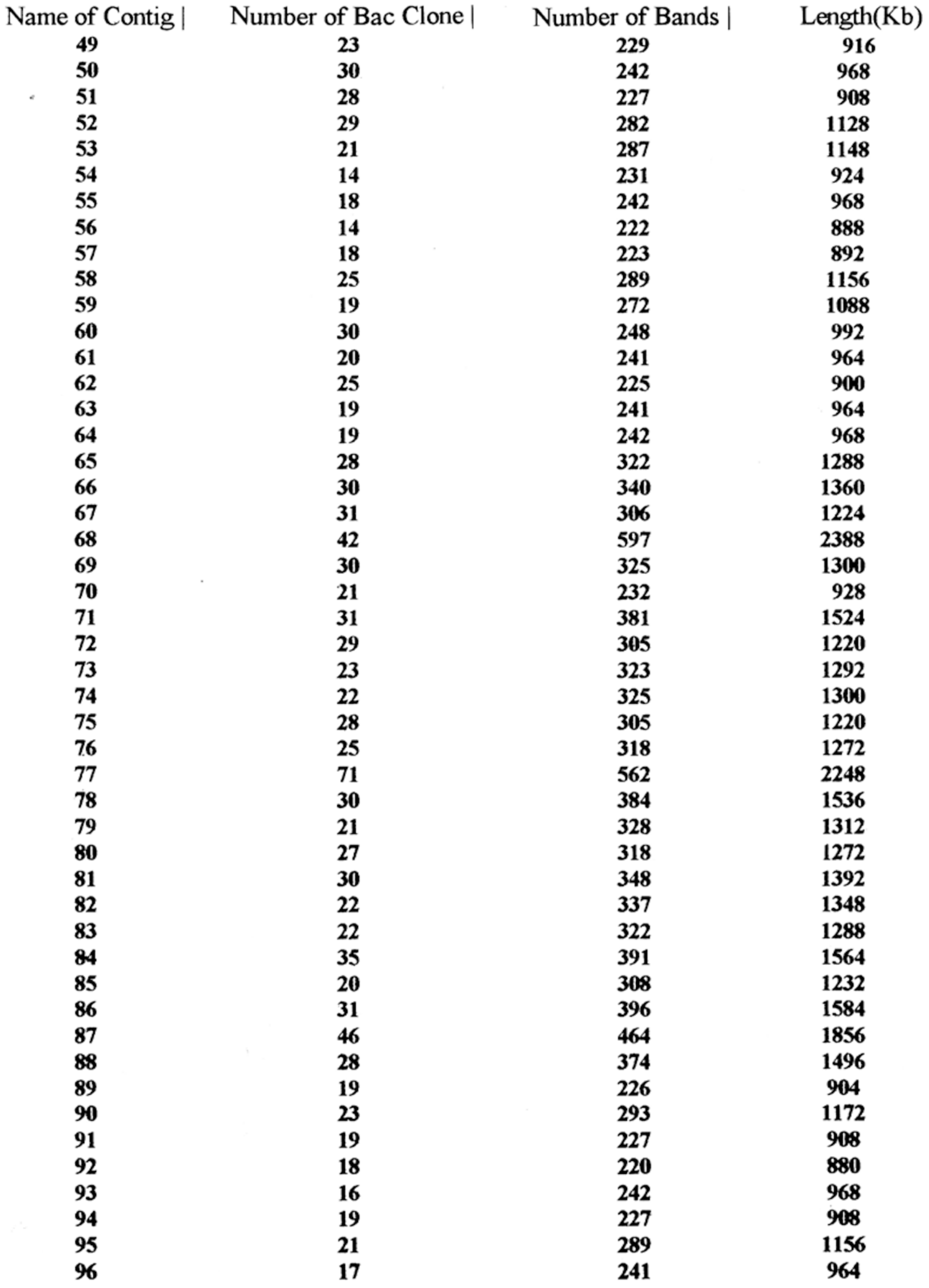
DISCUSSION

This was part of our project for contig map construction for the rice genome. In order to obtain accurate overlappings in contigs, we built up the contigs manually with one BAC clone added at a time in the similar manner as in the project for contig mapping for the C. *elegance* genome[6, 7, 9] . The use of combination of BAC-fingerprint strategy for contig construction has many advantages. It is most efficient to accumulate contigs; it produces contig maps with much fewer gaps than produced by using the strategy relying on genetic map; with the relatively smaller DNA insert in BAC, the contig maps produced have higher resolution, which will

Tab.1 Details of some major contigs.



Name of Contig |	Number of Bac Clone |	Number of Bands |	Length(Kb)
1	13	286	1144
2	23	260	1040
3	13	236	944
4	18	222	888
5	13	250	1000
6	23	227	908
7	18	243	972
8	21	256	1024
9	10	267	1068
10	24	286	1144
11	45	259	1036
12	16	236	944
13	24	234	936
14	19	256	1024
15	28	299	1196
16	19	296	1184
17	19	291	1164
18	23	278	1112
19	22	253	1012
20	15	232	928
21	22	238	952
22	17	235	940
23	26	273	1092
24	17	242	968
25	14	233	932
26	27	257	1028
27	22	235	940
28	12	227	908
29	14	231	924
30	15	229	916
31	15	233	632
32	17	239	956
33	14	223	892
34	19	227	908
35	20	240	960
36	17	267	1068
37	15	229	916
38	26	282	1128
39	23	250	1000
40	28	237	948
41	14	254	1016
42	31	295	1180
43	10	222	888
44	12	231	924
45	13	229	916
46	12	237	948
47	23	242	968
48	17	228	912



Name of Contig	Number of Bac Clone	Number of Bands	Length(Kb)
49	23	229	916
50	30	242	968
51	28	227	908
52	29	282	1128
53	21	287	1148
54	14	231	924
55	18	242	968
56	14	222	888
57	18	223	892
58	25	289	1156
59	19	272	1088
60	30	248	992
61	20	241	964
62	25	225	900
63	19	241	964
64	19	242	968
65	28	322	1288
66	30	340	1360
67	31	306	1224
68	42	597	2388
69	30	325	1300
70	21	232	928
71	31	381	1524
72	29	305	1220
73	23	323	1292
74	22	325	1300
75	28	305	1220
76	25	318	1272
77	71	562	2248
78	30	384	1536
79	21	328	1312
80	27	318	1272
81	30	348	1392
82	22	337	1348
83	22	322	1288
84	35	391	1564
85	20	308	1232
86	31	396	1584
87	46	464	1856
88	28	374	1496
89	19	226	904
90	23	293	1172
91	19	227	908
92	18	220	880
93	16	242	968
94	19	227	908
95	21	289	1156
96	17	241	964

facilitate the location of genes of interest by positioning cloning; with the easy manupulation of BAC DNA, the BAC contigs will be served as an ideal backbone for DNA sequencing of the entire chromosome. Small errors may exist in contig build-ups, which could be corrected by mapping contigs to chromosomes by marker hybridization.

ACKNOWLEDGMENT

We would like to thank A Coulson of Sanger Center of Cambridge, England, LJ Shi of Shanghai Institute of Cell Biology, CAS, TP Wang of Shanghai Institute of Biochemistry, CAS, for their advice and encouragement; J Li of Shanghai Institute of Plant physiology, CAS, for his contribution to the modification of enzymatic reactions; CZ Zheng of Shanghai Institute of Organic Chemistry for his support in computer work; L Shen of Shanghai Institute of Organic Chemistry for her assistance in data analysis; X Hu, WH Tao, W Li, C Su, XL Zhang, WH Zhao of National Center for Gene Research, CAS, WZ Chang, ST Liu, WH Huang, YY Lu, FD Ni, HM Cao of Shanghai Institute of Biochemistry for their providing some useful fingerprint auto radiograms. GFH is the recipient of Rockefeller Foundation's Biotechnology Career Fellowship. this work was supported by funds from State Commission of Science and Technology, the Chinese Academy of Sciences and the Commission of Science and Technology of Shanghai.

REFERENCES

[1] Arumuganathan K, Earle ED. Nuclear DNA content of some important plant species. Plant Mol Biol Reporter 1991; **9:**208.

[2] Hodges TK, Peng J, Lyznik LA, Koetje DS. Transformation and regeneration of rice protoplasts, PP. 155-174 in Rice Biotechnology, edited by Toennessen G and Khush G. CAB International, Tucson, Ariz. 1991.

[3] ibid.

[4] McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS et al. Molecular mapping of rice chromosome. Theor Appl Genet 1988; **76:**815.

[5] Kurata N, Nagamura Y, Yamamoto K, Harushima Y et al. A 300 Kilobase interval genetic map of rice including 883 expressed sequences. Nature Genetics 1994; **8:**365.

[6] Sulston J, Mallett F, Staden R, Durbin R, Horsnell T, Coulson A. Software for genome mapping by fingerprinting techniques. CABIOS 1988; **4:**125.

[7] Sulston J, Mallett F, Durbin R, Horsnell T. Image analysis of restriction enzyme fingerprint autoradiograms. CABIOS 1989; **5:**101.

[8] Tao QZ, Zhao HY, Qiu LF, Hong GF. Construction of a full bacterial artificial chromosome (BAC) library of Oryza sativa genome. Cell Research 1994; **4:** 127.

[9] Coulson A, Sulston J. Genome mapping by restriction fingerprinting, In: Genome Analysis: A practical approach, edited by Davies KE. IRL Press LTD Oxford UK 1988; PP. 19-40.

Received 2-11-1995. Revised 8-12-1995. Accepted 11-12-1995