Theor Appl Genet (2005) 111: 1596–1607
DOI 10.1007/s00122-005-0091-3

ORIGINAL PAPER

**Jetty S. S. Ammiraju · Yeisoo Yu · Meizhong Luo
Dave Kudrna · HyeRan Kim · Jose L. Goicoechea
Yuichi Katayose · Takashi Matsumoto · Jianzhong Wu
Takuji Sasaki · Rod A. Wing**

Random sheared fosmid library as a new genomic tool to accelerate complete finishing of rice (*Oryza sativa* spp. Nipponbare) genome sequence: sequencing of gap-specific fosmid clones uncovers new euchromatic portions of the genome

Received: 6 May 2005 / Accepted: 15 August 2005 / Published online: 1 October 2005

Abstract The International Rice Genome Sequencing Project has recently announced the high-quality finished sequence that covers nearly 95% of the *japonica* rice genome representing 370 Mbp. Nevertheless, the current physical map of *japonica* rice contains 62 physical gaps corresponding to approximately 5% of the genome, that have not been identified/represented in the comprehensive array of publicly available BAC, PAC and other genomic library resources. Without finishing these gaps, it is impossible to identify the complete complement of genes encoded by rice genome and will also leave us ignorant of some 5% of the genome and its unknown functions. In this article, we report the construction and characterization of a tenfold redundant, 40 kbp insert fosmid library generated by random mechanical shearing. We demonstrated its utility in refining the physical map of rice by identifying and in silico mapping 22 gap-specific fosmid clones with particular emphasis on chromosomes 1, 2, 6, 7, 8, 9 and 10. Further sequencing of 12 of the gap-specific fosmid clones uncovered unique rice genome sequence that was not previously reported in the finished IRGSP sequence and emphasizes the need to complete finishing of the rice genome.

Communicated by Q. Zhang

Fosmid library reported here is publicly available from our web site http://www.genome.arizona.edu/orders

J. S. S. Ammiraju · Y. Yu · M. Luo · D. Kudrna
H. Kim · J. L. Goicoechea · R. A. Wing (✉)
Department of Plant Sciences and BIO5 Institute,
Arizona Genomics Institute, The University of Arizona,
Tucson, AZ, 85721 USA
E-mail: rwing@ag.arizona.edu
Tel.: +1-520-6269595
Fax: +1-520-6211259

Y. Katayose · T. Matsumoto · J. Wu · T. Sasaki
National Institute of Agrobiological Sciences,
Tsukuba, Ibaraki, Japan

Introduction

The International Rice Genome Sequencing Project (IRGSP; http://rgp.dna.a rc.go.jp/IRGSP), a consortium of ten countries has recently announced a finished genome covering 95% of the rice genome. The IRGSP followed a gold-standard approach for rice genome sequencing that relied on first, establishing an exhaustive sequence-ready integrated genetic and physical map comprised of BAC fingerprints assembled into contigs using software FPC (Soderlund et al. 2000; Chen et al. 2002) and sequence-tagged connectors (Venter et al. 1996; Mao et al. 2000). Second, genetically anchored seed BACs or PACs are shotgun sequenced and finished. Minimally overlapping clones flanking the seed BAC/PACs are then selected based on the sequence-ready physical map and sequenced as above. This clone-by-clone approach is reiterated over and over until the complete genome is sequenced. Each base pair was sequenced ten times on an average, ensuring an error rate of less than one base in 10,000. The finished sequence now has 370 Mbp contig size.

Nevertheless, the present IRGSP integrated genetic and physical map of *japonica* rice still contains about 62 physical gaps that depending on the chromosome, range in size from 10 to 100 kbp in euchromatic regions and up to 1 Mbp in centromeric regions, corresponding to nearly 5% of the missed segments in the reported rice genome sequence. These include 9 centromeric, 17 telomeric and 36 arm-specific gaps that are distributed on several chromosomes, with an approximately estimated total size of 18.1 Mbp (IRGSP, 2005). Although, the rice genome is now considered ‘‘complete’’ and provides

unprecedented opportunities for a variety of research needs, it is imperative to identify and characterize the remaining (unusual) segments missing in the rice genome sequence. Each missed segment potentially contains an agronomically important gene and/or noncoding heterochromatic regions that might be essential for gene regulation. Without finishing these gaps, it is impossible to identify the complete complement of genes encoded by rice genome and will also leave us ignorant of some 5% of the genome and its unknown functions.

Gaps in the physical maps can be caused by at least two known phenomenon: (1) nonclonable or unstable segments in *Escherichia coli* hosts (Hagan and Warren 1982; Schroth and Ho 1995; Kang and Cox 1996; Ravin and Ravin 1999; Razon et al. 2001), which can probably be overcome by using an alternative cloning and host systems (Leem et al. 2004); (2) Cloning bias (Chen et al. 2002; Thorstenson et al. 1998). Due to nonrandom distribution of restriction sites for any enzyme in genomic DNA, libraries constructed with restriction enzymes will always have over- or underrepresentation of some regions and no coverage of a small fraction of the genome. These regions of no or low coverage are revealed as gaps in the BAC contig maps produced by analysis of restriction fingerprints of BAC clones.

Now, a major thrust of the IRGSP is to further refine the physical map of *japonica* rice by closing these gaps to accelerate the process of identifying and characterizing ''missing'' sequence in the rice genome. However, despite intensive e orts using seven di erent genomic library resources that include four deep coverage BAC libraries (made with restriction enzymes *Eco*RI, two libraries of *Hin*dIII, and *Mbo*I), a PAC library (*Sau3A*I) (Chen et al. 2002; Baba et al. 2000), two 10 kbp libraries (made with *Hae*III and *Sau*3AI), a large number of physical gaps remain. Therefore, there is a pressing need to generate new large insert genomic resources that are not based on enzyme digestion to fill the gaps caused by cloning bias and newer approaches to characterize unfinished portions of rice genome sequence.

A random genomic DNA library is ideally constructed from DNA that has been mechanically sheared. Unfortunately, techniques for cloning sheared DNA fragments in the 100–200 kbp size range are highly challenging. As an alternative, the fosmid cloning system is rapidly emerging as a method of choice to rapidly create high titer ''mini-BAC libraries'' with an average insert size of 40 kbp from a small amount of genomic DNA with cloning e ciencies comparable to cosmids (Kim et al. 1992). Although the genomic inserts are smaller, in comparison to BAC inserts, they still have been shown to be quite useful for whole genome physical mapping (Kim et al. 1995; Gingrich et al. 1996; Fitz-Gibbon et al. 1997; Kim et al. 2003), large-scale genomic DNA sequencing (Kim et al. 1995; Kawarabayasi et al. 1998), and for studies related to structure and organization of the genomes (Hughes et al. 1997; Hoyer et al. 1998a, b; Vergin et al. 1998; Millikan et al. 1999; Quaiser et al. 2002).

In this article we report the construction and characterization of a tenfold random sheared fosmid library of *japonica* rice variety Nipponbare, and demonstrate its e ectiveness in closing and extending physical gaps in the IRGSP physical map with particular emphasis on chromosomes 1, 2, 6, 7, 8, 9 and 10. Further we propose refinement of existing strategies that not only would e ciently improve the rice physical map but has broad application to any sequencing project that adopts a clone-by-clone sequencing strategy.

Materials and methods

Fosmid library construction

Megabase-size DNA was prepared by embedding nuclei from 30-day-old rice (*O. sativa* ssp. *japonica* cv. Nipponbare) in 0.5% low-melting agarose, followed by nuclei lysis in the presence of detergent and proteinase-K as described by Luo and Wing (2003). DNA in agarsoe was sheared using a freeze–thaw cycle as follows. A total of 15 DNA plugs were transferred to 1.5 ml individual eppendorf tubes containing 200 µl of TE bu er, one plug each tube. These tubes were frozen by dipping in the liquid Nitrogen 10–20 s and transferred immediately to a 45°C water bath to thaw for approximately about 2–3 min. This process was repeated for 22 times. After shearing, the DNA plugs were melted at 70°C and then immediately transferred to a 45°C water bath and treated with 1 U of Gelase (Epicentre, Madison, WI, USA) for every 100 µl of melted agarose for 1.5 h. These tubes were transferred to ice for 5 min and then centrifuged at 13,000 rpm at 4°C for 10 min. The supernatant fluid was ethanol precipitated followed by a 70% ethanol wash. The pellets were combined into a single tube and the DNA concentration was estimated by comparison with DNA size standards on a 1% mini agarose gel. The ends of the sheared DNA (~10 µg) were polished using a Blunt end repair kit (Epicentre) according to manufacturer's instructions, for 45 min at room temperature followed by heat inactivation at 70°C for 10 min. The sheared, end-repaired DNA was run on a 1% CHEF agarose gel using the following conditions: 16-h run time, 14°C, 5–15 s pulse, 120° angle in 0.5x TBE at 6 v/cm. DNA size standards were the 40 kbp T7 DNA marker from Epicentre Fosmid cloning kit (#CCFOS110) and the PFG midrange marker I (New England Biolabs, Beverly, MA, USA). DNA fragments in the size range of 25–45 kbp were excised from the gel and the agarose was removed by melting at 70°C followed by the addition of gelase as described above. The DNA fragments were precipitated and finally dissolved in 10 µl of sterile water. DNA concentration was estimated using known standards on a 1% agarose gel. The insert DNA was ligated to pCC1FOS copy control vector at a 10:1 (vector-insert) ratio for 2 h at room temperature using fast link ligase (Epicentre). The ligation reaction was heat inactivated at 70°C for 10 min

and then packaged into phage particles using MaxPlax lambda packaging extracts (Epicentre). The packaged fosmid clones were titered by serial dilutions and infected to 100 µl of Phage T1 resistant EPI300 *E. coli* plating strain cells grown to an A600 of 1.0 for 20 min at 37°C. Cells were plated onto LB/agar Q-trays (Genetix Ltd., New Milton, Hampshire, UK) containing 12.5 µg/ml chloroamphenicol. Colonies were picked and arrayed into 384-well microtiter plates (Genetix Ltd.) containing freezing media using a Genetix Q-bot (Genetix Ltd.). The library was replicated and stored at −80°C.

Fosmid clone characterization

Insert-sizing was performed by analyzing a total of 300 randomly picked clones across the library. Plasmids were isolated from 1.2 ml 2XYT cultures of individual clones that were grown overnight in 96 deep-well plates using a semiautomated liquid-handling machine (Tomtec Quadra 96, model 320, USA) with a simplified high-throughput method (Hye-Ran Kim and Rod Wing, unpublished) that is based on conventional alkaline lysis methods (Sambrook et al. 2001). Samples were digested with the restriction enzyme *Not*I, and electrophoresed on 1% agarose CHEF gels with PFG midrange marker I (New England Biolabs, Beverly, MA, USA) under the following conditions: 16 h, 14°C, 120° angle, 5–15 s linear ramp and using 0.5x TBE at 6 v/cm.

Estimation of genome representation using hybridization

The library was arrayed on to six high-density filters (22.5×22.5 cm^2, Hybond N + membrarnes, in a 4×4 double spotting pattern) on selective media containing 12.5 µg/ml chloramphenicol, grown overnight at 37°C. The resultant colonies were lysed as described by Sambrook et al. (2001) and the DNA was fixed to the filters by treatment with alkali. Overgo probes, designed previously (Chen et al. 2002) from the sequences of RFLP markers were chosen from each of the 12 chromosomes of rice. An overgo pooling strategy was used to initially pull out all the hits from the hybridization of combined probes and later reconfirmed by colony blot hybridization with individual probes. Radioactive labelling, hybridizations and washings were essentially similar to those described by Chen et al. (2002).

Organellar DNA contamination

Three barley chloroplast probes *ndhA*, *rbcL*, *psbA* (obtained from J. Mullet, Texas A & M University) and four mitochondrial probes *atpA*, *cob*, *atp9*, *coxE* from rice (obtained from T. Sasaki, MAFF) were used to screen the fosmid library to estimate the percentage of organellar contamination.

Physical map refinement

A total of 40 overgo probes were designed from the sequences flanking either both ends or one end of 28 physical gaps on chromosomes 1, 2, 6, 7, 8, 9 and 10 (Table 2) that includes nine telomeric gaps. The publicly available script ''overgo maker'' (http://genome.wustl.edu) was used to design the primer pairs for the probes. Overgo probe design, pooling hybridizations and a strategy to fill the physical gaps are adapted from Yang et al. (2003).

End sequencing and in silico anchoring

Positive fosmid clones resulting from the chromosome gap-specific probe hybridizations were sequenced using BigDye terminator chemistry v3.1 (ABI) with primers T7 (5′-TAATACGACTCACTATAGGG-3′) and pCC1-RP-2 (5′-TACGCCAAGCTATTAGGTGAGA-3′) in a 96-well format followed by capillary electrophoresis on an ABI 3730xl automated DNA sequencer. Base calling and vector trimming were automatically performed using PHRED and Crossmatch software (Ewing and Green 1998; Ewing et al. 1998). High-quality sequences were used for homology searches as described by Yang et al. (2003).

Full sequencing of fosmid clones and sequence assembly

Full sequencing and sequence assembly of each fosmid clones were performed with a similar method as used in the sequence analysis of PAC/BAC clones (Sasaki et al. 2002).

Sequence analysis

Repeat masked, finished fosmid sequences were annotated using FGENESH (Solovyev et al. 1995) for gene prediction, BLAT (Kent 2002) for rice FL cDNA sequences (Kikuchi et al. 2003) and TIGR OSGI assembled ESTs (http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species = rice), BLASTX for swissprot (http://au.expasy.org/sprot/) search. The Apollo program (Lewis et al. 2002) was used for editing and selecting gene models and translated proteins from the chosen gene models were searched against GenBank nr protein database using BLASTP. For the restriction enzyme frequency analysis, contiguous and minimally overlapping sequences from finished BAC sequences were retrieved from RGP and TIGR sequence databases. Repeat content of the fosmid clones was estimated using RepeatMasker (A.F.A. Smit and P. Green; http://ftp.genome.washington.edu/RM/RepeatMasker.html) program.

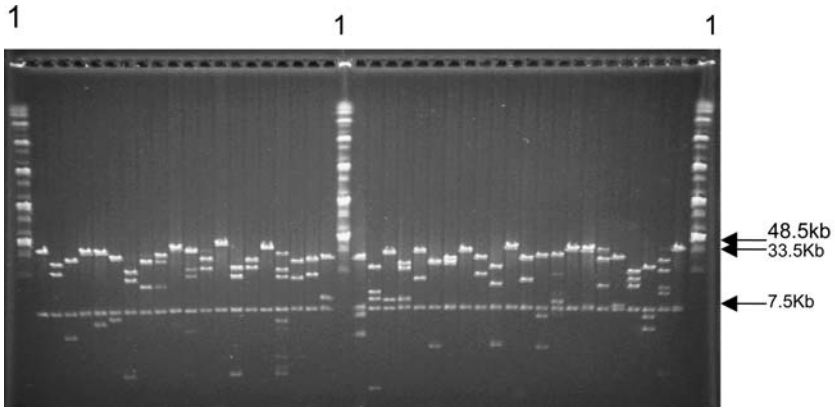

Fig. 1 *Not* I digests of randomly picked fosmid clones. *Lane marked with 1* is PFGE midrange marker I

Results

Fosmid library characterization

We successfully constructed the first sheared large insert library of *japonica* cv. Nipponbare rice. The entire library was constructed using 1.5 packaging reactions. Approximately 600–800 ng of end-repaired DNA was used for each ligation and subsequent packaging reactions. The packaging e ciencies were calculated to be approximately a little more than 10^6 cfu/μg. This ;packaging e ciency is comparable to the 10^7 cfu/μg that can be obtained using control fosmid insert (Epicentre). The library contains 110,592 clones, arrayed into 288, 384-well plates and was gridded onto six high-density filters for screening by hybridization.

The expected insert sizes should be in the range of 30–45 kbp, because the packaging extract automatically selects DNA molecules in that size range, but we did identify clones that are smaller and larger than this range. Figure 2 illustrates the insert size analysis of randomly picked clones digested with *Not*I enzyme followed by pulse field gel electrophoresis. From the analysis of 300 clones, we estimated the average insert size to be 40.8 kbp with an insert size range from 25 to 70 kbp (Fig. 2). Approximately 82% of the library contains inserts between sizes 30 and 45 kbp, 12.3% constitute inserts between 45 and 50 kbp and remaining 10% have inserts either smaller than 30 kbp or larger than 50 kbp (Fig. 2). The clones containing larger (> 50 kbp) and smaller (< 30 kbp) inserts are probably the escapees in the packaging reaction, and must have resulted from trapping of these fragments during excision and elution, since we employed only one size selection step. We have not observed any colonies with fosmids not containing inserts.

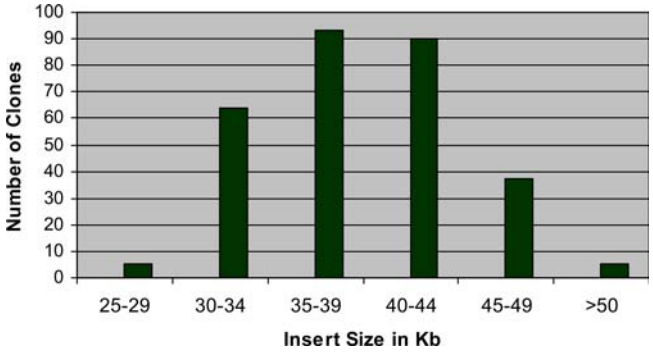

Fig. 2 Frequency distribution of fosmid clone insert sizes

To estimate the percentage of clones that contains chloroplast or mitochondrial DNA, the library was screened with probes from three barley chloroplast genes and four rice mitochondrial genes. Results showed that 1.5% of the fosmid library contained chloroplast DNA and 0.016% mitochondrial DNA sequences (data not shown). This level of organellar contamination is what we typically find using similar DNA isolation methods for plant BAC library construction (Luo et al. 2001; Budiman et al. 2000). Based on this insert size, the number of clones after subtracting the percentage of organellar contamination, the theoretical coverage of this library is approximately 10.3-fold of the rice genome.

To validate the genome coverage estimation based on average insert size and number of clones, the fosmid library was screened with a pool of 12 overgo probes derived from 12 mapped RFLP markers that belong to each of the 12 chromosomes (Kurata et al. 1994; Harushima et al. 1998). These probes have been successfully used in our laboratory to screen rice BAC libraries during the construction of a physical map of rice (Chen et al. 2002). A total of 101 hits were obtained from the combined hybridization of 12 probes indicating an average of 8.4 hits per probe. All these positive hybridization signals were reconfirmed by colony blot hybridization using individual probes separately. Table 1 shows a summary of the hybridization results. We found representation for every probe used with a range of 3–12 hits (Table 1). These results indicated more than a 99% probability of recovering any random sequence from this library.

Table 1 Summary of the hybridization results

Chromosome	Marker	No. of copies	No. of hits
1	R2277	1	10
2	S1072	2	10
3	R518	2–3	7
4	S11182	1	12
5	R3332	1	5
6	L688	1	12
7	C636	2	5
8	C1121	1	7
9	R1164	1	13
10	R1877	1	9
11	L833	1	3
12	L714	1	8

In silico physical mapping of gap-specific fosmid clones and refinement of the physical map

To test the utility of the library in filling gaps in the rice physical map, we tried to identify gap filling clones for 25 physical gaps on six rice chromosomes assigned to the RGP for complete sequencing—chromosomes 1, 2, 6, 7, 8 and 9 including nine telomeric gaps and three gaps on rice chromosome 10 assigned to ACWW (Table 2). The physical sizes of these gaps have been determined by a variety of approaches that include FISH and YAC physical mapping and from genetic maps (Wu et al. 2003; Rice chromosome 10 sequencing consortium 2003; Table 2). All gaps, except for those located on the centromeric regions, are less than 200 kbp in length, where as telomeric gaps are less than 100 kbp. Forty probes designed from the sequences flanking the 28 physical gaps either from one side or from both sides, that includes telomeric arm- specific probes (Fig. 3; Table 2) were used to screen the fosmid library using a pooling strategy. A total of 330 candidate clones were obtained, and these clones were reconfirmed by colony blot hybridizations using individual probes specific to each end of the gap. End sequences of these clones were determined and homology searches were conducted using ''BLAST'' (Altschul et al. 1997) to identify potential clones (showing more than 96% sequence identity) extending into or closing the gaps. The resulting analysis showed that 97 of these clones have extensions of di erent sizes into the gaps. Twenty-two of the 97 clones showed largest clone extensions into the gaps (Fig. 3, Table 2) and were mapped on to the physical map. Furthermore, gap 7 (Table 2) which lies between 157.1–157.6 cM on chromosome 1, has been successfully closed using two overlapping fosmid clones OSJNOa013M08 (AP006838) and OSJNOa108M02 (AP006839) (Fig. 3). Sixteen of the probe sequences flanking either both or one end of the gap, identified clones that have either no/non-significant extensions (Table 2) or repetitive sequences that cannot be used for in silico mapping. Such clones resulting from these hybridizations were excluded from further analysis.

E orts are underway to close the remaining gaps in the rice genome by chromosome-walking steps using end sequences derived from gap extending fosmid clones. We estimated the total expansion of the physical maps resulted from these gap filling or extending clones of the seven rice chromosomes to be approximately 389 kbp.

Analysis of gap-specific fosmid clone sequences

Twelve of the physical gap-specific fosmid clones were shotgun sequenced to tenfold coverage and the finished sequences were submitted to Genbank. In order to explore the di erent possibilities for the absence of these segments in the existing BAC and PAC libraries, the finished sequences were analyzed for repeat content using the software RepeatMasker version 3.0.5 (http://www.repeatmasker.org/). The total repeat contents of the fosmid clones range between 17 and 48% with LTR retro-transposons and DNA transposons as predominant class of repeats (data not shown), except for one particular fosmid clone AP006856, which consists of nearly 90% of its sequence with DNA transposons. Further, nearly 500 kbp to 1.2 Mbp of high quality contiguous sequences flanking the physical gaps were retrieved from the pseudomolecules of respective chromosomes. The frequency and distribution of recognition sites for the restriction enzymes *Hin*dIII, *Eco*RI, *Bam*HI, and *Mbo*I, used in the construction of existing BAC or PAC resources were estimated and compared with the frequency and distribution of these sites in the sequences of gap-specific fosmid clones. The analysis indicated that gap-specific fosmid clones are not devoid of any of the restriction sites for these enzymes and that both fosmid as well as gap flanking sequences have even distribution of the restriction sites for these enzymes (data not shown).

Sequencing of gap-specific fosmid clones reveal presence of euchromatic portion

Gap-specific fosmid sequences were annotated using a variety of methods (see Materials and methods). Homologs for several known important genes as well as hypothetical genes have been identified in all the gaps belonging to several chromosomes (Table 3). For example, coding sequences for drought responsive element (DRE)-binding proteins 1A and 1B known to have a role in drought stress (Liu et al. 1998), and putative arsenite stimulated ATPase (arsA-like protein hASNA-I) involved in arsenite resistance (Anjard et al. 2002) in physical gap #5 of chromosome 9, structural maintenance of chromosome (SMC2) gene protein involved in chromosome condensation and sister chromatid cohesion (Hirano 2002) in gap #6 of chromosome 1, putative homeobox-leucine zipper protein important for plant development and regulation (Schena et al. 1993) in gap #2 of chromosome 8, putative hydroxyl proline-rich

Table 2 Summary of physical gap filling using fosmid clones

Chromosome no.	Gap no.	Map position (cM)[a]	Genetic length (cM)	Predicted size (kbp)	Method	Fosmid clone identity		Fosmid clone extension(s) into the gap (Kbp)[b]
1	Tel (S arm)	0				264G09[c]		5
	1	52.7–53.9	1.2	40	FISH	N	–	–
						S	–	–
	2	60.9–63.5	2.6	160	FISH	N	173H09[c]	10
						S	ND	ND
	3	72.8–73.1	0.3	20	FISH	N	–	–
						S	ND	ND
	5	103.7–106.2	2.5	90	FISH	N	–	–
						S	–	–
	6	157.1–157.6	0.5	40	FISH	N	013M08(AP006838)	33
						S	108M02(AP006839)	36
	7	159.6–161.5	1.9	50	FISH	N	–	–
						S	–	–
	Tel (L arm)	181.8					275J08[c]	5
2	Tel (S arm)	0					183H18(AP006851)	5
	1	30.7–31.2	0.5	130	YAC map	N	–	–
						S	244C03[c]	35
	3	62.2–64.7	2.5	90	Genetic map	N	–	–
						S	257A21(AP006850)	6
	4	99.6–101.2	1.6	90	YAC map	N	283G08(AP006846)	7
						S	148N02(AP006847)	30
	Tel (L arm)	157.9					–	–
6	2	65.8–65.8	0	890	FISH	N	–	–
						S	ND	ND
	Tel (L arm)	124.4					230J22[c]	6
7	Tel (S arm)	0					219C16	2
	1	49.7–49.7	0	200	YAC map	N	ND	ND
	Tel (Larm)	118.6					136M23[c]	7
8	Tel (S arm)	0					–	–
	2					N	214P15[c]	10
						S	174H12(AP006845)	16
	Tel (L arm)	121.2					199K18(AP006848)	20
9	2	44.1–45.2	1.1	50	YAC map	N	054B01(AP007202)	
						S	ND	ND
	3	78.0–78.5	0.5	50	Genetic map	N	ND	ND
						S	018M17(AP007149)	15
	4	78.5–78.8	0.3	50	Genetic map	N	–	–
						S	178M01(AP006855)	14
	5					N	273B05(AP006859)	30
						S		
10	1	9.5–10.9	1.4	30	FISH	N	046E18	15
						S	–	–
	2	15.7	0	69	FISH	N	236K04	32
						S	137L01	11
	5	23.1	0	104	FISH	N	–	–
						S	040M24	22

'–'clones identified not to have extension or unidentified due to repetitive sequences from the fosmid clones
ND not determined
[a]Genetic distance of the flanking markers (Wu et al 2003; The Rice Chromosome 10 Sequencing Consortium 2003)
[b]After removing the overlapping regions, N and S indicate the north and south ends of the gaps
[c]Clones under sequencing

protein involved in drought, wounding and defence (Maytsik et al. 2002) in gap #4 of chromosome 5, putative RNA recognition motif (RRM) important for regulation of gene expression at post-transcriptional level (Lorkovic and Barta 2002) in the telomeric gap of long arm of chromosome 8, have been identified. Further, the role of Serine–threonine receptor-like kinases and leucine zipper proteins in defence has been well documented (Martin et al. 2003).The identification that all gap-specific fosmid sequences span gene encoding regions of rice genome is significant.

Discussion

Rice is the most important food crop in the world and has emerged as a model experimental system for plant biology and evolution primarily because of its compact 430 Mbp genome and syntenic relationships with other cereals (Gale and Devos 1998). Therefore, decoding the entire rice genome is of exceptional agricultural importance. Recent comparisons of draft versus finished rice sequences highlighted the need for a high quality

Table 3 List of putative coding regions identified in the physical gap spanning fosmid clone sequences

Fosmid accession no.	Insert size (kbp)	Extension in to the gap (kbp)	Gene	Start	End	Strand	Blastp	e-Value
AP006838	36.891	33	AP006838.1	50	3103	–	SMC2 protein	0
AP006838			AP006838.2	3936	8548	–	Unknown protein	0
AP006838			AP006838.3	16480	20014	–	Response regulator 11	9.00E-87
AP006838			AP006838.4	32138	36485	–	Putative copia-like retrotransposon polyprotein	0
AP006839	28.76	26	AP006839.1	2640	4420	+	Putative leucine zipper protein	6.00E-26
AP006839			AP006839.2	5430	7353	+	Putative leucine zipper protein	3.00E-37
AP006839			AP006839.3	9314	12526	–	Putative transcriptional factor B3 family	3.00E-07
AP006839			AP006839.4	13307	14551	–	Unknown protein	N/A
AP006839			AP006839.5	15674	16135	+	Unknown protein	N/A
AP006839			AP006839.6	23554	26372	+	Putative zinc finger protein	4.00E-78
AP006845	33.241	16	OSJNOa174H12.1	3340	4200	+	Hypothetical protein	7.00E-05
AP006845			OSJNOa174H12.4	9210	10070	–	Hypothetical protein	7.00E-05
AP006845			OSJNOa174H12.5	10983	12357	–	Putative homeobox-leucine zipper protein	6.00E-70
AP006845			OSJNOa174H12.6	14102	17072	+	Hypothetical protein	N/A
AP006845			OSJNOa174H12.7	18281	24068	+	Transcription factor like	1.00E-19
AP006845			OSJNOa174H12.8	27084	27446	–	Hypothetical protein	N/A
AP006845			OSJNOa174H12.9	28218	29585	–	Hypothetical protein	N/A
AP006845			OSJNOa174H12.12	31465	32136	–	Unknown protein	8.00E-14
AP006846	38.221	7	OSJNOa283G08.1	841	2317	–	Putative hydroxyproline-rich glycoprotein	1.00E-66
AP006846			OSJNOa283G08.2	3323	5246	+	Glycolipid transfer protein like	5.00E-86
AP006846			OSJNOa283G08.7	23259	23712	–	Hypothetical protein	N/A
AP006846			OSJNOa283G08.9	30078	34623	+	Hypothetical protein	N/A
AP006847	36.361	30	OSJNOa148N02.2	2863	3390	+	Unknown protein	N/A
AP006847			OSJNOa148N02.6	14188	15327	–	Hypothetical protein	1.00E-56
AP006848	39.45	20	OSJNOa199K18.1	4858	6399	–	Putative RNA recognition motif (RRM) containing protein	4.00E-66
AP006848			OSJNOa199K18.2	8458	10594	+	HGWP repeat containing protein like	e-116
AP006848			OSJNOa199K18.3	11603	12674	–	Hypothetical protein	N/A
AP006848			OSJNOa199K18.4	18618	21948	–	Methyl-CpG binding protein like	9.00E-41
AP006848			OSJNOa199K18.5	22762	26347	+	Hypothetical protein	1.00E-83
AP006848			OSJNOa199K18.8	33720	34636	+	Hypothetical protein	N/A
AP006848			OSJNOa199K18.9	36132	37898	–	Hypothetical protein	N/A
AP007149	36.7	15	OSJNOa018M17.1	521	3618	–	Cell wall protein like	N/A
AP007149			OSJNOa018M17.2	5232	7725	+	Calphotin-like protein	8.00E-16
AP007149			OSJNOa018M17.3	10324	14498	+	Hypothetical protein	N/A
AP007149			OSJNOa018M17.5	19938	23074	–	Putative Malonyl-CoA decarboxylase, mitochondrial precursor	e-151
AP007149			OSJNOa018M17.6	25355	33915	–	Putative *endo*-β-1,4-glucanase precursor	0
AP007149			OSJNOa018M17.7	34623	34847	+	Hypothetical protein	N/A
AP006850	34.833	6	AP006850.1	734	6647	–	Hypothetical protein	0
AP006851	7.69	5	AP006851.1	366	1092	+	Hypothetical protein	3.00E-10
AP006851			AP006851.2	3864	7430	–	Putative protein disulphide isomerase	4.00E-95
AP006855	34.955	14	AP006855.1	14693	16944	+	Unknown protein	7.00E-49
AP006855			AP006855.2	33098	33981	–	DNA polymerase delta subunit 4 family	5.00E-15
AP006856	131.21	10	AP006856.1	9	3375	–	Putative TNP-like transposable element	0
AP006856			AP006856.1	4568	7897	–	Putative TNP-like transposable element	0
AP006859	41.326	30	AP006859.1	1978	5574	+	Putative hydroxymethylglutaryl coenzyme A synthase	0
AP006859			AP006859.2	6203	9637	–	Putative arsA-like protein hASNA-I	e-137

Table 3 (Contd.)

Fosmid accession no.	Insert size (kbp)	Extension in to the gap (kbp)	Gene	Start	End	Strand	Blastp	e-Value
AP006859			AP006859.3	11302	18038	+	Putative non-LTR retroelement reverse transcriptase	e-112
AP006859			AP006859.4	24352	28339	+	DNA repair protein, putative	e-127
AP006859			AP006859.5	29937	30593	–	Putative DRE-binding protein 1B	1.00E-67
AP006859			AP006859.6	37970	38686	–	DRE-binding protein 1A	2.00E-89
AP007202	36.624	17	AP007202.1	4710	13797	+	Serine/threonine-specific receptor protein kinase like	0
AP007202			AP007202.2	20959	24418	+	Hypothetical protein	6.00E-66

finished sequence of rice (Sasaki et al. 2002; The Rice Chromosome 10 Sequencing Consortium 2003). The IRGSP employs a map-based genome sequencing strategy, for which complete and contiguous physical maps of the 12 rice chromosomes composed of large insert clones are an essential prerequisite. The presence of several physical gaps in the current physical map of rice impedes the progress towards meeting this goal of a complete sequenced cereal genome. Intensive e orts to fill the remaining gaps in the physical map using a comprehensive array of publicly available genomic tools that includes five large insert BAC/PAC libraries, constructed using partial digestions of HMW DNA with restriction enzymes, two 10 kbp plasmid libraries and both Monsanto and Syngenta draft sequences (Go et al. 2002) has proven to be quite di cult. We have recently implemented a new strategy which incorporates chromosome-walking steps and sequence reads from two 10 kbp insert size plasmid libraries (Yang et al. 2003). Although this strategy has been very e ective in closing small physical and sequencing gaps, gaps of medium to large size have proven too arduous to bridge, and involve several steps of chromosome-walking. Therefore, development of resources that can provide clone coverage across larger gaps in the BAC contig maps for the rice sequencing e ort has become extremely urgent and important.

Generation of a complete physical map of any organism is significantly a ected by the depth (genome representation) and insert size of the source library. Our own experience in the generation of BAC libraries for an array of organisms and physical mapping of rice, maize and tomato genomes (Chen et al 2002; http://www.arizona.edu/FPC/maize) along with several other physical mapping projects indicated that the true genome representation of a library is often less than the theoretical representation (estimated using Clarke and Carbon's formula = number of clones × average insert size/genome size). It has also been shown that, after a certain point, the genome representation of a library cannot be improved significantly by merely increasing the number of clones. There are several fundamental approaches to generate DNA fragments for library construction. One class involves enzymatic digestion, either partial digestion with a restriction endonuclease (Sambrook et al. 2001) or controlled degradation with DNase I (Anderson 1981). Another class involves physical stresses induced by sonication (Deininger 1983), atomization (Cavalieri and Rosenberg 1995), nebulization (Bodenteich et al. 1994) or point-sink shearing (Oefner et al. 1996).

The first approach, partial digestion of genomic DNA, is the most popular method which is easy to control and the fragments are relatively easy to clone. All present day large insert BAC, PAC, and YAC libraries are developed using this approach. However, the cloning e ciency is not uniform because of the nonrandom distribution of enzyme recognition sites across a given genome. Regions of genomes devoid or dense for

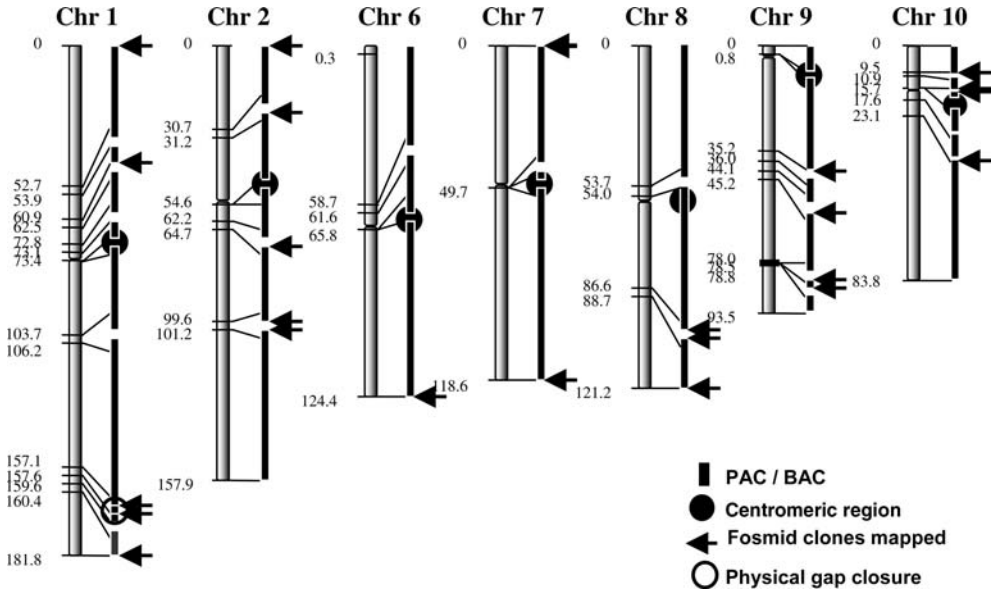

Fig. 3 The physical map refinement using fosmid clones. *Arrows* indicate distrubution of fosmid clones that extend into the physical gaps on respective chromosomes

particular enzyme result in fragments smaller than 40 kbp or larger than 400 kbp that are excluded during the fragment size-selection process of library construction, leading to under- or overrepresentation of some regions, a manifestation of which are gaps in the physical map. Restriction sites located in repeated elements pose another problem, since the end sequence data cannot be used for developing a minimum tile of BAC clones or merging the ends of FPC BAC contigs. A practical approach to solving the problem of underrepresentation of DNA fragments in a library is to simply employ two or three large insert libraries derived from multiple enzymes from the same genotype, that collectively represents as much as 26× genome coverage (Yim et al. 2002). Although the use of multiple libraries in physical mapping projects has been found to be essential, not all gaps in physical maps can be filled this way.

Random shearing generates fragments that are truly random and thereby reduces cloning bias (Thorstenson et al. 1998). However, methods for e ciently cloning sheared large insert BAC libraries are still in infancy. In this report, we have successfully constructed the first random sheared fosmid library for rice, and possibly all plants, using the *japonica* cv. Nipponbare. fosmid libraries, however, have earlier been used extensively in whole genome physical mapping and sequencing of a variety of organisms including bacteria, humans and animals. The stability of fosmid clones has been extensively studied and found to be comparable with BAC clones and are more user-friendly to work in comparison to BAC or YAC clones.

In this study, we demonstrated that closing the gaps can be achieved by developing a medium size insert library, with low or no cloning bias, in combination with library screening and in silico mapping. The fosmid library reported in this paper has total genome coverage of more than tenfold based on insert size and clone number. These results were confirmed with our hybridization data of 12 RFLP markers mapped to di erent chromosomes of rice. Each probe detected an average of 8.4 hits, indicating that the probability of finding any sequence in this library is more than 99.99%. Analysis of the organellar contamination in the library using hybridization with chloroplast and mitochondrial probes demonstrated that the fosmid library exhibits a low percentage of contamination with these sequences. The suitability of any large insert library for positional cloning, physical mapping and whole genome sequencing depends on the ability to recover clones from specific regions by screening. We have identified and mapped 22 fosmid clones to the physical map that are either gap extending or gap filling on di erent chromosomes. The total contig extension is estimated to be approximately 389 kbp. These results of physical gap filling and map refinement indicate that the fosmid library is less biased as compared to enzyme-digested libraries. In this respect, the current fosmid library has a more complete representation of the genome in comparison to the previously available BAC and PAC libraries.

In contrast to the high technical skills and cost required to develop a large-insert deep-coverage BAC library, fosmid library construction is quite trivial (De Tomaso and Weissman 2003). Moreover, the cloning vector we used (Epicentre) contains both the single-copy *E. coli* F-factor replicon and the inducible high-copy oriV origin of replication (Wild et al. 2002). Initiation of replication from oriV requires the trfA gene product that is on the *E. coli* copy control host cells (TransforMax EPI300 cells, Epicentre). This system provides the user with all advantages of single-copy cloning vector and a high-copy cloning vector without the disadvantages of either. Copy control clones can be induced to high copy number for high yields of DNA for all ''downstream'' applications. Although, we have only demonstrated the utility of the fosmid library in refining the physical map of rice, its applications are unlimited, for example,

positional cloning and genome organization studies. The primary disadvantage of fosmid libraries is inert size limitation to about 40 kbp. This is in contrast to BAC with an insert capacity up to 300–400 kbp.

The presence and even distribution of recognition sequences for enzymes that have been previously used in the development of BAC or PAC library resources of Nipponbare, in the gap-specific fosmid sequences suggests that the reasons for their absence in the existing BAC or PAC library resources is not entirely due to cloning bias due to restriction enzyme usage and may also relate to their secondary structural properties. Sequences with unsual DNA structures such as Z-DNA, long inverted terminal repeats, AT-rich sequences have been reported to be unstable in *E. coli* genomes (Hagan and Warren 1982; Schroth and Ho 1995; Kang and Cox 1996; Ravin and Ravin 1999; Razon et al. 2001). Also except for one fosmid clone, all the fosmid clones do not contain exceptionally high-repeat content that could have caused their nonclonability. A significant result of this study is the mapping of gaps to regions corresponding to genes. The fact that the gaps in the rice genome may correspond to chromosomal regions encoding functional genes emphasizes the importance of the final steps of genome sequencing. There are still several gaps with unknown sequences in the rice genome. We summarize here a new strategy that is being adapted by IRGSP to refine the physical map of rice in order to characterize the tough leftovers of the rice genome sequence. This involves screening the fosmid library with overgo probes flanking a physical gap, for which there is no representation in the enzyme-digested BAC/PAC libraries, identification of a single or several overlapping fosmid clones to fill small/medium size gaps that range in size 20–100 kbp. For large physical gaps, gap extending fosmid clones will further be used to screen large insert libraries like BAC/PAC to fill the gaps, and these clones will be further used to either fill gaps or rescreen the fosmid library to fill the gaps.

References

Altschul SF, Madden TL, Scha er AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Anderson S (1981) Shotgun DNA sequencing using DNase I-generated fragments. Nucleic Acids Res 9:3015–3027

Anjard C, Loomis WF, Dictyostelium Sequencing Consortium (2002) Evolutionary analyses of ABC transporters of *Dictyostelium discoideum*. Eukaryot Cell 1:643–52

Baba T, Katagiri S, Tanoue H, Tanaka R, Chiden Y, Saji S, Hamada M, Nakashima M, Okamoto M, Hayashi M, Yoshiki S, Karasawa W, Honda M, Ichikawa Y, Arita K, Ikeno M, Ohta T, Umehara Y, Matsumoto T, de Jong P, Sasaki T (2000) Construction and Characterization of Rice Genomic Libraries: PAC Library of *Japonica* Variety, Nipponbare and BAC Library of *Indica* Variety, Kasalath. Bull NIAR 14:41–49

Bodenteich AS, Chissoe Y, Wang F, Roe BA (1994) Shotgun cloning as the strategy of choice to generate templates for high throughput dideoxynucleotide sequencing. In: Adams MD, Fields C, Venter C (eds) Automated DNA sequencing and analysis techniques. Academic, London, pp 42–50

Budiman MA, Mao L, Wood TC, Wing RA (2000) A deep coverage tomato BAC library and prospects toward developing of an STC framework for genome sequencing. Genome Res 10:129–136

Cavalieri LF, Rosenberg BH (1995) Shear degradation of deoxyribonucleic acid. J Am Chem Soc 81:5136–5139

Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang G, Kim H, Frisch D, Yu Y, Sun S, Higingbottom S, Phimphilai J, Phimphilai D, Thurmond S, Gaudette B, Li P, Liu J, Hatfield J, Main D, Farrar K, Henderson C, Barnett L, Costa R, Williams B, Walser S, Atkins M, Hall C, Budiman MA, Tomkins JP, Luo M, Bancroft I, Salse J, Regad F, Mohapatra T, Singh NK, Tyagi AK, Soderlund C, Dean RA, Wing RA (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14:537–545

De Tomaso AW, Weissman IL (2003) Construction and characterization of large-insert genomic libraries (BAC and Fosmid) from the ascidian *Botryllus schlosseri* and initial physical mapping of a histocompatibility locus. Mar Biotechnol 5:103–115

Deininger PL (1983) Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. Anal Biochem 129:216–223

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8:186–194

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8:175–185

Fitz-Gibbon S, Choi AJ, Miller JH, Stetter KO, Simon MI, Swanson R, Kim UJ (1997) A fosmid-based genomic map and identification of 474 genes of the hyperthermophilic archaeon *Pyrobaculum aerophilum*. Extremophiles 1:36–51

Gale MD, Devos KM (1998) Plant comparative genetics after 10 years. Science 282:656–659

Gingrich JC, Boehrer DM, Garnes JA, Johnson W, Wong BS, Bergmann A, Eveleth GG, Langlosis RG, Carrano AV (1996) Construction and characterization of human chromosome 2 specific cosmid, Fosmid, and PAC clone libraries. Genomics 32:65–74

Go SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, and Briggs S (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296: 92–100

Hagan CE, Warren GJ (1982) Lethality of palindromic DNA and its use in selection of recombinanat plasmids. Gene 19:147–151

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, and Sasaki T (1998) A high density rice genetic linkage map with 2275 markers using a single F2 population. Genetics 148:479–494

Hirano T (2002) The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesion, and repair. Genes Dev 15:399–414

Hoyer LL, Payne TL, Bell M, Myers AM, Scherer S (1998a) *Candida albicans* ALS3 and insights into the nature of the ALS gene family. Curr Genet 33(6):451–459

Hoyer LL, Payne TL, Hecht JE (1998b) Identification of *Candida albicans* ALS2 and ALS4 and localization of als proteins to the fungal cell surface. J Bacteriol 189:5334–5343

Hughes DS, Felbeck H, Stein JL (1997) A histidine protein kinase homolog from the endosymbiont of the hydrothermal vent tubeworm *Riftia pachyptila*. Appl Envion Microbiol 63:3494–3498

International Rice Genome Sequencing Consortium (2005) The map-based sequence of the rice genome.Nature 436:793–800

Kang HK, Cox DW (1996) Tandem repeats 3′ of the IGHA genes in the human immunoglobulin heavy chain gene cluster. Genomics 35:189–195

Kawarabayasi Y, Sawada M, Horikawa H, Haikawa Y, Hino Y, Yamamoto S, Sekine M, Baba S, Kosugi H, Hosoyama A, Nagai Y, Saka M, Ogura K, Otsuka R, Nakazawa H, Takamiya M, Ohfuku Y, Funahashi T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Aoki K, Kikuchi H (1998) Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. DNA Res 30:55–76

Kent WJ (2002) BLAT—the BLAST-like alignment tool. Genome Res 12:656–664

Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, Hotta I, Kojima K, Namiki T, Ohneda E, Yahagi W, Suzuki K, Li CJ, Ohtsuki K, Shishiki T, Otomo Y, Murakami K, Iida Y, Sugano S, Fujimura T, Suzuki Y, Tsunoda Y, Kurosaki T, Kodama T, Masuda H, Kobayashi M, Xie Q, Lu M, Narikawa R, Sugiyama A, Mizuno K, Yokomizo S, Niikura J, Ikeda R, Ishibiki J, Kawamata M, Yoshimura A, Miura J, Kusumegi T, Oka M, Ryu R, Ueda M, Matsubara K, Kawai J, Carninci P, Adachi J, Aizawa K, Arakawa T, Fukuda S, Hara A, Hashizume W, Hayatsu N, Imotani K, Ishii Y, Itoh M, Kagawa I, Kondo S, Konno H, Miyazaki A, Osato N, Ota Y, Saito R, Sasaki D, Sato K, Shibata K, Shinagawa A, Shiraki T, Yoshino M, Hayashizaki Y, Yasunishi A (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 301:376–379

Kim UJ, Shizuya H, de jong P, Birren B, Simon MI (1992) Stable propagation of cosmid sized human DNA inserts in an F factor based vector. Nucleic Acids Res 20:1083–1085

Kim UJ, Shizuya H, Sainz J, Garnes J, Pulst SM, de Jong P, Simon MI (1995) Construction and utility of a human chromosome 22 specific fosmid library. Genet Anal 12:81–84

Kim CG, Fujiyama A, Saitou N (2003) Construction of a gorilla fosmid library and its PCR screening system. Genomics 82:571–574

Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J,Antonio BA, Shomura A, Shimizu T, Lin S-Y, Inoue T, Fukuda A, Shimano T, Kuboki Y, Toyama T, 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 (1994) A 300 kilobase interval genetic map of rice inclusing 883 expressed sequences. Nat Genet 8:365–372

Leem SH, Kouprina N, Grimwood J, Kim JH, Mullokandov M, Yoon YH, Chae JY, Morgan J, Lucas S, Richardson P, Detter C, Glavina T, Rubin E, Barrett JC, Larionov V (2004) Closing the gaps on human chromosome 19 revealed genes with a high density of repetitive tandemly arrayed elements. Genome Res 14(2):239–246

Lewis SE, Searle SM, Harris N, Gibson M, Lyer V, Richter J, Wiel C, Bayraktaroglir L, Birney E, Crosby MA, Kaminker JS, Matthews BB, Prochnik SE, Smithy CD, Tupy JL, Rubin GM, Misra S, Mungall CJ, Clamp ME (2002) Apollo: a sequence annotation editor. Genome Biol 3:RESEARCH0082–2

Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1998) Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in Arabidopsis. Plant Cell 10:1391–1406

Lorkovic ZJ, Barta A (2002) Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant *Arabidopsis thaliana*. Nucleic Acids Res 30:623–635

Luo M, Wing RA (2003) An improved method for plant BAC library construction. In: Grotewold E (ed) Plant functional genomics: methods and protocols. methods in molecular biology. Humana Press Inc. Totowa, pp 3–19

Luo M, Wang Y-H, Frish D, Joobeaur T, Wing RA, Dean RA (2001) Melon bacterial artificial chromosome(BAC) library construction using improved methods and identification of clones linked to the locus conferring resistance to melon *Fusarium* wilt (*Fom-2*). Genome 44:154–162

Mao L, Wood TC, Yu Y, Budiman MA, Tomkins J, Woo S, Sasinowski M, Presting G, Frish D, Go SA, Dean RA, Wing RA (2000) Rice transposable elements:a survery of 73,000 seqeunce-tagged-connectors. Genome Res 10:982–990

Martin GB, Bogdanove AJ, Sessa G (2003) Understanding the functions of plant disease resistance proteins. Annu Rev Plant Biol 54:23–61

Matysik J, Bhalu AB, Mohanty P (2002) Molecular mechanisms of quenching of reactive oxygen species by proline under stress in plants. Curr Sci 82:525–532

Millikan DS, Felbeck H, Stein JL (1999) Identification and characterization of flagellin gene from the endosymbiont of the hydrothermal vent tubeworm *Riftia pachyptila*. Appl Environ Microbiol 65:3129–33

Oefner PJ, Hunicke-Smith SP, Chiang L, Dietrich F, Mulligan J, Davis RW (1996) E cient random subcloning of DNA sheared in a recirculating point-sink flow system. Nucleic Acids Res 24: 3879–3886

Quiser A, Ochsenreiter T, Klenk HP, Kletzin A, Treusch AH, Meurer G, Eck J, Sensen CW, Schelper C (2002) First insights in to the genome of an uncultivated crenarchaeote from soil. Environ Microbiol 4:603–611

Ravin NV, Ravin VK (1999) Use of a llinear multicopy vector based on the mini-replicon of temperate coliphage N15 for cloning DNA with abnormal secondary structures. Nucleic Acids Res 1:e13

Razin SV, Loudinkova ES, Trifonov E, Scherrer K (2001) Nonclonability correlates with genome instability: a case study of unique DNA region. J Mol Biol 307:481–486

Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T, Katayose Y, Wu J, Niimura Y, Cheng Z, Nagamura Y, Antonio BA, Kanamori H, Hosokawa S, Masukawa M, Arikawa K, Chiden Y, Hayashi M, Okamoto M, Ando T, Aoki H, Arita K, Hamada M, Harada C, Hijishita S, Honda M, Ichikawa Y, Idonuma A, Iijima M, Ikeda M, Ikeno M, Ito S, Ito T, Ito Y, Ito Y, Iwabuchi A, Kamiya K, Karasawa W, Katagiri S, Kikuta A, Kobayashi N, Kono I, Machita K, Maehara T, Mizuno H, Mizubayashi T, Mukai Y, Nagasaki H, Nakashima M, Nakama Y, Nakamichi Y, Nakamura M, Namiki N, Negishi M, Ohta I, Ono N, Saji S, Sakai K, Shibata M, Shimokawa T, Shomura A, Song J, Takazaki Y, Terasawa K, Tsuji K, Waki K, Yamagata H, Yamane H, Yoshiki S, Yoshihara R, Yukawa K, Zhong H, Iwama H, Endo T, Ito H, Hahn JH, Kim HI, Eun MY, Yano M, Jiang J, and Gojobori T (2002) The genome sequence and structure of rice chromosome 1. Nature 420:312–316

Schena M, Lloyd AM, Davis RW (1993). The HAT4 gene of Arabidopsis encodes a developmental regulator. Genes Dev 7:367–379

Scroth GP, Ho PS (1995) Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. Nucleic Acids Res 23:1977–1983

Soderlund C, Humphray S, Dunham A, French L (2000) Contigs built with fingerprints, markers and FPC V4.7. Genome Res 10:1772–1787

Solovyev VV, Salamov AA, Lawrence CB (1995) Identification of human gene structure using linear discriminant functions and dynamic programming. Proc Int Conf Intell Syst Mol Biol 3:367–375

The Rice Chromosome 10 Sequencing Consortium (2003) In-depth view of structure, activity, and evolution of rice chromosome 10. Science 300:1566–1569

Thorstenson Y, Hunicke-Smith S, Oefner P, Davis R (1998) An automated hydrodynamic process for controlled, unbiased DNA shearing. Genome Res 8:848–855

Venter JC, Smith HO, Hood L (1996) A new strategy for genome sequencing. Nature 381:364–366

Vergin KL, Urbach E, Stein JL, DeLong EF, Lanoil BD, Giovannoni SJ (1998) Screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order Planctomycetales. Appl Environ Microbiol 64:3075–3078

Wild J, Hradecna Z, Szybalski W (2002) Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. Genome Res 12:1434–1444

Wu J, Mizuro H, Hayashi- Tsugane M, Ito Y, Chiden Y, Fujisawa M, katagiri S, Saji S, Yoshiki S, Karasawa W, Yoshihara R, Hayashi A, Kobayashi H, Ito K, Hamada M, Okamoto M, Ikeno M, Ichikawa Y, Katayose Y, Yano M, Matsumoto T, Sasaki T (2003) Physical maps and recombination frequency of six rice chromosomes. Plant J 36:720–730

Yang TJ, Yu Y, Nah G, Atkins M, Lee S, Frisch DA, Wing RA (2003) Construction and utility of 10-kb libraries for e cient clone–gap closure for rice genome sequencing. Theor Appl Genet 107:652–660

Yim Y-S, Davis GL, Duru NA, Musket TA, Linton EW, Messing JW, McMullen MD, Soderlund CA, Polacco ML, Gardiner JM, Coe EH Jr (2002) Characterization of three maize bacterial artificial chromosome libraries toward anchoring of the physical map to the genetic map using high density bacterial artificial chromosome filter hybridization. Plant Physiol 130:1686–1696