Genome physical mapping from large-insert clones by fingerprint analysis with capillary electrophoresis: a robust physical map of *Penicillium chrysogenum*

Zhanyou Xu, Marco A. van den Berg[2], Chantel Scheuring, Lina Covaleda, Hong Lu[1], Felipe A. Santos, Taesik Uhm, Mi-Kyung Lee, Chengcang Wu, Steve Liu[1] and Hong-Bin Zhang*

Department of Soil and Crop Sciences and [1]Department of Computer Science, Texas A&M University, College Station, TX 77843, USA and [2]DAI/INNO (624-0270), DSM, PO Box 425, 2600 AK Delft, The Netherlands

ABSTRACT

Physical mapping with large-insert clones is becoming an active area of genomics research, and capillary electrophoresis (CE) promises to revolutionize the physical mapping technology. Here, we demonstrate the utility of the CE technology for genome physical mapping with large-insert clones by constructing a robust, binary bacterial artificial chromosome (BIBAC)-based physical map of *Penicillium chrysogenum*. We fingerprinted 23.1× coverage BIBAC clones with five restriction enzymes and the SNaPshot kit containing four fluorescent-ddNTPs using the CE technology, and explored various strategies to construct quality physical maps. It was shown that the fingerprints labeled with one or two colors, resulting in 40–70 bands per clone, were assembled into much better quality maps than those labeled with three or four colors. The selection of fingerprinting enzymes was crucial to quality map construction. From the dataset labeled with ddTTP–dROX, we assembled a physical map for *P.chrysogenum*, with 2–3 contigs per chromosome and anchored the map to its chromosomes. This map represents the first physical map constructed using the CE technology, thus providing not only a platform for genomic studies of the penicillin-producing species, but also strategies for efficient use of the CE technology for genome physical mapping of plants, animals and microbes.

INTRODUCTION

Genome physical mapping, i.e. reconstruction of chromosomes or genomes from large-insert, arrayed bacterial artificial chromosome (BAC) or transformation-competent binary BAC (BIBAC) libraries, is becoming an active area of genomic studies (1–3). It has been demonstrated in human (4) and several model species (5–11) that physical maps are essential not only for large-scale genome sequencing, but also for many other aspects of genome research, such as transcript mapping, positional cloning, comparative genome analysis and efficient gene deletions (12–13). The technology of physical mapping by fingerprint analysis of large-insert clones allows the construction of genome maps even for those species for which development of genetic maps is impossible owing to the absence of appropriate mapping populations.

Restriction fingerprinting and contig assembly facilitated with computer technologies (14–25) is emerging as the method of choice for physical mapping of genomes from BACs. Several restriction enzyme-based methods have been developed to generate fingerprints from BACs for physical map construction, including the agarose gel-based method (17,23), the manual DNA sequencing gel-based method (14,20) and the automated DNA sequencing-based method (16,21,22,24,25). Using the first two methods, BAC-based physical maps have been constructed for several species (4–11,26). The application

of capillary electrophoresis (CE) will significantly increase the efficiency and accuracy of quality physical map construction. The CE technology enables labeling DNA fragments with fluorescent-ddNTPs, applying internal size standards in every channel, and fractionating the labeled fragments with capillary sequencers, thus increasing the accuracy and efficiency of determining band size. Three protocols have been developed (24,25) to generate fingerprints from BACs on capillary sequencers. Although a study was conducted to compare the CE-fingerprinting method against other fingerprinting methods *in silico* (24), little is known about the actual utility of CE technology for quality physical map construction. Up to four fluorescent-ddNTPs can be used to label each BAC (24,25), with one fluorescent dideoxynucleotide for each type of restriction fragment ends generated, but it is unclear how many restriction enzymes, labeling colors or bands per clone are needed to efficiently generate quality physical maps with the CE technology.

The relatively small genome of the filamentous fungus *Penicillium chrysogenum* provides an ideal model for optimizing the CE technology for whole-genome physical mapping from BACs. *P.chrysogenum* is an asexual fungus, has a genome size of 34.1 Mb distributed over four chromosomes (27) and is the primary commercial source of penicillinV and penicillinG, the major antibiotic applied during treatment of bacterial infections. The biosynthesis of penicillin and the genes involved have been characterized in detail using *P.chrysogenum* and a related fungus, *Aspergillus nidulans*, as model organisms [reviewed in (28–31)]; however, few other genes have been studied in *P.chrysogenum*. One of the crucial elements inhibiting the progress into other fields of *P.chrysogenum* research is the lack of efficient molecular tools. The low rate of homologous recombination, the multinucleate hyphal compartments and the lack of a sexual cycle (32) are especially hindering genomic and molecular biology studies of *P.chrysogenum*.

We demonstrated the utility of the CE technology for genome physical mapping with large-insert DNA clones by constructing a robust, BIBAC-based physical map of the asexual fungus, *P.chrysogenum*. The physical map will allow detailed analysis of the *P.chrysogenum* genome and its function in relation to penicillin production, and facilitate the understanding of the historic recombination events induced by classical mutagenesis in order to isolate strains with better production performance (27,33,34). This, in turn, will enhance the ability of microbiologists to further increase the organism's capacity to produce penicillin. The process of developing the physical map has demonstrated the utility of the CE technology and developed strategies for rapid and efficient construction of high-quality physical maps for plants, animals and microbes.

MATERIALS AND METHODS

Preparation of high molecular weight DNA in agarose plugs

The international laboratory strain of *P.chrysogenum*, Wisconsin 54–1255 (35), was used in this study. Fungal protoplasts were prepared by incubating freshly grown mycelium with glucanex, a lytic enzyme degrading the cell wall, in isotonic buffer (0.7 M KCl, 50 mM potassium phosphate buffer,

pH 5.8). After separation of the undigested mycelia and cellular debris, protoplasts were washed and stored in STC (218 g/l sorbitol, 7.35 g/l CaCl$_2$, 10 mM Tris–HCl, pH 7.5) at −80°C before use.

The frozen *P.chrysogenum* protoplasts (5×10^9) were thawed on ice and centrifuged at 5000 *g* for 20 min at 4°C. The protoplasts were washed in SCE buffer (1 M sorbitol, 20 mM sodium citrate, pH 5.8, 10 mM EDTA) with the addition of 7.38 g/l of CaCl$_2$ (SCEC) and centrifuged again at 5000 *g*, 4°C for 15 min. The protoplasts were resuspended in 1 ml SCEC and warmed to 45°C. A volume of 1.5 ml of pre-warmed (45°C) 1% low-melting-point agarose in SCEC was added to the protoplasts and gently mixed. The protoplast/agarose mixture was pipetted into 100 μl plug molds (Bio-Rad). After the plugs solidified, the nuclei were lysed and DNA was purified (11,36,37).

Construction of BIBAC libraries

Two BIBAC libraries were constructed from *P.chrysogenum* DNA partially digested with either BamHI or HindIII. The digested DNAs were subjected to a single size selection, ligated into the BIBAC vector pCLD0454 and transformed into DH10B cells (Life Technologies) (36,37).

BIBAC fingerprinting

BIBAC clones were grown in 96-deep well plates containing 1.2 ml Terrific Broth (Life Technologies) with 15 μg/ml tetracycline in an orbital shaker with shaking at 325 r.p.m. for 20 h at 37°C. DNAs were isolated using the DNA Isolation Robotic Workstation AutoGenprep 960 (AutoGen, Inc.). Typically, 0.8–1.2 μg of DNA was obtained per clone. Each DNA was dissolved in 100 μl H$_2$O, and 35 μl of each DNA sample was transferred into 96-microtube plates for digestion and labeling reactions.

The DNA was digested and end-labeled using a five-enzyme, four-color labeling kit consisting of (i) four 6 bp restriction endonucleases, BamHI, Hind III, XbaI and XhoI, producing four different 3′ recessed ends to be labeled, (ii) one 4 bp restriction endonuclease, HaeIII, producing blunted ends and (iii) the SNaPshot Multiplex Ready Reaction Mix (Applied Biosystems). The ends of the fragments produced by digestion with HindIII, BamHI, XbaI and XhoI were labeled with ddATP–dR6G, ddGTP–dR110, ddCTP–dTAMRA and ddTTP–dROX, respectively, and HaeIII was used to digest the labeled fragments for fractionation by CE. Each BIBAC fingerprinting reaction consisted of 35 μl of DNA (250–400 ng), 4.5 μl of 10× buffer 2 (New England Biolabs), 0.45 μl of 10 mg/ml BSA, 0.04 μl of 80 U/μl HindIII, 0.15 μl of 20 U/μl of XbaI, 0.15 μl of 20 U/μl of XhoI and 4.71 μl of H$_2$O in a total volume of 45 μl. The reaction was incubated at 37°C for 2 h and then inactivated by incubation at 65°C for 20 min. The sample was cooled on ice and centrifuged briefly. To the same reaction tube, a 15 μl mixture of 1.5 μl of 10× core buffer (Promega), 0.15 μl of 10 mg/ml BSA, 0.04 μl of 80 U/μl BamHI, 0.05 μl of 80 U/μl HaeIII, 2.5 μl of the SNaPshot Multiplex Ready Reaction Mix (Applied Biosystems) and 10.76 μl of H$_2$O was added. The reaction was incubated at 37°C for 2 h, immediately followed by incubation at 65°C for 1 h for enhanced labeling.

The labeled DNA was precipitated by adding 120 µl of 100% ethanol and 6 µl of 3 M NaAc and incubating at −80°C for at least 30 min. Samples were thawed and centrifuged at 3000 r.p.m. in a tabletop centrifuge for 30 min. The pellets were washed two times in 200 µl of 70% ethanol and then dissolved in a mixture of 9.8 µl of Hi-Di formamide and 0.2 µl of the internal size marker LIZ-500 (Applied Biosystems). The DNA was denatured at 95°C for 8 min and cooled on ice for 8 min. CE was carried out on the ABI 3100 DNA Analyzer (Applied Biosystems) using the default GeneScan module (36 cm capillaries, POP4, 60°C, 10 s for sample injection and 44 min/run). The raw fingerprint data in the window ranging from 35 to 500 bases were collected using the GeneScan V3.70 and the ABI 3100 data collection V1.0.1.

Physical map contig assembly

The raw fingerprint data from the ABI 3100 DNA Analyzer were transformed into band data based on the peak area and peak height, and edited both automatically and manually with the 'MultiColor Editor' program (H. Lu and S. Liu, unpublished data). This software distinguishes the peaks corresponding to restricted fragments from the peaks corresponding to background noise in the profile of each fingerprint and removes vector bands from raw fingerprint data. The band data were converted into FPC data by the 'SizeToFpc' software (H. Lu and S. Liu, unpublished data), which can be used in both Windows and UNIX systems with both automatic and manual functions.

The edited fingerprint data were used to assemble contigs using the FPC V6.2 software (18,19). The fingerprint of each clone was collected and compiled according to the labeled end colors, with the fingerprint labeled with one color per window (24). Because each clone was digested and labeled with four colors, a total of 15 sets of fingerprint data categorized into four groups were created according to the labeling colors (24): (i) 4 single-color fingerprint datasets (B for blue, G for green, Y for yellow and R for red), (ii) 6 two-color combinations (BG, BY, BR, GY, GR and YR), (iii) 4 three-color combinations (BGY, BGR, BYR and GYR) and (iv) 1 four-color combination (BGYR). A tolerance of five corresponding to 0.5 bp was used for contig assembly (24).

Anchoring of the contig map to chromosomes

Two approaches were used to anchor the contig map assembled from the BIBAC fingerprints to *P.chrysogenum* chromosomes. First, one half of a 100 µl agarose plug containing *P.chrysogenum* chromosome-sized DNA was subjected to PFGE on a 0.6% chromosomal grade agarose gel (Bio-Rad) in 0.5× TBE with *Schizosacchomyces pombe* chromosome-size DNA (Bio-Rad) as DNA molecular weight markers (38). The DNA in the gel was nicked and blotted onto Hybond N+ membrane (Amersham) using 1.5 M NaCl, 0.4 N NaOH for 24 h. The Southern blot was hybridized with probes generated from DNA of selected BIBACs labeled with [^{32}P]dCTP (7). Two to three BIBACs were selected from each contig of the physical map as probes. The blots were hybridized overnight, washed three times in 0.2× SSC, 0.1% SDS at 65°C, 15 min each wash, and exposed to X-ray film.

In the second approach, the BIBAC libraries of *P.chrysogenum* were robotically double-gridded and blotted onto Hybond N+ membrane (36) and then probed with chromosome-specific gene probes for chromosomes I and II of *P.chrysogenum*, with *pcbC* for chromosome I (27), and *pyrG* (27) and *niaD* (M. A. van den Berg, unpublished data) for chromosome II. Plates 1–8 of the BamHI library were hybridized with *pcbC*, plates 1–8 of the Hind III library with *pyrG*, and *niaD* was used for plates 5–8 of both BamHI and HindIII libraries. Labeling of probes, hybridization and detection of selected BIBAC filters were performed using the ECL labeling kit (Amersham).

RESULTS

### Source BIBAC library construction

Two BIBAC libraries were constructed from *P.chrysogenum* Wisconsin 54–1255 in the BIBAC vector pCLD04541 to develop the whole-genome physical map of *P.chrysogenum*. The two BIBAC libraries were constructed with BamHI and HindIII that have recognition sites with differing G/C contents to maximize the actual genome coverage of the libraries and thus the resulting physical map (Table 1). The BamHI library consisted of 3456 clones. Analysis of 73 random clones showed that it had an average insert size of 126 kb and 4.1% insert-empty clones, providing a coverage of 12.1× the 34.1 Mb genome (27). The HindIII library consisted of 4608 clones. Analysis of 68 random clones showed that it had an average insert of 130 kb and no insert-empty clones, providing a genome coverage of 17.3×. The combined libraries contained a total of 8064 clones with an average insert size of 128 kb and a coverage of 29.4×.

BIBAC fingerprinting

A combination of five enzymes, BamHI, HindIII, XbaI, XhoI and HaeIII, were used to fingerprint each BIBAC. The restricted fragments of the first four enzymes were end-labeled with a SNaPshot kit containing four fluorescent-ddNTPs, thus providing informative bands for each clone fingerprint. The frequently cutting enzyme, HaeIII, was only used to cut the labeled fragments into smaller sizes for CE-based analysis, and their resulting ends were not labeled, thus providing no information for each clone fingerprint. To estimate the impact of the number of bands per clone, or the number of restriction enzymes or labeling colors on the quality of the constructed physical map, a total of 7296 randomly selected clones

Table 1. The *P.chrysogenum* BIBAC libraries and their clones used in the physical map construction

Features	HindIII library	BamHI library	Total
Cloning vector	pCLD04541	pCLD04541	1
Cloning sites	HindIII	BamHI	2
No. of clones	4608	3456	8064
Average insert size (kb)	130	126	128
Insert-empty clones (%)	0.0	4.1	1.8
Genome equivalents	17.3×	12.1×	29.4×
No. of clones fingerprinted	4224	3072	7296
No. of clone fingerprints used	3849	2315	6164
Genome equivalents used	14.5×	8.6×	23.1×
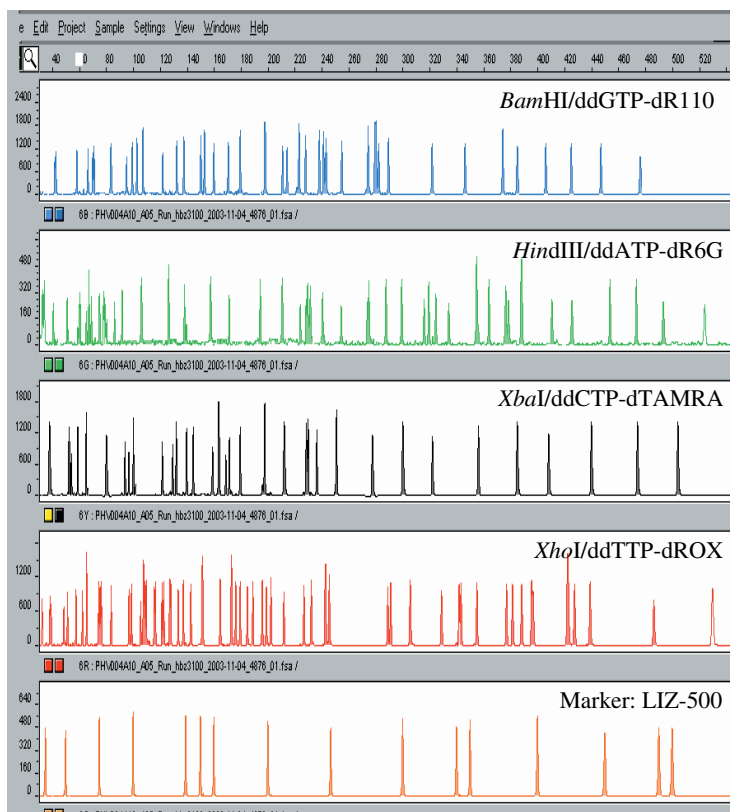
The fingerprint of clone PHV004A10

Figure 1. Example of a single BIBAC fingerprint generated with five restriction enzymes (the HaeIII fragment ends were not labeled) and four fluorescent-ddNTPs of the SNaPshot Multiplex Ready Reaction Mix on the ABI 3100 DNA analyzer. The blue peaks show the BamHI fragments labeled with ddGTP-dR110; the green peaks show the HindIII fragments labeled with ddATP-dR6G; the black peaks show the XbaI fragments labeled with ddCTP–dTAMRA; the red peaks show the XhoI fragments labeled with ddTTP–dROX; and the orange peaks show the internal size standard LIZ-500.

from both BIBAC libraries were each fingerprinted using the five enzymes and four labeling colors separated on an ABI 3100 DNA Analyzer (Table 1). Figure 1 shows an example of a BIBAC fingerprint generated with this method. Since 1132 (15.5%) of the clones had no inserts, produced less than five bands, or failed in fingerprinting, 6164 clones, equivalent to 23.1× of the *P.chrysogenum* genome, were converted into the FPC database. Of these 6164 clones, 3849 (14.5×) were from the HindIII library and 2315 (8.6×) from the BamHI library. Since each BIBAC was labeled with four individual colors, a total of 15 fingerprint datasets, based on single or combinations of up to four colors, were compiled and categorized into four groups (see Materials and Methods; Table 2). The single-color BIBAC fingerprints had an average number of bands ranging from 26.0 for the yellow fingerprint generated with XbaI to 40.2 for the red fingerprint generated with XhoI.

Physical map contig assembly and editing

Before contig assembly from the 15 possible fingerprint datasets, we explored various strategies of contig assembly by comparatively analyzing the datasets using the program FPC V6.2 (18,19) with number of clones equivalent to 5.3× (1442 clones), 10.9× (3106 clones), 16.0× (4549 clones) and 23.1× (6164 clones) at a cutoff value of 1×10^{-20}. Two results were apparent from this contig assembly pilot experiment.

The first one was that the percentages of questionable clones (Qs) and chimeric contigs increased significantly (although the number of contigs decreased) as the genome coverage increased from 5.3× to 23.1×. This result was in agreement with the previous study using *in silico* contig assembly (19). The second result was that the percentages of Qs and chimeric contigs also increased significantly from the fingerprint data-sets labeled with one color to those labeled with more than one color, though the numbers of contigs assembled from the single-color datasets were significantly larger than those assembled from combined datasets labeled with more than one color (Supplementary Table 1). This pilot experiment indic-ated that it was impossible to make reasonable comparisons among the 15 fingerprint datasets for quality map construction by assembling contigs at a single cutoff value. Therefore, we decided to assemble contig maps using variable cutoffs and controlling the number of Qs, a significant parameter for qual-ity contig maps (19,24).

The 15 sets of fingerprint data were then individually assembled into contigs using the FPC program. A cutoff value of 1×10^{-37} was used for the initial automated contig assem-bly. Cycles of automated contig assembly were continued with successively higher cutoff values while simultaneously main-taining Qs ≤1%. The automated contig assembly results are summarized in Table 2. In comparison among the four groups of datasets, the two-color fingerprint datasets resulted in the

Table 2. Contig assembly from the fingerprints of 6164 BIBACs (23.1×) generated with single enzyme or labeling colors, or the combinations of multiple enzymes or labeling colors by lowering the cutoff settings until ≤1% Qs were generated

Combination[a] of color(s)	No. of bands per clone	Size represented per band (bp)	Cutoff needed for 1% Qs	No. of contigs	No. of clones in the contigs	No. of singles	No. of Qs	Mean length/ contig (kb)[b]
B	34.4	3721	5×10^{-25}	216	5344	820	15	231
G	37.8	3386	1×10^{-24}	213	4521	1642	21	293
Y	26.0	4923	1×10^{-29}	161	5294	867	46	278
R	40.2	3184	1×10^{-23}	89	5851	313	4	426
Mean	34.6	3699		170	5252	910	22	307
BG	73.0	1753	1×10^{-30}	246	5040	1124	36	245
BR	74.6	1716	1×10^{-27}	42	5969	195	44	758
BY	60.4	2119	2×10^{-28}	98	5744	420	28	406
RY	66.2	1934	5×10^{-28}	66	5824	340	19	553
GR	78.9	1622	5×10^{-32}	255	5176	988	41	236
GY	64.7	1978	1×10^{-28}	211	4897	1267	40	270
Mean	69.6	1839		153	5441	722	35	411
BGR	113.2	1131	1×10^{-35}	299	5373	791	48	216
BGY	99.0	1293	1×10^{-35}	395	4934	1230	29	196
BRY	100.6	1272	3×10^{-30}	65	5913	251	32	564
GRY	104.9	1220	1×10^{-35}	340	5152	1012	22	209
Mean	104.4	1226		275	5343	821	33	296
BGYR	139.3	919	1×10^{-35}	281	5439	725	30	225

[a]Letter 'B' indicates 'blue' fingerprints generated with BamHI; 'G' for 'green' fingerprints generated with HindIII; 'Y' for 'yellow' fingerprints generated with XbaI; and 'R' for 'red' fingerprints generated with XhoI.
[b]The mean length of the contigs from each assembly was calculated by the product of the mean size that each unique band (the average insert size of the clones divided by the average number of bands per clone) is equivalent to and the total number of unique bands of all contigs divided by the number of contigs of the assembly.

lowest average number of contigs (153 contigs), followed by the single-color fingerprint datasets (170 contigs), the three-color datasets (275 contigs) and the four-color datasets (281 contigs). Within each of the single-, two- and three-color fingerprint dataset groups, the number of resultant contigs varied significantly. For example, in the single-color group, the number of contigs assembled from the fingerprints labeled with red color was less than half of the number of contigs assembled from the fingerprints labeled with each of the other three colors. Among the 15 fingerprint datasets, the BR combination generated with BamHI and XhoI was assembled into the smallest number of contigs (42 contigs), having the largest average contig length (758 kb/contig).

Because the R, BR and BRY fingerprint datasets were assembled into the fewest and largest automatic contigs in the single-, two-color and three-color groups (Table 2), respectively, the contigs assembled from these datasets were selected for further study. The contigs resulting from automated assembly of the R, BR and BRY datasets were further analyzed manually using the Calc CB map and Fp Order CB map functions of the FPC program as Ren *et al.* (11). Chimeric contigs were disassembled and reassembled at higher stringency, and Qs were removed manually. Neighboring contigs were merged, and singletons added by gradually reducing the cutoffs to 1×10^{-15}. As a result, 11 contigs were obtained from each of the R and BR fingerprint contigs, and 12 from the BRY fingerprint contigs. Table 3 shows the R fingerprint map contigs, ranging from 294 to 7661 kb per contig and spanning a total of 38.4 Mb in physical length.

Contig verification and anchoring to chromosomes

Because the R, BR and BRY fingerprint datasets were assembled into similar number of contigs, but generation of single-color fingerprints is much more efficient in general, we focused our contig verification and contig anchoring efforts

Table 3. Summary of the *P.chrysogenum* physical map

Chromosomes	Size (Mb)[a]	Contig code	Evidence[b]	No. of clones	Map length (kb)
I	10.4	6	Deduction	842	5466
		7	*pcbC* probe ctg probes	784	5280
		33	Deduction	303	1398
		Subtotal		1929	12 144
II	9.6	25	*niaD* probe ctg probes	360	2442
		39	*pyrG* probe ctg probes	1005	7661
		Subtotal		1365	10 103
III	7.3	2	ctg probes	495	3930
		21	ctg probes	895	5952
		Subtotal		1390	9882
IV	6.8	1	ctg probes	643	4339
		13	ctg probes	242	1629
		23	ctg probes	52	294
		Subtotal		937	6262
Genome	34.1	Total		5621	38 401

[a]The sizes of the chromosomes were estimated by PFGE (27).
[b]The contigs 6 and 33 were anchored to chromosome I or II by using contig BIBAC probes, but it was difficult to clearly locate them to individual chromosomes. The positions of the two contigs were deduced according to the total contig length and the chromosome I and II lengths.

on the R fingerprint map. We first analyzed the R fingerprint contigs against those of the remaining 14 fingerprint datasets to test whether the contigs assembled from the BR, BRY and the other 12 fingerprint datasets could be used to further refine the R fingerprint contig map. The result showed that the contigs assembled from the BR and BRY fingerprint datasets were essentially the same as the R fingerprint contigs in terms of clone content and order, while the contigs of the other datasets were found to be subsets of the R fingerprint contigs. Therefore, the R fingerprint contig map could not be further

refined using the contigs assembled from the other datasets. Although the attempt failed, the results provided evidence that the assembled map contigs were reliable.

To further verify and anchor the R fingerprint contigs to the *P.chrysogenum* chromosomes, we screened a subgroup of plates of the contig map source BIBAC libraries (see Materials and Methods) with *pcbC*, *pyrG* and *niaD* previously located to chromosomes I and II of *P.chrysogenum* [(27); M. A. van den Berg, unpublished data]. A total of 16 positive clones were obtained with *pcbC*, with 12 of them located to a segment of contig 7 (ctg7) and four clones not in the fingerprint database (Supplementary Figure 4). Similarly, *niaD* identified 11 positive clones, with 8 of them located to a segment of contig 25, 1 in a singleton, and 2 not in the fingerprint database (Supplementary Figure 8). Using *pyrG*, nine clones were identified, of which seven were located to a segment of contig 39, and two were not in the fingerprint database (Supplementary Figure 10). These results not only further verified the accuracy of the contigs but also anchored them to chromosomes I and II, respectively (Table 3).

For contig verification and chromosome anchoring, we also fractionated the chromosome-sized DNA of *P.chrysogenum* by PFGE and conducted Southern analysis with BIBACs randomly selected from the contig map, with 2–3 BIBACs per contig (Supplementary Figures 1–10). The hybridization anchored 5 of the 11 contigs to chromosomes III and IV and the remaining contigs to chromosome I or II (Supplementary Figure 11). This result, in combination with the *pcbC*, *pyrG* and *niaD* hybridizations, anchored all of the 11 contigs to the chromosomes of the fungus. Furthermore, all probe BIBACs randomly selected from a single contig were found to hybridize with the same chromosome, indicating that the contigs were assembled properly. At this point, we attempted to further merge the contigs anchored to the same chromosomes by lowering the cutoff to 1×10^{-13}. Consequently, two contigs anchored to chromosome I were merged into one contig, and the number of the R fingerprint map contigs was reduced to 10 (Table 3; Supplementary Figures 1–10).

DISCUSSION

We have developed a robust physical map of the *P.chrysogenum* genome from BIBAC fingerprints generated by the CE technology and anchored it to corresponding chromosomes. The map consists of 10 contigs, with each contig covering an average of 3.84 Mb and 2–3 contigs per chromosome. The map is highly reliable, as indicated by comparative analysis of the contig maps assembled from 15 fingerprint datasets, source library screening and contig anchoring. The total physical length of the map is ∼38.4 Mb, 12.6% larger than the estimated 34.1 Mb size of the *P.chrysogenum* genome (27). This difference may indicate that some contigs still overlap adjacent contigs even though these overlaps could not be detected with the probes and stringency used, and/or the genome size of *P.chrysogenum* was underestimated (27). Because of the difficulty of constructing a genetic map for an organism with asexual reproduction (34), the physical map reported here will provide an especially useful tool for genomics research of *P.chrysogenum* and related species.

The *P.chrysogenum* physical map represents the first reported genome physical map constructed from clone fingerprints generated by the use of the CE technology. In the process of constructing the physical map using BIBAC fingerprints generated by the technology, we demonstrated several useful points and developed strategies for quality physical map construction. As shown in Table 2, a range between 40 and 70 bands per clone is desirable for quality map construction. This result is in agreement with that obtained in our previous *in silico* study (24) and is also supported by two other ongoing physical mapping projects in our laboratory (unpublished data). While attempts were made to refine the contig map constructed from one enzyme fingerprint dataset using the maps assembled from other enzyme and enzyme combination fingerprint datasets, this strategy was not effective for map refinement. These results are in agreement with those obtained from the soybean BAC fingerprints generated by the manual sequencing gel-based fingerprinting method (26).

The results of the contig assembly experiments indicate that the datasets derived from labeling with one or two colors (two or three enzymes, including HaeIII) resulted in many fewer contigs than the assemblies with datasets derived from three or four colors (four or five total enzymes). It is much easier to correctly assemble a contig from fingerprints with fewer bands above a threshold number. Using an excess number of restriction enzymes results in many more fingerprint bands to analyze, making fingerprint comparisons and contig assembly more difficult. The use of excess number of enzymes and labeling colors to generate fingerprints increases the cost and reduces the throughput of fingerprint production. Importantly, it makes the fingerprinting protocols more complicated. The fingerprints generated with a more complicated protocol are likely to be lower in quality than those generated with a simpler protocol (24). Increasing the number of enzymes and labeling colors used for fingerprinting is expected to increase the possibility of partial digestion, star activity and/or low labeling efficiency. The lower fingerprint quality would lead to a lower quality contig map, regardless of which dataset is used. Therefore, the fingerprinting protocols using one or two 6 bp enzymes, along with a 4 bp cutter, and one or two labeling colors, such as those of Gregory *et al*. (16) and Xu *et al*. (24), are recommended for constructing quality physical maps from fingerprints generated using CE technology and the FPC program (19).

In addition, protocols in which each clone is labeled with a single dye allow multiplex fingerprint production, depending on the availability of the number of migration rate-matched dyes. Since at least three such dyes, including ddATP-HEX, ddATP-NED and ddATP-TET (16), are commercially available for fingerprint analysis, then a combination of three BACs each labeled with a single unique dye could be analyzed on a single channel of a capillary sequencer. Therefore, both CE throughput could be increased and cost could be reduced by as much as 3-fold if each BAC is labeled with a single dye (16,24). Using a fingerprinting kit consisting of three enzymes (HindIII, BamHI and HaeIII) and one labeling dye (24), we have constructed a contig physical map of the cotton genome (Z. Xu, L. Covaleda, R. Kohel, J. Z. Yu, H.-B. Zhang and M.-K. Lee, unpublished data). In this study, because two BACs labeled with different colors were run in a single capillary channel, the fingerprinting throughput was doubled,

the cost decreased and the quality of the resulting map increased. Moreover, for further enhanced fingerprint production throughput, a higher-throughput instrument, such as the 96-capillary ABI 3730 or MegaBACE 1000, could be used for fingerprint fractionation.

Nevertheless, it should be noted that the effect of number of bands per clone or number of enzymes on quality map construction with the CE technology varied significantly (Table 2). This is probably due to the distribution of the bands within the data collection window (35–500 bases) and/or fingerprint quality. Therefore, a pilot experiment of random clones digested with different enzymes will be helpful to select suitable enzymes for fingerprint production for a particular genome of interest. This study suggests that the enzymes selected should result in no partial or non-specific (star activity) digestion, and should generate fingerprints with clearly defined sharp peaks with a good distribution in the 35–500 base size window and low background noise. For example, the digestion of *P.chrysogenum* BIBACs with the enzyme XhoI resulted in the best fingerprints for physical map construction, considering fingerprint quality, cost and throughput.

The techniques and strategies demonstrated here are applicable for whole-genome physical mapping of other species, including plants, animals and microbes. Although the genome of *P.chrysogenum* is much smaller and simpler than the genomes of plants and animals, previous physical mapping studies (4–11,26) consistently indicated that genome physical mapping with large-insert clones by fingerprint analysis does not seem significantly influenced by the size and complexity of the target genomes. This study showed that the quality of the contig map, indicated by contig length and ≤1% Qs, constructed from the fingerprints generated with one or two labeling colors (two or three enzymes including HaeIII) was much higher than contig maps generated with three or four colors (four or five enzymes), suggesting that it is unnecessary to use five enzymes and four labeling colors to generate BAC fingerprints for quality map construction using CE technology and the PFC program (19).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Zhang,H.-B. and Wu,C. (2001) BAC as tools for genome sequencing. *Plant Physiol. Biochem.*, **39**, 195–209.
2. Meyers,B.C., Scalabrin,S. and Morgante,M. (2004) Mapping and sequencing complex genomes: let's get physical. *Nature Rev. Genet.*, **5**, 578–588.
3. Wu,C., Sun,S., Lee,M.-K., Xu,Z., Ren,C. and Zhang,H.-B. (2005) Whole genome physical mapping: an overview on methods for DNA fingerprinting. In Meksem,K. and Kahl,G. (eds), *Handbook of Plant Genome Mapping: Genetic and Physical Mapping*. Wiley-VCH Verlag GmbH, Weinheim, Germany, pp. 257–284.
4. The International Human Genome Mapping Consortium (2001), A physical map of the human genome. *Nature*, **409**, 934–941.
5. Marra,M., Kucaba,T., Sekhon,M., Hillier,L., Martienssen,R., Chinwalla,A., Crockett,J., Fedele,J., Grover,H. *et al.* (1999) A map for sequence analysis of the *Arabidopsis thaliana* genome. *Nature Genet.*, **22**, 265–270.
6. Hoskins,R.A., Nelson,C.R., Berman,B.P., Laverty,T.R., George,R.A., Ciesiolka,L., Naeemuddin,M., Arenson,A.D., Durbin,J., David,R.G. *et al.* (2000) A BAC-based physical map of the major autosomes of *Drosophila melanogaster*. *Science*, **287**, 2271–2274.
7. Chang,Y.L., Tao,Q., Scheuring,C., Meksem,K. and Zhang,H.-B. (2001) An integrated map of *Arabidopsis thaliana* for functional analysis of its genome sequence. *Genetics*, **159**, 1231–1242.
8. Tao,Q., Chang,Y.L., Wang,J., Chen,H., Islam-Faridi,M.N., Scheuring,C., Wang,B., Stelly,D.M. and Zhang,H.-B. (2001) BAC-based physical map of the rice genome constructed by restriction fingerprint analysis. *Genetics*, **158**, 1711–1724.
9. Chen,M., Presting,G., Barbazuk,W.G., Goicoechea,J.L., Blackmon,B., Fang,G., Kim,H., Frisch,D., Yu,Y., Sun,S. *et al.* (2002) An integrated physical and genetic map of the rice genome. *Plant Cell*, **14**, 537–545.
10. Gregory,S.G., Sekhon,M., Schein,J., Zhao,S., Osoegawa,K., Scott,C.E., Evans,R.S., Burridge,P.W., Cox,T.V., Fox,C.A. *et al.* (2002) A physical map of the mouse genome. *Nature*, **418**, 743–750.
11. Ren,C., Lee,M.-K., Yan,B., Ding,K., Cox,B., Romanov,M.N., Price,J.A., Dodgson,J.B. and Zhang,H.-B. (2003) A BAC-based physical map of the chicken genome. *Genome Res*., **13**, 2754–2758.
12. Chaveroche,M.K., Ghigo,J.M. and d'Enfert,C. (2000) A rapid method for efficient gene replacement in the filamentous fungus *Aspergillus nidulans. Nucleic Acids Res*., **28**, e97.
13. Hamer,L., Adachi,K., Montenegro-Chamorro,M.V., Tanzer,M.M., Mahanty,S.K., Lo,C., Tarpey,R.W., Skalchunes,A.R., Heiniger,R.W. *et al.* (2001) Gene discovery and gene function assignment in filamentous fungi. *Proc. Natl Acad. Sci. USA*, **98**, 5110–5115.
14. Coulson,A., Sulston,J., Brenner,S. and Karn,J. (1986) Toward a physical map of the genome of the nematode *Caenorhabditis elegans. Proc. Natl Acad. Sci. USA*, **83**, 7821–7825.
15. Sulston,J., Mallett,F., Staden,R., Durbin,R., Horsnell,T. and Coulson,A. (1988) Software for genome mapping by fingerprinting techniques. *Comput. Appl. Biosci*, **4**, 125–132.
16. Gregory,S.G., Howell,G.R. and Bentley,D.R. (1997) Genome mapping by fluorescent fingerprinting. *Genome Res.*, **7**, 1162–1168.
17. Marra,M., Kucaba,T., Dietrich,N.L., Green,E.D., Brownstein,B., Wilson,R.K., McDonald,K.M., Hillier,L.W., McPherson,J.D. and Waterston,R.H. (1997) High throughput fingerprint analysis of large-insert clones. *Genome Res.*, **7**, 1072–1084.
18. Soderlund,C., Longden,I. and Mott,R. (1997) FPC: a system for building contigs from restriction fingerprinted clones. *Comput. Appl. Biosci.*, **13**, 523–535.
19. Soderlund,C., Humphray,S., Dunham,A. and French,L. (2000) Contigs built with fingerprints, markers, and FPC V4.7. *Genome Res.*, **10**, 1772–1787.
20. Zhang,H.-B. and Wing,R.A. (1997) Physical mapping of the rice genome with BACs. *Plant Mol. Biol.*, **35**, 115–127.
21. Ding,Y., Johnson,M.D., Colayco,R., Chen,Y.J., Melnyk,J., Schmitt,H. and Shizuya,H. (1999) Contig assembly of bacterial artificial chromosome clones through multiplexed fluorescence-labeled fingerprinting. *Genomics*, **56**, 237–246.
22. Ding,Y., Johnson,M.D., Chen,W.Q., Wong,D., Chen,Y.-J., Benson,S.C., Lam,J.Y., Kim,Y.-M. and Shizuya,H. (2001) Five-color-based high-information-content fingerprinting of bacterial artificial chromosome clones using Type IIS restriction endonucleases. *Genomics*, **74**, 142–154.
23. Olson,M.V., Dutchik,J.E., Graham,M.Y., Brodeur,G.M., Helms,C., Frank,M., MacCllin,M., Scheinman,R. and Frank,T. (1986) Random-clone strategy for genomic restriction mapping in yeast. *Proc. Natl Acad. Sci. USA*, **83**, 7826–7830.
24. Xu,Z., Sun,S., Covaleda,L., Ding,K., Zhang,A., Scheuring,C. and Zhang,H.-B. (2004) Genome physical mapping with large-insert bacterial clones by fingerprint analysis: methodologies, source clone genome coverage and contig map quality. *Genomics*, **84**, 941–951.
25. Luo,M.-C., Thomas,C., You,F.M., Hsiao,J., Ouyang,S., Buell,C.R., Malandro,M., McGuire,P.E., Anderson,O.D. and Dvorak,J. (2003) High-throughput fingerprinting of bacterial artificial chromosomes using

the SNaPshot labeling kit and sizing of restriction fragments by capillary electrophoresis. *Genomics*, **82**, 378–389.

26. Wu,C., Sun,S., Nimmakayala,P., Santos,F., Meksem,K., Springman,R., Ding,K., Lightfoot,D.A. and Zhang,H.-B. (2004) A BAC- and BIBAC-based physical map of the soybean genome. *Genome Res.*, **14**, 319–326.

27. Fierro,F., Gutiérrez,S., Díez,B. and Martín,J.F. (1993) Resolution of four large chromosomes in penicillin-producing filamentous fungi: the penicillium gene cluster is located on chromosome II (9.6 Mb) in *Penicillium notatum* and chromosome I (10.4 Mb) in *Penicillium chrysogenum. Mol. Gen. Genet.*, **241**, 573–578.

28. Gutierrez,S., Fierro,F., Casqueiro,J. and Martin,J.F. (1999) Gene organization and plasticity of the β-lactam genes in different filamentous fungi. *Antonie van Leeuwenhoek*, **75**, 81–94.

29. Martin,J.F., Casqueiro,J., Kosalkova,K., Marcos,A.T. and Gutierrez,S. (1999) Penicillin and cephalosporin biosynthesis: mechanism of carbon catabolite regulation of penicillin production. *Antonie van Leeuwenhoek*, **75**, 21–31.

30. Litzka,O., Then Bergh,K., Van den Brulle,J., Steidl,S. and Brakhage,A.A. (1999) Transcriptional control of expression of fungal β-lactam biosynthesis genes. *Antonie van Leeuwenhoek*, **75**, 95–105.

31. Van de Kamp,M., Driessen,A.J. and Konings,W.N. (1999) Compartmentalization and transport in β-lactam antibiotic biosynthesis by filamentous fungi. *Antonie van Leeuwenhoek*, **75**, 41–78.

32. Elander,R.P., Espenshade,M.A., Pathak,S.G. and Pan,C.H. (1973) The use of parasexual genetics in an industrial strain improvement program with *Penicillium chrysogenum*. In Demainn,A.L. and Davies,J.E. (eds), *Manual of Industrial Microbiology and Biotechnology*. ASM Press, Washington, DC, pp. 239–253.

33. Fierro,F. and Martin,J.F. (1999) Molecular mechanisms of chromosomal rearrangement in fungi. *Crit. Rev. Microbiol.*, **25**, 1–17.

34. Pontecorvo,G. and Sermonti,G. (1953) Recombination without sexual reproduction in *Penicillium chrysogenum. Nature*, **172**, 126–127.

35. Elander,C. (1983) Strain improvement and preservation of β-lactam-producing microorganisms. In Demain,A.L. and Salomon,N. (eds), *Antibiotics Containing the* β-*lactam Structure I*. Springer-Verlag, NY, pp. 97–146.

36. Zhang,H.-B. (2000) *Construction and Manipulation of Large-Insert Bacterial Clone Libraries: Manual*. Texas A&M University, College Station, TX.

37. Wu,C., Xu,Z. and Zhang,H.-B. (2004) DNA libraries. In Meyers,A. (ed.), *Encyclopedia of Molecular Cell Biology and Molecular Medicine*. 2nd edn. Wiley-VCH Verlag GmbH, Weinheim, Germany, vol. 3, pp. 385–425.

38. Färber,F. and Geisen,R. (1994) Antagonistic activity of the food relevant filamentous fungus *Penicillium nalgiovense* is due to the production of penicillin. *Appl. Environ. Microbiol.*, **60**, 3401–3404.