Plant Molecular Biology **35:** 115–127, 1997.

Physical mapping of the rice genome with BACs

Hong-Bin Zhang[1,*] and Rod A. Wing[2]
[1]*Department of Soil and Crop Sciences, Crop Biotechnology Center, Texas A&M University, College Station, TX 77843-2123, USA (*author for correspondence);* [2]*Department of Agronomy and Biological Sciences, 275 Poole Ag Center, Box 340359, Clemson University, Clemson, SC 29634, USA*

Key words: physical mapping, rice, bacterial artificial chromosome

Abstract

The development of genetics in this century has been catapulted forward by several milestones: rediscovery of Mendel's laws, determination of DNA as the genetic material, discovery of the double-helix structure of DNA and its implications for genetic behavior, and most recently, analysis of restriction fragment length polymorphisms (RFLPs). Each of these milestones has generated a huge wave of progress in genetics. Consequently, our understanding of organismal genetics now extends from phenotypes to their molecular genetic basis. It is now clear that the next wave of progress in genetics will hinge on genome molecular physical mapping, since a genome physical map will provide an invaluable, readily accessible system for many detailed genetic studies and isolation of many genes of economic or biological importance. Recent development of large-DNA fragment (>100 kb) manipulation and cloning technologies, such as pulsed-field gel electrophoresis (PFGE), and yeast artificial chromosome (YAC) and bacterial artificial chromosome (BAC) cloning, has provided the powerful tools needed to generate molecular physical maps for higher-organism genomes. This chapter will discuss (1) an ideal physical map of plant genome and its applications in plant genetic and biological studies, (2) reviews on physical mapping of the genomes of *Caenorhabditis elegans*, *Arabidopsis thaliana*, and man, (3) large-insert DNA libraries: cosmid, YAC and BAC, and genome physical mapping, (4) physical mapping of the rice genome with BACs, and (5) perspectives on the physical mapping of the rice genome with BACs.

Introduction

The development of genetics in this century has been exponential due to several milestones: rediscovery of Mendel's laws, determination of DNA as the genetic material, discovery of the double-helix structure of DNA and its implications for genetic behavior, and most recently, analysis of restriction fragment length polymorphisms (RFLPs). Each of these milestones had led to a huge wave of progress in genetics. Consequently, our understanding of organismal genetics now extends from phenotypes to their molecular genetic basis. It is now clear that the next wave of progress in genetics will hinge on molecular physical mapping of the genomes, since a genome physical map will provide an invaluable, readily accessible system for many detailed genetic studies. Investigations will extend bidirectionally: not only from the phenotypes to the molecules, but also from the molecules to the phenotypes. Recent development of large DNA fragment (>100 kb) manipulation and cloning technologies, such as pulsed-field gel electrophoresis (PFGE)), and yeast artificial chromosome (YAC) [9] and bacterial artificial chromosome (BAC) [46] cloning, has provided the powerful tools needed to generate molecular physical maps for higher-organism genomes. Once generated, the physical map will provide virtually unlimited numbers of DNA markers from any chromosomal region for gene tagging, gene manipulation, and genetic studies. It will also provide an on-line framework for studies in genome molecular structure, genome organization and evolution, gene regulation, and gene interaction. The identification, isolation, characterization, and manipulation of genes will become far more feasible than

ever before. The physical map, therefore, will become central to all types of genetic and molecular inquiry and manipulation, including genome analysis, gene cloning, and crop genetic improvement.

An ideal physical map of a plant genome and its applications in plant genetic and biological studies

A physical map actually consists of continuously overlapping contigs of large-insert genomic DNA clones, in which the distance between two DNA markers is measured in kb. For organisms with multiple chromosomes, each chromosome would ideally consist of just one continuous contig, but in reality often consists of several contigs. Figure 1 presents an ideal molecular physical map of one chromosome of a genome. The physical map is marked with anchor DNA markers, or landmarks, selected from the developed RFLP maps. The anchor markers are a group of selected DNA markers that are fairly evenly distributed across the genome, can cross-hybridize to DNAs from the genomes of related species, and thus can be mapped to their respective RFLP maps. Some of these anchor DNA markers are also mapped to the corresponding chromosome by means of chromosome *in situ* hybridization of anchor marker-associated large-insert clones.

The ideal physical map presented in Figure 1 is integrated with genetic and cytogenetic maps. Clearly, the reliability of a physical map will be much higher if its construction is guided by genetic and cytogenetic maps, which can minimize the mapping errors from chimaeric clones and/or DNA fragment duplications. Chimaeric clones have been discovered in many large-insert DNA libraries [3, 48]. DNA fragment duplication occurs in all plant genomes (for example, see [42]). For filling any gap in the physical map, the cytogenetically defined markers flanking the gap can be used as references to isolate DNA from the gap by chromosome microdissection, providing markers that can be used as probes to identify large-insert clones from the gap region. The integration of a genetic linkage map divides the physical and cytogenetic maps into many sections which could greatly facilitate research in specific regions of the genome. Most importantly, integration with the genetic map places traits of economic importance onto the physical and cytogenetic maps, even if those genes are known only by phenotypes. It is unlikely with current technologies that these traits can be mapped to the physical or cytogenetic map without integration with genetic linkage maps.

A physical map is extremely useful for plant genetic and biological studies. It is a 'highway' for rapid isolation of numerous genes and for many genetic and biological studies. Once generated, the physical map can be readily accessed for many research purposes. For example, one can easily investigate genome organization, evolution, gene regulation, or gene interaction using contigs from a specific chromosomal region. If one wants to isolate the genes from a chromosomal region of interest, he can obtain a piece of a contig for that region and use it as a 'hook' to fish the genes from that region by 'gene fishing' (contig-based cDNA library screening). Since many agronomically important genes are known only by their phenotypes, map-based cloning has become an efficient and widely used strategy in isolation of such genes [4, 6, 11, 26, 34, 36, 39, 43]. However, the map-based cloning strategy is time-consuming and unreliable, because it requires many technical steps and its success is inversely related to the size and complexity of the genome [4, 6, 11, 26, 34, 36, 41, 51, 54]. The physical map will provide a powerful, simple, and rapid means to isolate numerous economically or biologically important genes by 'gene golfing'.

'Gene golfing' is a new strategy based on physical maps and that is ideally suited for rapidly cloning a large number of genes known only by phenotypes. We term this new strategy 'gene golfing' because a clone containing a target gene can be isolated as if one is playing golf. Suppose that a DNA marker has been identified, 10 cM away from the gene of interest. Using the available DNA marker, it is difficult, if not impossible, to approach the target gene by the currently used map-based cloning strategy, because repeated sequences could block chromosome walking toward the target gene. However, the gene can be readily approached by the gene golfing strategy using a developed physical map. First, the available DNA marker is used to probe the large-insert DNA library used for the physical map so as to determine the position of the DNA marker in the physical map. According to the position of the DNA marker-hybridized clone(s) in the physical map and the physical/genetic distance ratio in the region from the DNA marker to the target gene, a piece of contig containing the target gene is then taken directly from the library. To verify the relationship between the contig segment and the target gene, single-copy DNA fragments from the contig are isolated and used as probes to perform RFLP mapping against the target gene. This is the first stroke of gene golfing. If the RFLP mapping indicates that the contig segment does

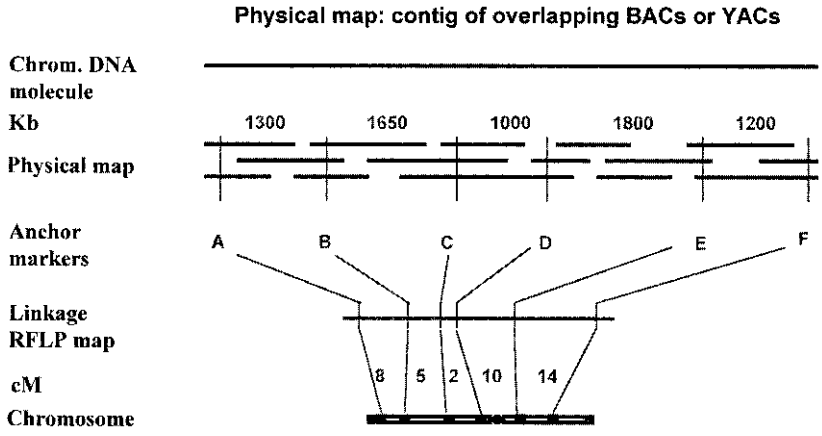

Figure 1. An ideal physical map of a rice chromosome. The overlapping bars in the physical map are large-insert genomic DNA clones such as BACs and/or YACs. Letters from A to F indicate anchor DNA markers selected from the developed RFLP linkage map. The markers are integrated into the physical map by probing its source library with them and mapped to the chromosome by FISH of the marker-associated BACs. The distances between the markers are measured in cM in the genetic map, in kb in the physical map, and in μm on the chromosome cytogenetic map.

not contain the target gene, a second golf stroke is performed. Note that the golf ball is much closer to the target hole at the second stroke than at the first one. Thus, the likelihood of putting the ball into the target hole is markedly increased at the second stroke. Additional strokes can be taken, if needed, until the clone containing the target gene is isolated. In comparison with the currently used map-based cloning, gene golfing is much simpler and faster. Map-based gene cloning often includes the following steps: gene tagging with a DNA marker, identification of DNA markers closely linked to the target gene, physical mapping with PFGE to estimate feasibility of approaching the target gene by chromosome walking using a large-insert DNA library, chromosome walking, isolation of clones containing the target gene, high resolution gene mapping, and gene identification by genetic complementation, suppression, and/or nucleotide sequence analysis [36, 51, 54]. It is not necessary to perform closely linked DNA marker identification, PFGE-based physical mapping, and chromosome walking to approach the target gene, when using the gene golfing strategy. Therefore, it can be used to approach the genes in any chromosomal regions, including those rich in repeated sequences, in which it is difficult to approach a target gene by chromosome walking.

Reviews on physical mapping of the genomes of *Caenorhabditis elegans, Arabidopsis thaliana* and man

The nematode *C. elegans* and the mustard *A. thaliana* are model species for studies of a variety of problems in genetics and biology. In addition to their short life cycles and reproductivity that is well-suited for genetic analysis, the most significant features of these two higher species for genome research are that they have small genome sizes and thus, relatively simple genome organizations [20, 29, 30]. To generate the physical maps of the genomes of these two model species and ourselves, a huge effort has been made over the past decade. From these genome research projects, many significant findings, technologies, research experiences, and lessons have been obtained, which are invaluable in physical mapping of the genomes of rice and other organisms.

Physical mapping of the C. elegans *genome*

Caenorhabditis elegans has a genome size of about 100 Mb/1C, one pair of sex chromosomes (X), and five pairs of autosomes [30]. The generation of the physical map of the genome was begun in 1984 [18, 30]. To generate the physical map of this species, cosmid clones of the *C. elegans* genomic DNA were fingerprinted and assembled systematically into contigs. A total of 17 000 cosmid clones with an average insert size of ca. 35 kb, representing about 6x haploid genome equi-

valents, was fingerprinted according to Coulson *et al.* [18] and continuous contigs were assembled from the cosmid fingerprints by fingerprint matching between cloned DNA fragments. This phase of the project led to a coverage of over 80% of the genome, with cosmids assembled into about 700 island contigs. To link the cosmid contigs, the cosmids at the ends of each contig were selected and hybridized to YACs with an average insert size of 225 kb and *vice versa* [17, 19, 30]. Hodgkin *et al.* [30] reported in late 1995 that the physical map of the nematode genome consisted of 13 YAC and cosmid contigs, separated by 7 gaps. Chromosomes X and II consisted of single contigs each, chromosomes IV and V consisted of two contigs each, chromosome III consisted of three contigs, and chromosome I consisted of four contigs. The physical map has been used for isolation of many genes and nucleotide sequencing of the genome. However, cosmid clones are the main DNA resource of these projects because of the ease of cosmid DNA purification versus the extreme difficulty of purifying YAC DNA from its host DNA [30].

Physical mapping of the A. thaliana *genome*

Arabidopsis thaliana has a genome size of 145 Mb/1C [5] and five pairs of chromosomes. To generate a genome physical map for this plant species, a strategy that was basically the same as that used for physical mapping of the *C. elegans* genome was employed. Hauge *et al.* [29] fingerprinted over 17 000 *A. thaliana* DNA cosmid clones with an average insert size of about 40 kb. In 1991, they reported that some 750 contigs representing 90–95% of the *Arabidopsis* genome were generated from the fingerprints of the 17 000 cosmids [29]. In an independent experiment to generate the physical map of the *Arabidopsis* genome, Schmidt *et al.* [45] generated a physical map of chromosome 4 by directly hybridizing mapped DNA markers to all four available *Arabidopsis* YAC libraries [20, 22, 27, 50] by colony hybridization. The average insert sizes of these four YAC libraries range from 150 to 420 kb. The markers used were 112 DNA markers mapped to chromosome 4, 20 unmapped genes, random genomic DNA fragments, sequences flanking transposable elements, and the 180 bp repeated elements. Since chromosome 4 is about 21.5 Mb in size, these markers were spaced less than 160 kb apart on average on this chromosome. The direct colony hybridization of YAC clones with the DNA markers led to 14 YAC contigs for chromosome 4. The YAC contigs were linked through chromosome walking, using the cloned ends of YACs at each end of the contigs as probes. As reported by Schmidt *et al.* [45] in late 1995, the physical map of chromosome 4 consisted of four contigs that collectively covered 90–95% of the chromosome.

Physical mapping of the human genome

The human genome is about 30-fold larger than that of *C. elegans* and consists of one pair of sex chromosomes and 22 pairs of autosomes. The basic resource to generate the physical map of the human genome was a YAC library with an average insert size of one megabase and a coverage of $10x$ haploid genome equivalents [14]. Numerous strategies were used to assemble and position these YACs onto their chromosomes of origin [14, 16, 21, 25, 31]. Basically, the physical mapping was performed by mapped or chromosome-specific DNA marker-based 'chromosome landing' of the YACs by means of PCR- and hybridization-based library screening, supplemented with YAC fingerprinting and overlapping analysis, and cross-hybridization of individual YACs. The DNA markers included sequence-tagged sites (STSs), microsatellite markers, chromosome-specific sequences, and mapped RFLP markers. The physical distances between neighboring markers on chromosomes, on average, ranged from 70 kb to less than 1000 kb. By 1995, these strategical efforts had led to physical maps of YAC contigs that covered from 75 to 93% of individual chromosomes in the genome.

Large-insert DNA libraries: cosmid, YAC and BAC, and genome physical mapping

A desirable physical map of a genome should be generated from a DNA library that can be maintained stably for a long term, has a suitably large average insert size, and can be readily used for genetic and biological studies. The large insert size, stability, and easy manipulation of the source library are important for development and subsequent use of the physical map for genetic and biological research. The physical map is generated to facilitate a variety of research. Stability of the source library represents the reliability and long-term value of the physical map. As described above, two types of large-insert DNA libraries, cosmid and YAC, have been used for generation of the physical maps for the *C. elegans* [18, 19, 30], *A. thaliana* [29, 45] and human [14, 16, 21, 25, 31] genomes. Cosmid vectors were introduced in 1978 by Collins and Hohn

[15]. The DNA fragment cloned in a cosmid vector is suitably stable for long-term maintenance, and cosmid DNA is readily purified for manipulation and application of the source library in research. However, the DNA fragment cloning capacity of cosmids is less than 50 kb, which is not well suited for generation of physical maps even though they were used to generate the physical maps of *C. elegans* [18, 30] and *A. thaliana* [29]. This is especially true for organisms with large genomes. The YAC system was introduced in 1987 by Burke *et al.* [9]. The dominant feature of this system is its DNA fragment cloning capacity of over 1000 kb [9, 14, 35, 48], which clearly facilitates generation of a physical map. However, YAC DNA is very difficult to purify from the yeast host genomic DNA because yeast has 17 chromosomes ranging from 210 to 1900 kb, which form a ladder on a pulsed-field gel (Figure 2). Many plant [7, 20, 22, 23, 24, 27, 32, 37, 48, 50] and animal YAC libraries [e.g., 12, 14, 35] have been constructed since this large fragment cloning system was introduced (for plant YAC libraries). However, all the YAC libraries but one mouse YAC library (in yeast strain 3a which is mutated in the gene RAD52) [12] were hosted in *Saccharomyces cerevisiae* recombination positive strain AB1380 due to its high transformation efficiency [8], which may lead to an *in vivo* instability of the cloned large DNA fragment during long-term maintenance.

The BACs are a relatively new, large DNA fragment cloning system that was introduced in 1992 by Shizuya *et al.* [46]. This system has the advantages of both cosmid and YAC systems [55]. It has been demonstrated that the BAC system has permitted cloning of DNA fragments of up to 350 kb [46, 52, 53], which is 7-fold larger than that of cosmids (50 kb) but it is much smaller than that of YACs (>1000 kb). Theoretically, the BAC host *Escherichia coli* is able to accommodate a BAC clone of >1.0 Mb since its own genome size is about 4.6 Mb. Since 1994, several plant BAC libraries have been constructed [13, 47, 49, 52, 53]. The average insert sizes of these BAC libraries range from 100 to 160 kb, which are much smaller than those of animal YAC libraries [e.g., 12, 14, 35], but comparable to those of most available plant YAC libraries. Most importantly, BAC DNA can be purified as easily as cosmid DNA, in contrast to the difficulty of YAC DNA purification (Figure 2). The attribute of BACs with much larger insert sizes than cosmids is very important for generation and subsequent applications of physical maps of plant and animal genomes. As seen in the *C. elegans* genome sequencing project [30], the cosmids have been used as the major DNA templates for genome sequencing while the physical map also contains YAC clones. This is because of the difficulty of YAC DNA purification from the host yeast genomic DNA. In the human genome project [14, 16, 21, 25, 31], the YAC contigs of the genome physical map have covered over 75% of the human genome. To facilitate isolation of genes from the genome and aid other studies, the YACs and/or YAC contigs containing the chromosome region of interest have to be subcloned into BAC or cosmid vectors. One of the reasons for this may be that YAC inserts are too large to manipulate but the more important reason is again the unavoidable contamination of YAC DNA with the yeast host DNA. The human genome project is clearly being run in the direction from YAC contigs to BAC or cosmid subclones to applications. In contrast, the physical maps constructed with BACs are easy to use in genome research. In addition to a strict control of a single-copy BAC in a cell [46], all BAC libraries constructed are hosted in the *E. coli rec*- strain DH10B, which makes the BACs more stable in the host, than YACs hosted in yeast recombination positive strain AB1380. Although a high level of chimaerism has been found in many YAC libraries [3, 48], a very low level of chimaerism has been found in the available BAC libraries [52, 53], which further adds the advantages of BACs over YACs for generation of high-quality physical maps.

Physical mapping of the rice genome with BACs

To facilitate rice genome research, four BAC libraries [47, 49, 53] and one YAC library [48] have been constructed for rice. The BAC libraries have average insert sizes from 120 to 150 kb, with a range from 40 to 350 kb and the YAC library has an average insert size of 350 kb with a range from 40 to over 1000 kb. Generation of a physical map of the rice genome with YACs has been described in the previous chapter. Generation of a physical map of the rice genome with BACs is discussed here.

Many strategies have been used to generate the physical maps of the genomes of *C. elegans*, *A. thaliana* and human from large-insert genomic DNA libraries. In summary, these strategies fall into three types: (1) fingerprinting of large-insert clones and contig assembly from the fingerprints [14, 18, 21, 25, 29, 30], (2) chromosome landing of large-insert clones using mapped or chromosome-specific DNA markers such as sequence-tagged sites (STSs), RFLP, and

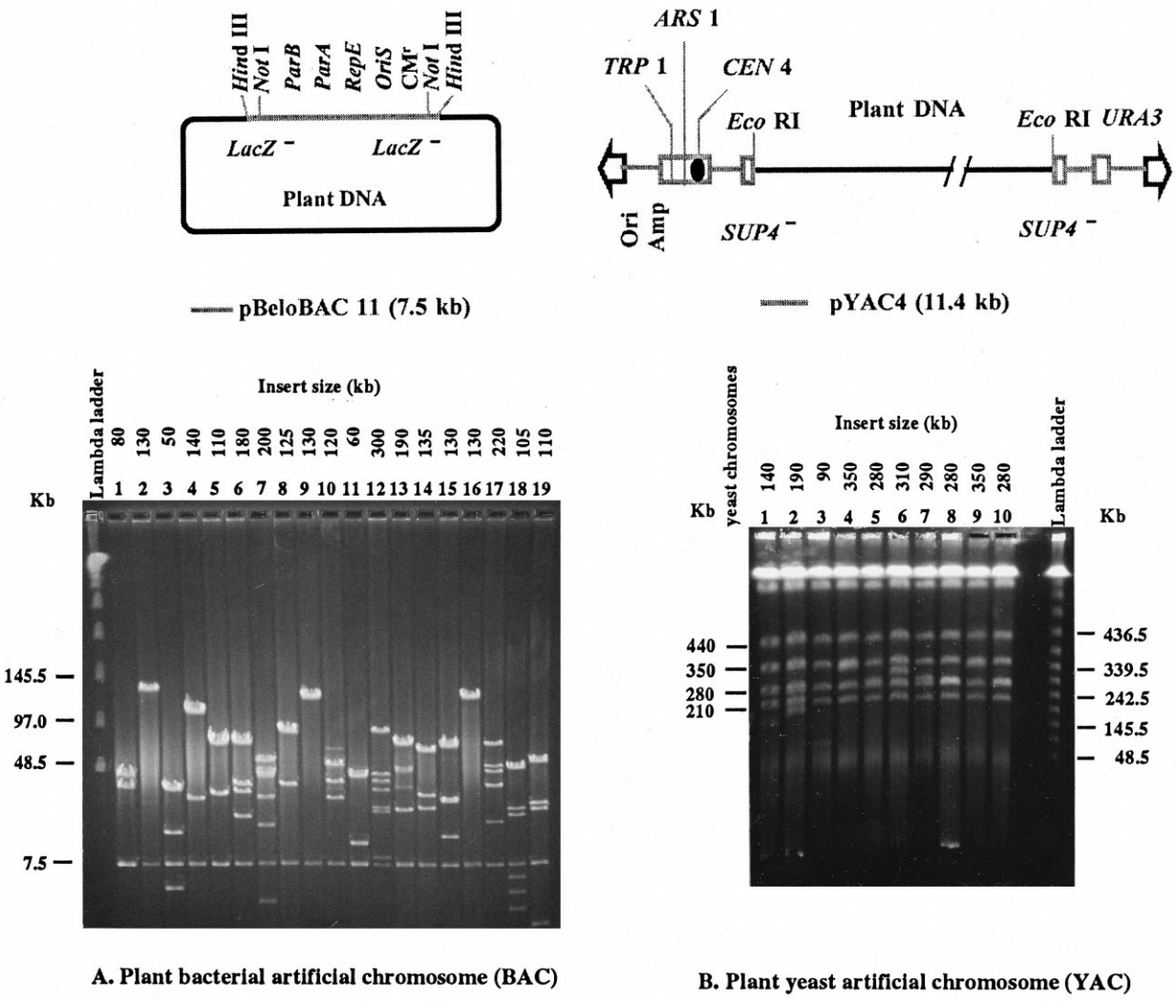

Figure 2. Examples of plant BACs and YACs. A. A BAC molecular construct (top) and rice BACs analyzed by PFGE (bottom). The BAC DNA was isolated by the conventional alkaline lysis method, digested with *Not*I to release the insert DNA from the cloning vector, and fractionated on a pulsed-field gel. The 7.5 kb band crossing all lanes of the gel is from the cloning vector pBeloBAC 11 and the remaining bands are from the cloned rice genomic DNA insert in each respective BAC. The insert size of each BAC clone is sum of sizes of all the rice DNA fragment bands in a lane. B. A YAC molecular construct (top) and tomato YACs analyzed by PFGE (bottom). Yeast spheroplasts were isolated, embedded, and lysed in low-melting-point agarose plugs. The yeast and YAC total DNA was purified within the agarose plugs and subjected to PFGE. The bands crossing all lanes are yeast chromosomes and the additional bands, or the more intense bands that comigrate with yeast chromosomes, indicate YACs.

microsatellite markers [14, 21, 16, 25, 31, 45], and (3) cross-hybridization of individual large-insert clones [14]. The first approach requires easy purification of a large number of cloned DNAs and suitably large insert sizes of the clones. The BACs are well-suited for this strategy, because of their suitably large insert sizes and easy purification of their DNAs as mentioned above. The fingerprinting strategy was also used for construction of YAC contigs in the human genome project, but human genome-specific, interspersed repeated sequences were needed as probes to detect the human genomic DNA fragments cloned in a YAC, because in this case YAC and yeast host total DNA had to be used [14, 21, 25]. For the second approach, a high-density molecular map is required. A great effort is needed to develop the high-density molecular map so that the mapped markers can be used to 'land' the marker-containing large-insert clones to their chromosomes of

origin. To assure accuracy of the physical map, the markers should be single-copy or of single-locus origin, if multiple-copy. It was discovered that duplicated DNA fragments are a universal phenomenon in plant genomes. In the rice RFLP map developed by the Japan Rice Genome Group [33], it has been discovered that ca. 6% of the mapped DNA markers are present in multiple copies in the rice genome [42]. This redundancy can mislead physical mapping by means of marker-based chromosome landing. Therefore, this strategy needs special caution. In addition to the requirement of this strategy for a high DNA marker-density map, the physically even distribution of these markers in the genome is required to complete the physical map. Unfortunately, it is difficult, if not impossible, to develop such a high-density molecular map, because recombination, which is the genetic basis of RFLP mapping, occurs unevenly along chromosomes. A good example is the physical mapping of *Arabidopsis* chromosome 4 described above [45]. To generate the physical map for this chromosome of 21.5 Mb, over 130 DNA markers, spaced on average every 160 kb on this chromosome, were used to screen the four *Arabidopsis* YAC libraries [20, 22, 27, 50]. The physical map developed using these DNA markers consisted of fourteen contigs, covering only about 85% of the chromosome. One of the reasons for this clearly resulted from the uneven distribution of the makers on the chromosome because the four YAC libraries used collectively have about $20x$ *Arabidopsis* haploid genome equivalents and one of them, the CIC YAC library, has an average insert size of 420 kb [20]. The combined YAC library should have a much higher genome coverage than 85%. This was proven by the fact most gaps between YAC contigs were filled in by chromosome walking [45]. That the three remaining gaps could not be filled in by this strategy did not seem due to absence of some YACs in the libraries, but was due to repeated sequences at the ends of the YAC contigs. This example indicates the limitation of the marker-based chromosome landing strategy for genome physical mapping. The third approach was used in human genome physical mapping, but only single-copy sequences from a clone can be used as a probe, or its repeated sequences should be blocked to prevent cross-hybridization [14].

Since BACs have suitably large-insert sizes, a low frequency of chimaeric clones, a high stability *in vivo*, and can be purified readily, the fingerprinting procedure has become the most rapid, simple and efficient approach to generate a physical map of the rice genome with BACs. Rice has a genome size of 430 Mb/1C [5] and 12 pairs of chromosomes. Two rice high density RFLP maps have been developed [10, 33], which contain a total of about 2000 DNA markers. Additionally, fluorescent *in situ* hybridization (FISH) with BAC probes has been used to map DNA marker- or gene-associated BACs to chromosomes (Islam-Faridi *et al.*; Zhang *et al.*, unpublished) [28]. The available DNA markers and developed BAC-FISH technologies have provided tools to facilitate generation and verification of the physical map.

To generate a physical map of the rice genome with BACs, we constructed two rice BAC libraries with average insert sizes of 130 and 150 kb, respectively [53]. The two BAC libraries have a total of about 22 000 clones, equivalent to $7x$ haploid rice genomes. Currently, we are developing the physical map of the rice genome from these two rice BAC libraries by the fingerprinting procedure as described by Coulson *et al.* [18]. BAC DNA was purified with the conventional alkaline lysis procedure as used for plasmid DNA purification [44], digested with *Hin*dIII, the restriction enzyme used for construction of the BAC libraries, end-labeled with ^{32}-P-dATP, digested with *Sau*3AI, and fractionated on a denaturing polyacrylamide gel. The gel was dried to chromatography paper and exposed to X-ray film. The fingerprints are directly put into a computer through a scanner and then converted into database records according to the positions of each band relative to the bands in the closest molecular-weight marker lane on a gel. In the computer, the incoming database of fingerprints are first compared against each other to assemble contigs if overlapped, and then compared against all existing database to place the incoming BACs and BAC contigs in established contigs if overlapped. The physical length of a contig in kb is estimated according to the number of restriction sites of the enzyme used for the first digestion prior to fragment end labeling (for a detailed procedure, see Chapter 10, this volume).

Feasibility

A major consideration regarding physical mapping of the rice genome with BAC fingerprinting is the variability of BAC fingerprints and influence of repeated nucleotide sequences on the fingerprinting results. Actually, these questions have been answered when the fingerprinting procedure was developed by Coulson *et al.* [18]. Through analysis of the cosmid clones containing the rDNA clusters and other multiple gene families by the fingerprinting procedure, they concluded that there

was no obvious influence of repeated sequences on the fingerprinting results. This is because each band in a cloned fragment fingerprint results from a 10 bp specificity of one 4 bp and one 6 bp restriction enzymes, which ensures that most restricted fragments containing repeat elements are either too long, or too short, to be detected in the context of the cloned fragment fingerprint. The use of polyacrylamide sequencing gels for fractionating the double enzyme-digested fragments in the procedure further minimizes the influence of repeated sequences on the cloned DNA fragment fingerprint analysis, because the sequencing gel has a resolution of one nucleotide. Any DNA fragments differing by one or more nucleotides in length can be separated on a sequencing gel, resulting in different bands in the fingerprint. In the human genome project, fingerprinting technologies have been successfully used to generate YAC contigs [14, 21, 25]. Since it is difficult to purify a large number of YAC DNAs from the yeast host genomic DNA, YAC and yeast total DNA was used. The DNA was digested with several restriction enzymes, blotted and probed with human genome-specific, interspersed repeated sequences to detect the human DNA fragments cloned in a YAC. This further demonstrated the feasibility of the fingerprinting procedure to generate physical maps of genomes. The limitation on the physical mapping of the human genome by means of the YAC fingerprinting strategy was not the feasibility of the strategy itself but the availability of interspersed repeated sequence probes that can be used to detect sufficient numbers of DNA fragments in the fingerprints. Figure 3 shows the fingerprints of some rice BACs isolated from the Teqing BAC library [53] using a rice BAC that contains centromere region-specific repeated sequences as a probe and generated according to the fingerprinting procedure described above. The result indicates the fingerprint of each BAC consists of many distinct bands and no band crossing all BAC lanes is found, even though all the BACs contain common repeated sequences. This result has further demonstrated that repeated sequences have no obvious influence on the fingerprinting results. Since each BAC fingerprint consists of many more bands than each cosmid fingerprint as revealed by Coulson *et al.* in the physical mapping of the *C. elegans* genome [18] and by Hauge *et al.* in the physical mapping of the *Arabidopsis* genome [29], the BAC fingerprinting is far more desirable than cosmid fingerprinting to generate physical maps.

Hodgkin *et al.* [30] reported that the average size of each cosmid contig in the *C. elegans* physical map

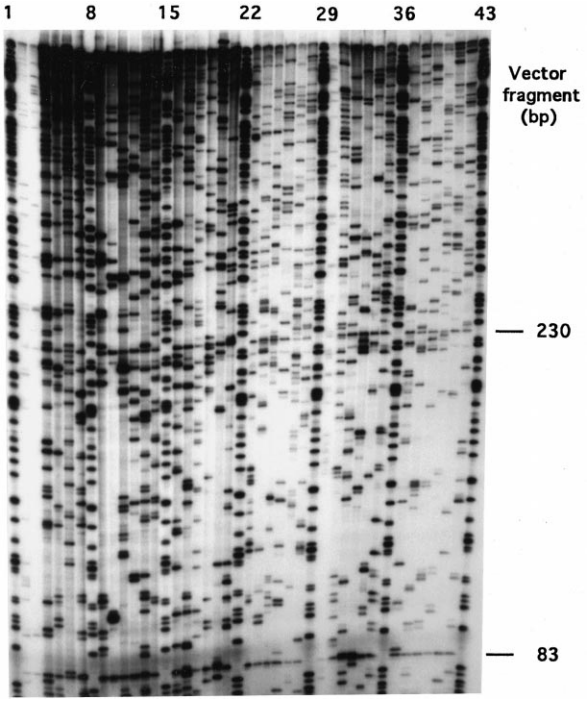

Figure 3. Fingerprints of BACs isolated from the rice cultivar Teqing BAC library using the insert DNA of a BAC containing centromere region-specific repeated sequences as a probe. High-density colony filters were prepared from the rice cultivar Teqing BAC library and probed with the insert DNA of the BAC containing centromere region-specific repeated sequences. More than 1000 BAC clones were isolated, and 36 of them were randomly selected for fingerprinting. BAC DNA was isolated, digested with *Hin*dIII, end-labeled with ^{32}P-dATP, digested with *Sau*3AI, and fractionated on a 4% denaturing polyacrylamide gel. The gel was dried onto chromatography paper and autoradiographed overnight. From left to right: Lanes 1, 8, 15, 22, 29, 36, and 43 are λ/*Sau*3AI DNA markers and the remaining lanes are BACs. Less DNA was loaded in lanes 2 and 3 than in other lanes. The two bands crossing all BAC lanes are vector sequences flanking the *Hin*dIII cloning site in pBeloBAC 11.

was about 200 kb and the longest continuous contig covered about 2000 kb. Since the rice BAC libraries have average insert sizes of 120 to 150 kb, which is about 4-fold larger than that of the cosmid library used in the physical mapping of the *C. elegans* genome, it is highly likely that the sizes of contigs that can be constructed by the BAC fingerprints will range from several hundreds to 8000 kb (2000×4), with an average size of 800 kb (200×4), if the size of a contig is proportional to the insert size of the clones used. Because 1 cM corresponds, on average, to about 300 kb in the rice genome, the BAC contigs could range from 1.0 to 25 cM in size. In addition, the BAC libraries have an insert size range from 40 to 350 kb,

which is much larger than that from 20 to 50 kb of the cosmids used in the *C. elegans* genome mapping. The large insert size range of BAC library may help improve its representation for the genome because of less stringent size selection in BAC cloning than that in cosmid cloning. Variation in insert size may also increase the size of BAC contigs because the clones having insert sizes larger than the average insert size of the library may play a role to link neighboring contigs. These hypotheses has been proven by the fact that the cosmid islands resulting from the fingerprinting were bridged by the YACs with an average insert size of 225 kb to generate the physical map of the *C. elegans* genome [17, 19, 30]. If the BACs with the largest insert sizes (e.g. 350 kb) are compared to the cosmids with the largest insert sizes (e.g. 50 kb), the average size of the BAC contigs could be as large as 1400 kb (200 kb of the average cosmid contig size $\times$ 7) and the longest BAC contig could span about 14 000 kb (2000×7). Therefore, the contigs constructed by BAC fingerprints could be up to 46 cM (14 000/300). The chromosome landing of BACs and BAC contigs facilitated by mapped RFLP marker-based hybridization and BAC-FISH will help inter-link the contigs constructed by BAC fingerprinting to create fewer, larger contigs.

Probability

The probability of success to generate a physical map of a genome by random clone fingerprinting depends on many factors, the most important of which include the average insert size of the source library, number of clones contained in the library, and the genome size of the organism. Assuming that the clones in a library are randomly distributed in the genome, the probability to complete a continuous overlapping contig could be calculated as follows.

Let *x* be the size (in Mb) of a contig to be generated and *i* be the average insert size (in Mb) of a source library, and assume 50% overlapping rate between neighboring clones in the contig, then the expected number of clones needed in the contig (*t*) would be:

$$t = \frac{2x - i}{i}$$

Let *c* be the number of clones contained in the source library and *s* be the genome size (in Mb) of the organism under study, the probability of completing the contig (P) would be:

$$P = \left(1 - \left(1 - \frac{i}{s}\right)^{c}\right)^{t}$$

To test these equations, the physical mapping of the *A. thaliana* and *C. elegans* genomes by cosmid fingerprinting are used as examples. In the case of *C. elegans*, the realized average size of the cosmid contigs constructed was about 200 kb [30]. If the above equations are suitable to estimate the probability of completing a contig of a certain size, the *P* value of completing the 200 kb cosmid contig calculated by the equations should be close to 1.0 (100%). Since $x = 0.2$ Mb, $i = 0.035$ Mb, $c = 17\,000$, and $s = 100$ Mb, then $P = 0.98$. In the case of *A. thaliana*, because some 750 cosmid contigs were constructed, covering 90–95% [29] of the genome of 145 Mb [5], the realized average size of contigs should be about 180 kb. Since $i = 0.04$ Mb, and $c = 17\,000$, then $P = 0.93$. In the cases of both *C. elegans* and *A. thaliana*, the *P* values are close to 1.0. These simple calculations have demonstrated that the above formulas are suitable to estimate the probability of completing a genome contig by random clone fingerprinting.

Figure 4 plots the probability of completing a continuous contig of 1.0 Mb of the rice genome with BACs versus the average insert sizes of BAC libraries and the number of clones needed to build up the contig according to the above equations. The probability varies with the average insert sizes and the clone numbers of the source BAC libraries. Interestingly, the probability of completing a contig of 1.0 Mb does not increase much when the average insert sizes of the libraries are larger than 200 kb if the libraries have library sizes of over 99% probability (over $4.5x$ haploid genome equivalents) to cover any sequence in the genome, even though the number of clones needed in the 1.0 Mb contig decreases with increase of the insert sizes of clones. However, the insert size of library is critical to complete the 1.0 Mb contig when it is smaller than 200 kb. When the average insert size of a library is 50 kb or smaller, the probability of completing the 1.0 Mb contig with it is extremely low even though it has $10x$ rice haploid genome equivalents. Therefore, a library with a suitably large average insert size is required to complete the 1.0 Mb contig. While it is desirable to have a large-insert BAC library for physical mapping of the rice genome, it is suitable to have a BAC library with an average insert size between 120 and 250 kb if the library has an over $7x$ haploid genome coverage. Since the average size of contigs constructed by fingerprinting of cosmids with an average insert size of 35 kb in the physical mapping of the *C. elegans* genome as described above was about 200 kb [30], it is practical to generate contigs with an average size of 1.0 Mb from

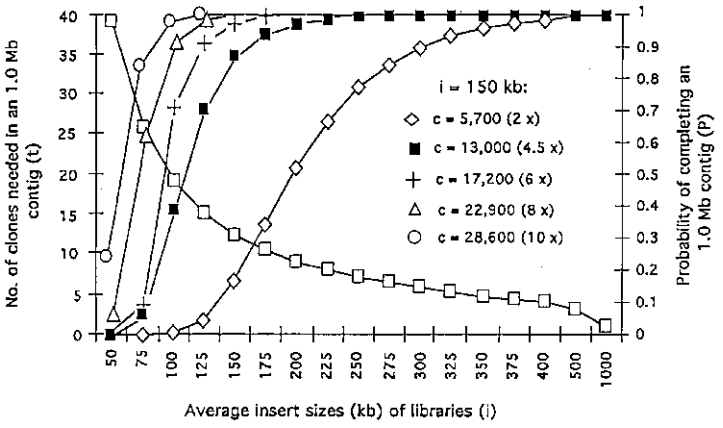

Figure 4. Generation of BAC contigs for the rice genome: average insert size of a source BAC library, number of clones in the library, number of clones needed to construct an 1.0 Mb contig, and probability of completing the contig. The genome equivalents are based on $i = 150$ kb, $s = 430$ Mb, and the number of clones in a library. To attain a 99% probability of containing any particular genomic DNA sequence, the library must cover more than $4.5x$ haploid genome equivalents.

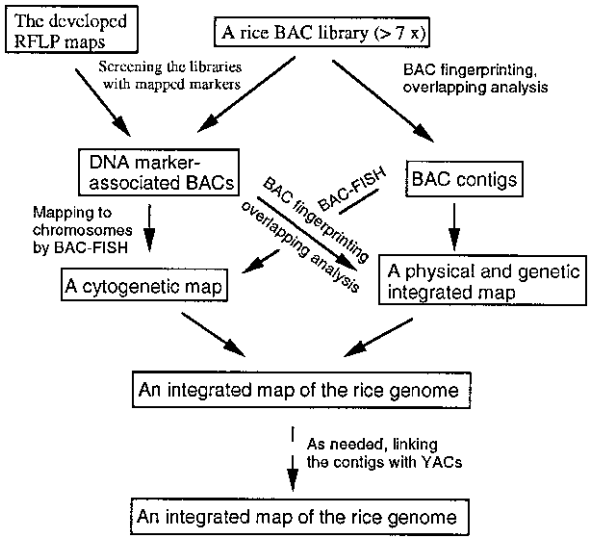

Figure 5. A strategy for physical mapping of the rice genome with BACs.

the BACs with an average insert size of 120–250 kb. In the rice genome, because 1.0 Mb is equivalent to about 3.5 cM [10, 33], the 1.0 Mb contig could contain 2–4 DNA markers mapped to the rice maps [10, 33] and thus could be linked with other contigs into larger continuous contigs.

Strategies

For generation of physical maps, BAC fingerprinting is a new strategy, but it can be predicted that it will be successful, because cosmid fingerprinting has been successfully used in the physical mapping of the genomes of *C. elegans* [30] and *A. thaliana* [29]. However, supplemental strategies are required to complete the physical map of the rice genome. These strategies include chromosome landing of BACs or BAC contigs by BAC-FISH and/or mapped DNA marker-based hybridization. Thus, the physical map developed is integrated with and reinforced by the genetic and cytogenetic maps of the rice genome. The screening of the BAC library with mapped DNA markers will bridge the developed RFLP maps with the physical and cytogenetic maps, and provide landmarks for them. Importantly, the genes mapped to the RFLP maps can be positioned on the physical map and chromosomes. The Japan Rice Genome Group has found that ca. 6% of their DNA markers are duplicated in the rice genome [42]. These DNA markers could lead to problems for placement of the BAC contigs to chromosomes by DNA marker-based chromosome landing, in that one marker could hybridize to two or more contigs, or a single contig could contain DNA sequences mapped to different locations on the linkage map. BAC-FISH, a proven tool for molecular cytogenetic mapping of the plant genomes, provides a straightforward approach to these problems and will play a special role in mapping and maintaining accuracy of the physical map. The direct mapping of the RFLP marker-associated BACs will provide a reliable guide for the construction of continuous contigs from the BAC fingerprints. The FISH of a BAC selected from a BAC contig will directly place the BAC contig to a specific chromosome region

and establish linkage relationships of this BAC contig to other BAC contigs.

It was found in the genome mapping project of *C. elegans* that some DNA sequences in certain genome regions were difficult to clone in cosmid vectors [30], which consequently resulted in some of the gaps in the physical map constructed by cosmid fingerprinting. This may also become a problem in physical mapping of the rice genome with BACs. However, this problem would not be serious as that encountered in the physical mapping of the *C. elegans* genome with cosmid fingerprinting since BAC vectors have much less stringent size selection than cosmid vectors, and as a result, BAC libraries have much larger insert size ranges than cosmid libraries. This argument has been reinforced by the fact that the cosmid cloning problem could be overcome by vectors with smaller inserts in some cases in the *C. elegans* genome project [30]. The BAC cloning vector pBeloBAC 11 has three cloning sites, *Bam*HI, *Hin*dIII, and *Sph*I [55]. A source BAC library constructed in different cloning sites may minimize the gaps in a physical map constructed with BAC fingerprinting. Alternatively, a rice YAC library will help link contigs generated with BAC fingerprints because YACs have a large insert cloning capacity, a large insert size range, and are hosted in yeast - a different system from the bacterial host of BACs. Practically, YACs were successfully used to link the cosmid contigs in the *C. elegans* physical map [18, 19, 30].

Practice

The key to physical genome mapping by the BAC fingerprinting strategy is the practical ability to isolate and fingerprint a large number of BAC DNAs. To this end, we have optimized procedures, which are suitable for automation, to purify and fingerprint a large number of BAC DNAs. These procedures routinely permit one technician to isolate over 200 BAC DNAs, or fingerprint about 100 BACs (run two sequencing gels of 56 wells) within a working day. Therefore, a technician can comfortably isolate and fingerprint 200 BAC DNAs within five working days and 20 000 BAC DNAs within two years, which represents over $6x$ rice haploid genome equivalents. If BAC DNA is isolated automatically, a technician will be able to fingerprint at least 300 BACs within a week. Therefore, the physical map of the rice genome could be completed within two to three years, which is practical and even faster than the time needed to construct a high-density RFLP map.

Perspectives on physical mapping of the rice genome with BACs

The RFLP analysis of genomes has revolutionized genetic and biological studies of organisms in the past 17 years. RFLP maps have been developed for almost all major crops and model plant species, and many genes of agronomic importance have been mapped. In rice, two high-density RFLP maps have been developed [10, 33]. The initial goals of these efforts are use of the developed RFLP maps to tag, map, clone, and manipulate genes and quantitative trait loci (QTLs) of economic importance for crop genetic improvement, and to facilitate studies in genome organization and evolution [38]. The RFLP maps have provided a great framework for these studies, but it is time-consuming and unreliable to perform such investigations with the current RFLP maps, because they require many technical steps and their success is inversely related to the size and complexity of the genome. The question now is where we should go: continue to develop higher-density molecular maps, or stay where we are and develop readily accessible, simple systems, i.e. physical maps for genome research. Should we go ahead clone the mapped genes by map-based cloning, or develop integrated maps of the genomes and then use them to clone the mapped genes easily and rapidly by gene golfing as described above? The integrated physical, genetic, and cytogenetic map of the rice genome will provide a 'highway' for isolation of a large number of genes and for many genetic and biological studies. The physical map can be readily accessed and will provide virtually unlimited numbers of DNA markers from any chromosome region or closely linked to any gene of interest. In addition to its economic importance, rice is also a model species for grass crop genome research. Recent comparative mapping has demonstrated that the gene content and order are highly conserved in the grass genomes [1, 2, 40]. Therefore, the physical map of the rice genome could also be used for isolation of agronomically important genes in other grass crops. Since the BAC fingerprinting technology is simple, rapid, economical, and does not require such a high-density RFLP map as the marker-based chromosome landing technology, the BAC fingerprinting used in physical mapping of the rice genome will be also applicable to physical mapping of the genomes of other crops.

Acknowledgments

The rice genome research project and the chapter were supported by the Rockefeller Foundation (RF95001#334 to H.B.Z. and R.A.W.) and the Rockefeller Foundation Technology Transfer Fellowships in Rice Biotechnology (RF94001#305 to H.B.Z.). H.B.Z. thanks Drs David M. Stelly and Charles F. Crane, Texas A&M University, Texas, USA, for critically discussing and reading the manuscript.

References

1. Ahn S, Anderson JA, Sorrells ME, Tanksley SD: Homologous relationships of rice, wheat and maize chromosomes. Mol Gen Genet 241: 483–490 (1993).
2. Ahn S, Tanksley SD: Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci USA 90: 7980–7984 (1993).
3. Anderson C: Genome shortcut leads to problems. Science 259: 1684–1687 (1993).
4. Arondel V, Lemieux B, Hwang I, Gibson S, Goodman HN, Somerville CR: Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Arabidopsis*. Science 258: 1353–1355 (1992).
5. Arumuganathan K, Earle ED: Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9: 208–218 (1991).
6. Bent AF, Kunkel BN, Dahlbeck D, Brown KL, Schmidt R, Giraudat J, Leung J, Staskawicz BJ: *RPS2* of *Arabidopsis thaliana*: A leucine-rich repeated class of plant disease resistance genes. Science 265: 1855–1860 (1994).
7. Bonnema G, Hontelez J, Verkerk R, Zhang YQ, van Daelen R, van Kammen A, Zabel P: An improved method of partially digesting plant megabase DNA suitable for YAC cloning: application to the construction of a 5.5 genome equivalent YAC library of tomato. Plant J 9: 125–133 (1996).
8. Burgess PM, Percival KJ: Transformation of yeast spheroplasts without cell fusion. Anal Biochem 163: 391–397 (1987).
9. Burke DT, Carle GF, Olson MV: Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors: Science 236: 805–811 (1987).
10. Causse MA, Fulton TM, Cho YG, Ahn AN, Chunwongse J, Wu K, Xiao J, Yu Z, Ronald PC, Harrington SE, Second G, McCouch SR, Tanksley SD: Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138: 1251–1274 (1994).
11. Chang C, Kwok SF, Bleecker AB, Meyerowitz EM: *Arabidopsis* ethylene-response gene *ETR1*: similarity of product to two-component regulators. Science 252: 539–544 (1993).
12. Chartier FL, Keer JT, Sutcliffe MJ, Henriques DA, Mileham P, Brown DM: Construction of a mouse artificial chromosome library in a recombination-deficient strain of yeast. Nature Genet 1: 132–136 (1992).
13. Choi S, Creelman RA, Mullet JE, Wing RA: Construction and characterization of a bacterial artificial chromosome library of *Arabidopsis thaliana*. Plant Mol Biol Rep 13: 124–129 (1995).
14. Chumakov IM, Rigault P, Gall IL, *et al.*: A YAC contig map of the human genome. Nature 377: 157–298 (1995).
15. Collins J, Hohn B: Cosmids: A type of plasmid gene-cloning vector that is packagable *in vitro* in bacteriophage lambda heads. Proc Natl Acad Sci USA 75: 4242 (1978).
16. Collins JE, Cole CG, Smink LJ, *et al.*: A high-density YAC contig map of human chromosome 22. Nature 377: 367–371 (1995).
17. Coulson A, Kozono Y, Lutterbach B, Shownkeen R, Sulton J, Waterson R: YACs and the *C. elegans* genome. BioEssays 13: 413–417 (1991).
18. Coulson A, Sulton J, S Brenner S, Karn J: Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. Proc Natl Acad Sci USA 83: 7821–7825 (1986).
19. Coulson A, Waterson R, Kiff J, Sulton J, Kohara Y: Genome linking with yeast artificial chromosomes. Nature 335: 184–186 (1988).
20. Creusot F, Fouilloux E, Dron M, Lafeuriel J, Picard G, Billaut A, Le Paslier D, Cohen D, Chaboute M, Durr A, Fleck J, Gigot C, Camilleri C, Bellini C, Caboche M, Bouchez B: The CIC library: a large insert YAC library for genome mapping in *Arabidopsis thaliana*. Plant J 8: 763–770 (1995).
21. Doggett NA, Goodwin LA, Tesmer JG *et al.*: An integrated physical map of human chromosome 16. Nature 377: 335–366 (1995).
22. Ecker JR: PFGE and YAC analysis of the *Arabidopsis* genome. Methods 1: 186–194 (1992).
23. Edwards KJ, Thompson H, Edwards D, de Saizieu A, Sparks C, Thompson JA, Greenland AJ, Eyers M, Schuch W: Construction and characterization of a yeast artificial chromosome library containing three haploid maize genome equivalents. Plant Mol Biol 19: 299–308 (1992).
24. Eyers M, Edwards K, Schuch W: Construction and characterization of a yeast artificial chromosome library containing two haploid *Beta vulgaris* L. genome equivalents. Gene 121: 195–201 (1992).
25. Germmill RM, Chumakov I, Scott P *et al.*: A second-generation YAC contig of human chromosome 3. Nature 377: 299–320 (1995).
26. Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, Goodman HM: Isolation of the *Arabidopsis 3AB1* gene by positional cloning. Plant Cell 4: 1251–1251 (1992).
27. Grill E, Somerville C: Construction and characterization of a yeast artificial chromosome library of *Arabidopsis* which is suitable for chromosome walking. Mol Gen Genet 226: 484–490 (1991).
28. Hanson RE, Zwick MS, Choi S, Islam-Faridi MN, McKnight TD, Wing RA, Price HJ, Stelly DM: Fluorescent in situ hybridization of a bacterial artificial chromosome. Genome 38: 646–651 (1995).
29. Hauge BM, Giraudat J, Hanley S, Hwang I, Kohchi T, Goodman HM: Physical mapping of the *Arabidopsis* genome and its applications. In: Hermann RG, Larkins B (eds) Plant Molecular Biology 2, pp. 239–248. Plenum Press, New York (1991).
30. Hodgkin J, Plasterk RHA, Waterston RH: The nematode *Caenorhabditis elegans* and its genome. Science 270: 410–414 (1995).
31. Kauter K, Montgomery K, Yoon S-J, *et al.*: A second-generation YAC contig of human chromosome 12. Nature 377: 321–334 (1995).
32. Kleine M, Michalek W, Graner A, Herrmann RG, Jung C: Construction of a barley (*Hordeum vulgare* L.) YAC library and isolation of a *Hor1* specific clone. Mol Gen Genet 240: 255–272 (1993).
33. Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shomura A, Shimizu T, Lin S-Y, Inoue T,

Fukuta A, Shimano T, Kuboki Y, Toyama, Miyamoto Y, Kirihara T, Hayasaka K, Miyao A, Monna L, Zhong HS, Tamura Y, Wang Z-X, Momma T, Umehara Y, Yano M, Sasaki T, Minobe Y: A 300 kilobase interval genetic map of rice including 883 expressed sequences. Nature Genet 8: 365–372 (1994).

34. Leung J, Bouvier-Durand M, Morris P-C, Guerrier D, Chefdor F, Giraudat J: *Arabidopsis* ABA response gene *ABIl*: features of a calcium-modulated protein phosphatase. Science 264: 1448–1452 (1994).
35. Libert F, Lefort A, Okimoto R, Womack J, Georgies M: Construction of a bovine genomic library of large yeast artificial chromosome clones. Genomics 18: 270–276 (1993).
36. Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD: Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 252: 1432–1438 (1993).
37. Martin GB, Ganal MW, Tanksley SD: Construction of a yeast artificial chromosome library of tomato and identification of cloned segments linked to two disease resistance loci. Mol Gen Genet 223: 25–32 (1992).
38. McCouch SR, Tanksley SD: Development and use of restriction fragment length polymorphism in rice breeding and genetics. In: Khush GS, Toenniessen GH (eds) Rice Biotechnology, pp. 109–124. CAB International Wallingford, UK (1991).
39. Meyer K, Leube MP, Grill E: A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*. Science 154: 1452–1455 (1994).
40. Moore G, Devos KM, Wang Z, Gale MD: Grasses, line up and form a circle. Curr Biol 5: 737–739 (1995).
41. Putterill J, Robson F, Lee K, Coupland G: Chromosome walking with YAC clones in *Arabidopsis*: isolation of 1700 kb of contiguous DNA on chromosome 5, including a 300 kb region containing the flowering-time gene *CO*. Mol Gen Genet 239: 145–157 (1993).
42. Rice Genome Research Program NIAR Staff: Current activities in genetic mapping group. Rice Genome 4: 3 (1995).
43. Sakai H, Medrano LJ, Meyerowitz EM: Role of *SUPERMAN* in maintaining *Arabidopsis* floral whole boundaries. Nature 375: 199–203 (1995).
44. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
45. Schmidt R, West J, Love K, Lenehan Z, Lister C, Thompson HY, Bouchez D, Dean C: Physical map and organization of *Arabidopsis thaliana* chromosome 4. Science 270: 480–483 (1995).
46. Shizuya H, Birren B, Kim U-J, Mancino V, Slepak T, Tachiiri Y, Simon M: Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc Natl Acad Sci USA 89: 8794–8797 (1992).
47. Tao Q, Zhao H, Qiu L, Hong G: Construction of a full bacterial artificial chromosome (BAC) library of *Oryza sativa* genome. Cell Res 4: 127–133 (1994).
48. Umehara Y, Miyazaki A, Tanoue H, Yasukochi Y, Saji S, Otsuki Y, Fujimura T, Kurata N, Minobe Y: Construction and characterization of rice YAC library for physical mapping. Plant Breed 1: 79–89 (1995).
49. Wang G, Holsten TE, Song W-Y, Wang H-P, Ronald PC: Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa-21* disease resistance locus. Plant J 7: 525–533 (1995).
50. Ward ER, Jen GC: Isolation of single-copy-sequence clones from a yeast artificial chromosome library of randomly-sheared *Arabidopsis thaliana* DNA. Plant Mol Biol 14: 561–568 (1990).
51. Wing RA, Zhang H-B, Tanksley SD: Map-based cloning in crop plants: tomato as a model system: I. Genetic and physical mapping of *jointless*. Mol Gen Genet 242: 681–688 (1994).
52. Woo SS, Jiang J, Gill BS, Paterson AH, Wing RA: Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. Nucl Acids Res 22: 4922–4931 (1994).
53. Zhang H-B, Choi S, Woo S-S, Li Z, Wing RA: Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. Mol Breed 2: 11–24 (1996).
54. Zhang H-B, Martin GB, Tanksley SD, Wing RA: Map-based cloning in crop plants: Tomato as a model system II. Isolation and characterization of a set of overlapping YACs encompassing the *jointless* locus. Mol Gen Genet 244: 513–521 (1994).
55. Zhang H-B, Woo S-S, Wing RA: BAC, YAC and cosmid library construction. In: Foster GD, Twell D (eds) Plant Gene Isolation: Principles and Practice, pp. 75–99. John Wiley, New York (1996).