CELLULAR & MOLECULAR BIOLOGY LETTERS
Volume 10, (2005) pp 425 – 437
http://www.cmbl.org.pl

Received 11 January 2005
Accepted 20 June 2005

CROSS-SPECIES BACTERIAL ARTIFICIAL CHROMOSOME (BAC) LIBRARY SCREENING VIA OVERGO-BASED HYBRIDIZATION AND BAC-CONTIG MAPPING OF A YIELD ENHANCEMENT QUANTITATIVE TRAIT LOCUS (QTL) *YLD1.1* IN THE MALAYSIAN WILD RICE *Oryza rufipogon*

BENG-KAH SONG[1], KALAIVANI NADARAJAH[2]*, MICHAEL N. ROMANOV[3] and WICKNESWARI RATNAM[1]

[1]School of Environmental and Natural Resource Sciences, Faculty of Science and Technology, Kebangsaan University Malaysia, 43600 Bangi, Selangor, Malaysia; [2]School of BioScience and Biotechnology, Faculty of Science and Technology, Kebangsaan University Malaysia, 43600 Bangi, Selangor, Malaysia; [3]Genetics Division, CRES (Conservation and Research for Endangered Species), Zoological Society of San Diego, Arnold and Mabel Beckman Center for Conservation Research, 15600 San Pasqual Valley Road, Escondido, CA 92027-7000, USA

Abstract: The construction of BAC-contig physical maps is an important step towards a partial or ultimate genome sequence analysis. Here, we describe our initial efforts to apply an overgo approach to screen a BAC library of the Malaysian wild rice species, *Oryza rufipogon*. Overgo design is based on repetitive element masking and sequence uniqueness, and uses short probes (~40 bp), making this method highly efficient and specific. Pairs of 24-bp oligos that contain an 8-bp overlap were developed from the publicly available genomic sequences of the cultivated rice, *O. sativa*, to generate 20 overgo probes for a 1-Mb region that encompasses a yield enhancement QTL *yld1.1* in *O. rufipogon*. The advantages of a high similarity in melting temperature, hybridization kinetics and specific activities of overgos further enabled a pooling strategy for library screening by filter hybridization. Two pools of ten overgos each were hybridized to high-density filters representing the *O. rufipogon* genomic BAC library. These screening tests succeeded in providing 69 PCR-verified positive

* Corresponding author; tel.: +60-3-89213465, fax: +60-3-89252698, e-mail: vani@pkrisc.cc.ukm.my

Abbreviations used: BAC – bacterial artificial chromosome; QTL – quantitative trait locus; PAC – P1-derived artificial chromosome; STS – sequenced-tagged site; UTR – untranslated regions

hits from a total of 23,040 BAC clones of the entire *O. rufipogon* library. A minimal tilling path of clones was generated to contribute to a fully covered BAC-contig map of the targeted 1-Mb region. The developed protocol for overgo design based on *O. sativa* sequences as a comparative genomic framework, and the pooled overgo hybridization screening technique are suitable means for high-resolution physical mapping and the identification of BAC candidates for sequencing.

Key Words: Overgo, BAC Library Screening, *Oryza rufipogon*, *Oryza sativa*, Pooled Probes Hybridization, Yield QTL *yld1.1*, BAC-contig Physical Mapping

INTRODUCTION

The development of large insert fragment (>100 kb) cloning technologies, such as BAC [1] and P1-derived artificial chromosome (PAC) [2], has provided a powerful tool for the construction of representative libraries of the entire genomes of any organism. To date, many BAC libraries have been constructed and utilized for various genome analyses. Major applications of BAC libraries include restriction fingerprinting for the construction of contigs towards genome sequencing [3, 4] and map-based cloning [5]. In order to facilitate the contig construction of a physical map, different strategies have been developed and used in the screening of BAC libraries. Two major approaches, PCR and filter hybridization-based screenings, have been used to generate contigs of large genomes. The technique of screening by filter hybridization is based on the use of radioactively labelled probes. These radiolabelled probes often originate from mapped DNA markers as oligonucleotides or PCR amplification products that are used in screening high-density BAC colony filters [6, 7]. However, there may be a higher probability that such probes contain repeat DNA regions and will produce multiple hits on the filters.

Novel hybridization probes termed overgos (overlapping oligonucleotides) [8] have proven to be more efficient than the aforementioned conventional probes. Overgo design is based on repetitive element masking and sequence uniqueness, and uses short probes, making this method highly efficient and specific. Overgos are made by annealing two overlapping primers (22-26 bases long) and extending the overhanging primer ends with Klenow enzyme in the presence of normal and radiolabelled nucleotides. The length of complementary overlap is usually 8 bp and the resulting overgo probe is 36 to 44 bp long. As a result, the overgo method is significantly less affected by sequence repeats, and is much more sensitive and easier to handle, with a higher throughput, lower rates of false positives and a lower background effect compared to the other probe hybridization methods [6]. This method promises to provide a powerful means for economical and rapid genome physical mapping [9-12].

Here, we report our efforts with the application of the overgo hybridization technique for screening of the *O. rufipogon* BAC library using the comparative genomic tools available for the related cultivated rice species *O. sativa*. The

yield enhancement QTL *yld1.1* region from chromosome 1 [13, 14] was selected for overgo design and library screening in this project. This study was conducted to generate a sequenced-tagged site-based (STS-based) BAC-contig physical map of this region, which we hope will facilitate map-based cloning of yield improving genes within this region. The long-term goal of the study is to facilitate the improvement of grain yield and quality in Malaysian rice varieties within the framework of the Malaysian Rice Project (http:// pkukmweb.ukm.my/~rice/).

MATERIALS AND METHODS

Database searching and data mining

Database information on genetic/physical maps, integrated maps, mapped molecular markers, yield related contigs, sequence-ready BACs/PACs and annotation for genome sequences in GenBank records were accessed mainly from two database resources: Gramene (http://www.gramene.org/) and the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/). Bioinformatics applications, such as PCR primer design, sequence alignment, repetitive DNA sequence masking, and overgo design softwares were also employed; we used publicly available web-based tools (as listed at http://pkukmweb.ukm.my/~rice/biodb.html) to facilitate the process of overgo probe development.

Design of overgo probes

Overlapping oligonucleotide probes were derived from the available *O. sativa* genome sequence for a region around the microsatellite marker RM5 (GenBank accession number AF344007), which is linked to QTL *yld1.1* [14] as shown in Fig. 1.

Sequences of nine BACs/PACs [contig 15, Gramene Rice-I-Map (FPC)] were selected to design overgo probes using the online software OvergoMaker 40 (http://www.genome.wustl.edu/tools/index.php?overgo=1). Based on the annotation in the GenBank files on the BAC/PAC clones for *O. sativa* ssp. *japonica* cv. Nipponbare, untranslated regions (UTR) for coding sequences of putative proteins were selected for probe design due to their uniqueness in the plant genome. The interval distance between probes was set at 40-60 kb to gain full BAC coverage of the region of interest. Selection criteria for overgos included: 8-bp complementary overlap of two overgo primers, a ~50% GC content of probe sequences, a 40-bp total overgo length, and a melting temperature (T_m) ranging between 67 and 70°C after Klenow fill-in, as checked with the OligoAnalyzer 3.0 program (http://scitools.idtdna.com/scitools/ Applications/OligoAnalyzer/). Also, at least ten Gs and Ts on both overhang strands are necessary to incorporate a sufficient number of radioactive Cs and As. The primer sequence was then checked against repetitive DNA sequences using the RepeatMasker (http://www.repeatmasker.org/) and BLASTN (http:// www.ncbi.nlm.nih.gov/ BLAST/) programs.

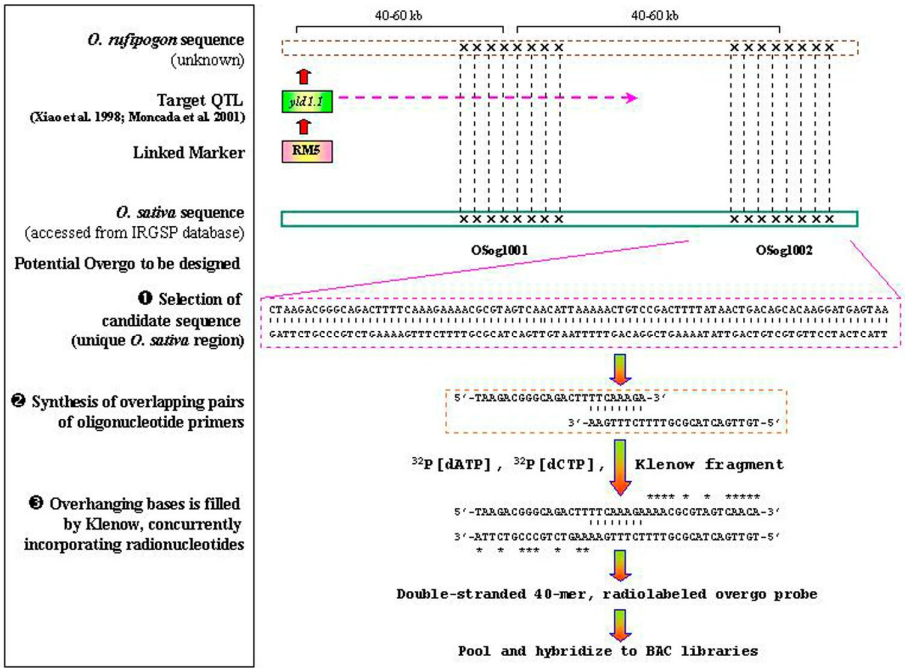

Fig. 1. The strategy for designing overgo hybridization probes based on the *O. sativa* sequence.

BAC library screening

Two-and-a-half high-density filters that contain 9,216 BAC clones per filter that represent the entire genomic library of the Malaysian wild rice species *O. rufipogon* (accession number IRGC 105491) were prepared by Amplicon Express (Pullman, WA, USA; http://www.genomex.com/downloads/AEX_List_BAC_Library.xls). Two sets of ten pooled overgo probes each were labelled with ^{32}P, hybridized to high-density filters, and subsequently washed using the overgo hybridization protocol [8]. The filters were then exposed to X-ray films for 3-7 days, depending on the strength of their hybridization signals. The overgo hybridization procedure used in this study was slightly modified from the original protocol [8], where the hybridization was done at 58°C instead of the recommended 60°C. In addition, the hybridization period was also increased to two days to allow stronger signals; this was especially useful when older filters were re-used. The change in hybridization temperature and period yielded a clearer and reduced background of hybridization after exposure of the filters for one week at -70°C.

Overgo probe performance validation by PCR

The positive BAC clones were subjected to a PCR-STS verification test using STS primer sets derived from *O. sativa* ssp. *japonica* sequences. One primer was

picked from the primer set for the appropriate overgo probe; another primer was newly designed using *O. sativa* ssp. *japonica* sequences as a framework. Primer pairs that failed to give a single amplification product were re-designed within the 5- to 9-kb regions adjacent to the overgo sites. Primer sequence information can be obtained from the authors upon request. Optimized PCR amplification was performed in a 20-µl reaction mix containing 20 mM of Tris-HCl (pH 8.0), 50 mM of KCl, 1.5 mM of $MgCl_2$, 0.5 µM of each primer, 0.2 mM dNTP mix, 0.7 U of *Taq* DNA polymerase (Invitrogen) and 1 µl of diluted (1:100) BAC overnight culture. All amplifications were carried out for 40 cycles of 1 min at 94°C (denaturation), 1 min at 64°C (annealing), and 2 min at 72°C (elongation), with a final 10 min extension at 72°C. The template genomic DNA of *O. rufipogon* and *O. sativa* was used as a positive control in the PCR reaction for the confirmation of BAC identity. The PCR amplification products were loaded on 1.5% agarose gels and electrophoresed at 100V.

RESULTS

Data mining and overgo probe design

We aligned *in silico* the *yld1.1*-linked marker RM5 sequence (AF344007) with available *O. sativa* ssp. *japonica* contigs and BACs/PACs located on rice chromosome 1. Nine PAC/BAC clones were identified on the basis of their proximity to the microsatellite RM5 marker (AP003216, AP002844, AP003314, AP002866, AP002897, AP003328, AP002843, AP002819, and AP002744; Tab. 1). A region of the approximately 1 Mb in length estimated on the *O. sativa* sequence was selected for overgo development. Based on annotation in the GenBank files on the *O. sativa* BAC/PAC clones, we designed 20 overgo probes from the UTR sequences of putative genes within the *O. rufipogon yld1.1* region. The probes were designed and spaced at an average distance of 40-60 kb using *O. sativa* sequences as a model for the spacing and the order of overgo probes in the *O. rufipogon* orthologous region. 5' UTR or 3' UTR regions of coding sequences were selected from their affiliated *O. sativa* BAC/PAC clones (Tab.1). All the designed probe sequences were checked using BLAST search against the NCBI dataset for *O. sativa* (http://www.ncbi.nlm.nih.gov/BLAST/ Genome/ PlantBlast.shtml?4) to confirm their uniqueness in the *O. sativa* genome. In the case of the putative CENPC-like protein, both the 5' and 3' UTR regions were found to contain repetitive elements in the rice genome. Using the first intron of the putative protein allowed us to compose a unique probe (OSog1012) for this gene.

Overgo hybridization of high-density BAC filters

In order to evaluate the efficiency of the overgo probe design strategy and the performance of the newly designed overgo probes in this study, we applied the first pool of ten overgos (OSog1001-1010) onto the high-density filter of our BAC library. The screening test generated 38 positive clones, and thus proved

Tab. 1. Description of the 20 overgo probes used for the cross-species hybridization.

Probe designation	GenBank accession no. of affiliated BAC/PAC	Predicted gene/CDS[a]	Input sequence region for overgo design	GC content (%)/ T_m (°C)	Left (L) and right (R) primer sequence
Osog1001	AP003216 (PAC)	Receptor-like protein kinase	3' UTR	52.5/ 71.6	L : GAGTGTGAAGCTAGCAAACTGCAG R : AGTGTGGTCGTATCCGCTGCAGTT
Osog1002	AP003216 (PAC)	Cyst nematode resistance protein	3' UTR	42.5/ 68.3	L : TAAGACGGGCAGACTTTTCAAAGA R : TGTTGACTACGCGTTTTCTTTGAA
Osog1003	AP002844 (PAC)	ABC transporter protein	3' UTR	52.5/ 71.2	L : GAATGGAATCCGTTCCGTTGGAAC R : CGTTCCATTCCGTTCCGTTCCAAC
OSog1004	AP002844 (PAC)	H^+-transporting ATPase	3' UTR	45.0/ 67.2	L : ATTTCGGGGAGACCTATCAGAAGA R : CAACTGGACCAGATATTCTTCTGA
OSog1005	AP003314 (PAC)	GTP-binding protein	5' UTR	50.0/ 72.0	L : AAGCCCTTTGACGTTGGAACAGTC R : AGCTTCTCGACGTTTGGACTGTTC
OSog1006	AP003314 (PAC)	DNA methyltransferase	5' UTR	50.0/ 70.5	L : CACGAATCATACCCACGAACCAAC R : ATCACCCTAAAGCCAGGTTGGTTC
OSog1007	AP002866 (PAC)	Unknown protein[b]	5' UTR	47.5/ 68.7	L : CAGAGTCACATTCTCCTACATCAC R : AAGCTATCTGCAGCCTGTGATGTA
OSog1008	AP002866 (PAC)	Unknown protein[b]	5' UTR	45.0/ 68.9	L : TTTGCATCAGTCTAGTGTTCACTG R : TCAACTGAGCACTGGACAGTGAAC
OSog1009	AP002866 (PAC)	Acetyl transferase	5' UTR	47.5/ 68.4	L : GATGAGATCCTTACACACATCGTC R : CGTGCCATATTGCCATGACGATGT
OSog1010	AP002897 (PAC)	Hypothetical protein[c]	5' UTR	55.0/ 71.0	L : CATCACCCCGCATATATGTCGATC R : GTATAGCGACCGGACCGATCGACA
OSog1011	AP002897 (PAC)	CTP synthase	5' UTR	42.5/ 68.0	L : AGCAAAAAGGGAATGTGTAAATCC R : TCAGTAGGTGGCCTAAGGATTTAC
OSog1012	AP002897 (PAC)	CENPC-like protein	Intron #1	45.0/ 68.5	L : TGACGAACTACTGTTCTGTTTGTG R : GCTCGAGCAAAATCCACACAAACA
OSog1013	AP003328 (BAC)	Cell division protein FtsH protease-like	5' UTR	47.5/ 69.8	L : TTACAGCTGGGTCTACTGCCTTGA R : GGACCTTTTTGTTGCATCAAGGCA
OSog1014	AP003328 (BAC)	Hypothetical protein	5' UTR	42.5/ 67.0	L : GTGGCTTGCGAAAGAAGGAAGAAG R : GATTGAAACATGTGTTCTTCTTCC
OSog1015	AP003328 (BAC)	UTP-glucose glucosyltransferase	5' UTR	52.5/ 72.0	L : AAAATGTTACTGCCGCTGGCTTAG R : CAGAGGAACGCGCTACCTAAGCCA
OSog1016	AP002843 (PAC)	Hypothetical protein	5' UTR	50.0/ 69.8	L : CGGAGCTTTGGTGTTTGTACGTAC R : GAGCACAACACGTATGGTACGTAC
OSog1017	AP002819 (PAC)	Uroporphyrinogen decarboxylase	5' UTR	47.5/ 67.5	L : AGCTTGCAGACAAAACTACGATGG R : GCTGATCCATGGTGTACCATCGTA
OSog1018	AP002819 (PAC)	Hypothetical protein	5' UTR	45.0/ 67.3	L : TATGTACAGCTAACGACGAGATAG R : TCGACCCCAAGTATAGCTATCTCG
OSog1019	AP002819 (PAC)	Endopolygalacturonase	5' UTR	42.9/ 67.6	L : TCCGTTGTATCTTCGTCCATGTGTA R : ACCGAGTACTCACTATTTACACATG
OSog1020	AP002744 (PAC)	Heat shock factor protein 1	5' UTR	50.0/ 70.4	L : GGGATTTGACGTGTAGGCAAAAGC R : ACTGCTCCGAGCTATAGCTTTTGC

[a]Coding sequence predicted from the genome sequence of *O. sativa* ssp. *japonica* cv. Nipponbare, [b]Gene without significant homology to any protein but with homology to EST, [c]Gene predicted with a gene prediction program.

the feasibility of our strategy of overgo design and hybridization method. Another pool of hybridization probes comprising the remaining ten overgos (OSog1011 to 1020) was used in the second screen, and this produced 38 positive hits (Tab. 2). After two successful hybridizations, we obtained 71 positive BACs (3.6 clones per overgo) from a total of 23,040 clones, which represents the whole *O. rufipogon* BAC library. Five clones were found in both screens (10N19, 24L02, 35N20, 50B24 and 59D14; Tab. 2) due to their location adjacent to each other within a distance of 46.8 kb on the rice physical map [contig 15, Gramene Rice-I-Map (FPC), http://www.gramene.org/].

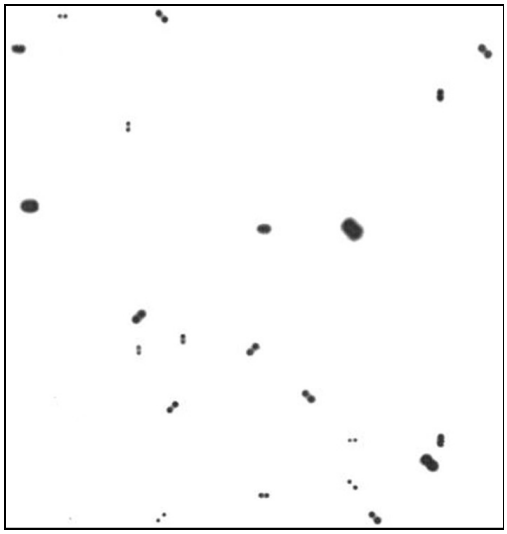

Fig. 2. A sample autoradiogram of a high-density filter (9,216 clones) hybridized with a pool of ten overgo probes, OSog1011 to 1020 (Overgo OSog Pool II). Positive clones were determined if the hybridization signals had double-spot pattern for each clone.

Overgo probe verification and STS-based BAC-contig map development

The subsequent PCR tests linked sixty-nine out of 71 positive clones to 19 overgo probes. The number of positive BACs per probe ranged between three and twelve. The appropriate BAC-probe assignments are available from the authors upon request. Each PCR-amplified fragment was of the size predicted based on the *O. sativa* ssp. *japonica* sequence (data not shown). Based on the overlapping clone information, an STS map with a minimal tilling path of BACs was assembled (Fig. 3). In the case of the OSog1010 locus, although the designed PCR primer set failed to amplify, this did not affect the STS-based BAC-contig map construction. Two clones, 17N09 and 50L10, were not assigned to any probe using the PCR tests. However, it remains unclear whether they were false positives or were due to PCR failure, corresponding to the unassigned OSog1010 locus.

Tab. 2. BAC library screening results for hybridization of two sets of ten overgo pools (Pool I and Pool II). The clones which were positive after hybridizations with both Overgo Pool I and II are given in bold.

Pools of overgo probes (pool designation)	No. of positive clones	Library addresses of positive clones
OSOG1001 TO OSOG1010 (Overgo OSog Pool I)	38	04O19, 06E05, 05I17, 09M04, **10N19**, 16B11, 14K24, 17M05, 17N09, 24B23, 24N24, **24L02**, 24L06, 23J12, 21P22, 26I13, 31B03, 30D08, 32H18, 34N18, 35C12, **35N20**, 36M04, 38K17, 38L21, 40O02, 47B11, **50B24**, 50L10, 51H21, 54C02, 54F04, 53P17, **59D14**, 59L21, 58N10, 23A17, 19K06
OSOG1011 TO OSOG1020 (Overgo OSog Pool II)	38	01A17, 02K14, 04A08, 04D04, 05D22, 06H18, **10N19**, 10O15, 11L15, 12B05, 13D10, 16D02, 15O01, 19D18, 19N17, 21C06, 21I17, 21K10, 21N12, 22G18, 24G10, **24L02**, 27K02, 33O13, 34A04, 34I10, **35N20**, 36J09, 40L20, 44N05, 47O09, 48P03, **50B24**, 53P14, 55C10, 56H10, **59D14**, 60F23
BAC clones picked by both Pool I & II	5	**10N19**, **24L02**, **35N20**, **50B24**, **59D14**

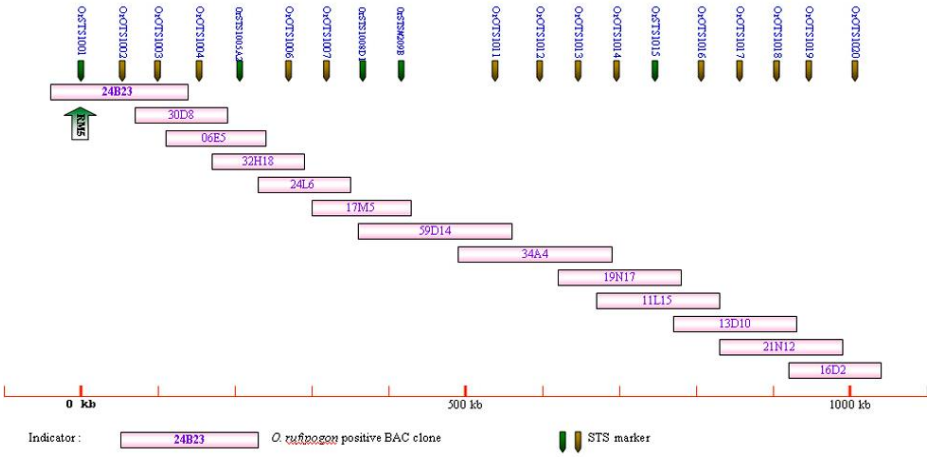

Fig. 3. A minimal tilling path of BAC clones constructed using STS PCR markers.

DISCUSSION

Probe composition is a crucial element in implementing the overgo approach to screen BAC libraries. The OvergoMaker 40 program employed in this study for probe design produced overgos with 40-60% GC content, and their T_m was then checked to ensure that they were within an interval of 67-72°C using the OligoAnalyzer 3.0 program. Some overgo sequences, which were found to have a T_m above 72°C or below 67°C, were discarded and a second candidate region was selected for another round of overgo design procedure. This is a critical point in the whole overgo hybridization procedure, as a T_m outside the prescribed range (67-72 °C) can cause failure of hybridization of a radiolabelled overgo to filter-bound DNA fragments, leading to the selection of non-specific BAC clones. This would also subsequently lead to problems in the washing step after hybridization [15]. Probes with similar T_m can be pooled together for a single hybridization step, as was reported for various partial or whole genome physical mapping projects done for human [16], chicken [17], mouse and zebrafish [12] and some mammalian species [18].

Another important selection criterion in the overgo development procedure is the identification of possible secondary structures and the number of radionucleotide incorporating sites. Checking for hairpins and self-dimers was done using the OligoAnalyzer 3.0 program, while radionucleotide incorporation sites were manually counted. Since ^{32}P-dCTP and ^{32}P-dATP nucleotides were used for preparation of radioactive labelled overgos, only overgos with more than ten G and T bases in the overhanging ends were considered and accepted.

The reported advantages of the overgo hybridization [6] and our experience provided the following additional reasons to choose overgos as hybridization probes: (1) they can be easily and effectively removed from high-density filters by washing at 80°C; (2) calibration of the concentration of individual probes and the radiolabel incorporation test can be omitted due to the consistency of the quality and yield of oligonucleotides from custom synthesis; and (3) large amounts of overgos can be purchased at more reasonable prices than longer oligonucleotides. The only prerequisite for this technique is the availability of suitable sequence data to generate probes. However, the rapid increase in the amount of reported sequence data for various organisms and the co-linearity of genomes among related species would allow this method to be performed for the purpose of cross-species hybridization leading to BAC-contig map construction [11]. Furthermore, the short length feature of overgos is advantageous when there is limited sequence data from which to design probes [8].

This study has demonstrated that the overgo design based on *O. sativa* genomic sequences is highly reliable and specific, and is suitable for *O. rufipogon* library screening. It was found that 74 to 84% of common bands in RFLP/RAPD are shared by both species [19]. SINE (short interspersed element) analysis on both species suggested that the *indica* strain of *O. sativa* might have originated from *O. rufipogon* via recent domestication [20]. A genetic affinity between the two

species was also demonstrated by others authors using AP-PCR [21], Southern and PCR analyses [22], MITE-AFLP [23], and chloroplast-microsatellite variation analysis [24]. Our recent efforts on sequencing 15 of 19 STS PCR products and a BAC 24B23 harbouring OSog1001, 1002 and 1003, have shown that *O. sativa* ssp *japonica* and *O. rufipogon* share a very high similarity (95-99%), at least in the selected *yld1.1* genomic region (B.K. Song *et al.*, unpublished data). This data concurs with the results obtained from the *O. rufipogon* BAC end sequencing done by Rod Wing's group in the targeted *yld1.1* region (Rod A. Wing, Arizona Genomics Institute, unpublished data; http://www.gramene.org/Oryza_sativa/contigview?highlight=&chr=1&vc_start=23891650&vc_end=24054190&x=0&y=0). In view of the above findings, it is reasonable to expect that *O. sativa* genomic sequences from either coding, UTR or intronic regions can be used as a source for our cross-species probe design strategy in *O. rufipogon*, including the ordering and spacing of the overgos.

The success of the overgo hybridization technique relies on generating probes of short sequences (40 bp). This property of the overgo design increases the probability of finding suitable unique regions from the *O. sativa* genomic sequence in GenBank. The overgo design requirements should be very stringent to avoid repeat DNA elements. The use of the RepeatMasker and BLAST programs as an integral part of our overgo design was sufficient to select unique candidate regions which are suitable for overgo design. If the probes contained repetitive sequences, we would see multiple hits after filter hybridization [25]. Input sequences of 5' and 3' UTRs of putative proteins showed a high degree of uniqueness and were found useful as a sequence source for overgo composition. Of the 20 overgo probes, four were designed from 3' UTR, 15 from 5' UTR and only one from the first intron of a putative gene (Tab. 1). The intron region was used to design the OSog1012 probe when we were unable to avoid repetitive sequences in 5' and 3' UTR regions of the selected sequence coding a CENPC-like protein. However, the first intron turned out to be unique and suitable for designing the OSog1012 probe, which, as shown, was able to hybridise to a few BAC clones.

An average of 3.6 positive BACs were picked per overgo probe, which is close to the 3.8-fold coverage of the library represented by the 23,040 clones, with an average BAC insert size of 125 kb and a *O. rufipogon* genome size of 760 Mb (data provided by Amplicon Express). Sixty-nine out of the 71 clones were found to be true positives and this demonstrates that the cross-species probes and primer design strategy employed here was successfully implemented. The efficacy of the overgo probes was further verified by a non-redundant STS-based BAC-contig map that covered approximately 1 Mb of the targeted region and was generated using the PCR confirmation data (Fig. 3).

CONCLUSIONS

The key to the success of performing overgo-based hybridization on a BAC genomic library largely relies on the construction of hybridization probes with

similar hybridization kinetics and unique specificities. The selection of specific sequences with similar GC contents can increase the specific activity of the overgo probes and eliminate overgos that may hybridize non-specifically. The properties of the designed overgos, such as specificity, similar melting temperature, and similar specific activities, allows a set of pooled overgos (up to 236 overgos, [6]) to be effectively used in a single hybridization. The results in this study indicate that the *O. sativa* genomic sequence information is applicable in designing overgo probes for cross-species BAC library screening of the Malaysian wild rice *O. rufipogon*. This further implicates that a BAC-contig physical map of a chromosomal region of interest in *O. rufipogon* may be assembled with a high degree of efficiency based on the *O. sativa* sequence framework. Thus, a comparative genomic approach makes it possible to examine a less studied genome using sequence information from a related species.

Acknowledgements. We would like to thank Dr. Robbie Waugh, Dr. McPherson, Dr. Cliff Han, Dr. Simon Griffiths, Dr. Cha Thye San and Dr. Hu Pingsha for providing us with valuable advice regarding our research. This work is supported by the Malaysian Ministry of Science and Technology, grant number IRPA 01-03-03-0001-BTK/ER/001.

REFERENCES

1. Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiri, Y. and Simon, M. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Echerichia coli* using an F-factor-based vector. **Proc. Natl. Acad. Sci. U.S.A.** 89 (1992) 8794-8797.
2. Ioannou, P.A., Amemiya, C.T., Garnes, J., Kroisel, P.M., Shizuya, H., Chen, C., Batzer, M.A. and de Jong, P.J. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. **Nat. Genet.** 6 (1994) 84-88.
3. Marra, M.A., Kucara, T.A., Green, E. D., Brownstein, B., Wilson, R.K., McDonald, K.M., Hillier, L.W., McPherson, J.D. and Waterston, R.H. High throughput fingerprint analysis of large-insert clones. **Genome Res.** 7 (1997) 1072-1084.
4. Osoegawa, K., Mammoser, A.G., Wu, C., Frengen, E., Zeng, C., Catanese, J.J., and de Jong, P.J. A bacterial artificial chromosome library for sequencing the complete human genome. **Genome Res.** 11 (2001) 483-496.
5. Zhang, H.B. and Wing, R.A. Physical mapping of the rice genome with BACs. **Plant Mol. Biol.** 35 (1997) 115-127.
6. Han, C.S., Sutherland, R.D., Jewett, P.B., Campbell, M.L., Meincke, L.J., Tesmer, J.G., Mundt, M.O., Fawcett, J.J., Kim, U.J, Deaven, L.L. and Doggett, N.A. Construction of a BAC contig map of chromosome 16q by two-dimensional overgo hybridization. **Genome Res.** 10 (2000) 714-721.

7. Hong, G. A rapid and accurate strategy for rice contig map construction by combination of fingerprinting and hybridization. **Plant Mol. Biol.** 35 (1997) 129-133.
8. Ross, M.T., LaBrie, S., McPherson, J. and Stanton Jr., V.P. Screening large-insert libraries by hybridization. In: **Current Protocols in Human Genetics**, (Boyl, A., Ed.), Wiley, New York, 1999, 5.6.1-5.6.5.
9. Arlt, M.F., Li, M., Herzog, T.J., and Goodfellow, P.J. A 1-Mb bacterial clone contig spanning the endometrial cancer deletion region at 1p32–p33. **Genomics** 57 (1999) 62-69.
10. Cai, W.W., Reneker, J., Chow, C.W., Vaishnav, M. and Bradley, A. An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization. **Genomics** 54 (1998) 387-397.
11. Kim, J., Gordon, L., Dehal, P., Badri, H., Christensen, M., Groza, M., Ha, C., Hammond, S., Vargas, M., Wehri, E., Wagner, M., Olsen, A. and Stubbs, L. Homology-driven assembly of a sequence-ready mouse BAC contig map spanning regions related to the 46-Mb gene-rich euchromatic segments of human chromosome 19. **Genomics** 74 (2001) 129-141.
12. Liu, S., Lu, W., Obara, T., Kuida, S., Lehoczky, J., Dewar, K., Drummond, I.A. and Beier, D.R. A defect in a novel Nek-family kinase causes cystic kidney disease in the mouse and in zebrafish. **Development** 129 (2002) 5839-5846.
13. Moncada, P., Martinez, C.P., Borrero, J., Chatel, M., Gauch, H., Jr., Guimaraes, E., Tohme, J. and McCouch, S. R. Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. **Theor. Appl. Genet.** 102 (2001) 41-52.
14. Xiao, J., Li, J., Grandillo, S., Ahn, S.N., Yuan, L., Tanksley, S.D. and McCouch, S.R. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. **Genetics** 150 (1998) 899-909.
15. Brettschneider, R., RFLP analysis. In: **Molecular Tools for Screening Biodiversity — Plants And Animals**, (Karp, A., Isaac, P.G. and Ingram, D.S., Eds.), Chapman & Hall, London, 1998, 83-95.
16. McPherson, J.D., Marra, M., Hillier, L., Waterston, R.H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E.R., Wilson, R.K., Fulton, R., Kucaba, T.A., Wagner-McPherson, C., Barbazuk, W.B., Gregory, S.G., Humphray, S.J., French, L., Evans, R.S., Bethel, G., Whittaker, A., Holden, J.L., McCann, O.T., Dunham, A., Soderlund, C., Scott, C.E., Bentley, D.R., Schuler, G., Chen, H.C., Jang, W., Green, E.D., Idol, J.R., Maduro, V.V., Montgomery, K.T., Lee, E., Miller, A., Emerling, S., Kucherlapati, Gibbs, R., Scherer, S., Gorrell, J.H., Sodergren, E., Clerc-Blankenburg, K., Tabor, P., Naylor, S., Garcia, D., de Jong, P.J., Catanese, J.J., Nowak, N., Osoegawa, K., Qin, S., Rowen, L., Madan, A., Dors, M., Hood, L., Trask, B., Friedman, C., Massa, H., Cheung, V.G., Kirsch, I.R., Reid, T., Yonescu, R., Weissenbach, J., Bruls, T., Heilig, R., Branscomb, E., Olsen, A., Doggett, N., Cheng, J.F., Hawkins, T., Myers, R.M., Shang, J., Ramirez, L.,

Schmutz, J., Velasquez, O., Dixon, K., Stone, N.E., Cox, D.R., Haussler, D., Kent, W.J., Furey, T., Rogic, S., Kennedy, S., Jones, S., Rosenthal, A., Wen, G., Schilhabel, M., Gloeckner, G., Nyakatura, G., Siebert, R., Schlegelberger, B., Korenberg, J., Chen, X.N., Fujiyama, A., Hattori, M., Toyoda, A., Yada, T., Park, H.S., Sakaki, Y., Shimizu, N., Asakawa, S., Kawasaki, K., Sasaki, T., Shintani, A., Shimizu, A., Shibuya, K., Kudoh, J., Minoshima, S., Ramser, J., Seranski, P., Hoff, C., Poustka, A., Reinhardt, R. and Lehrach, H. A physical map of the human genome. The international human genome mapping consortium. **Nature** 409 (2001) 934-941.

17. Romanov, M.N., Price, J.A. and Dodgson, J.B. Integration of animal linkage and BAC contig maps using overgo hybridization. **Cytogenet. Genome Res.** 102 (2003) 277-281.
18. Thomas, J.W., Prasad, A.B., Summers, T.J., Lee-Lin, S.Q., Maduro, V.V.B., Idol, J.R., Ryan, J.F., Thomas, P.J., McDowell, J.C., and Green, E.D. Parallel construction of orthologous sequence-ready clone contig maps in multiple species. **Genome Res.** 12 (2002) 1277-1285.
19. Bautista, N.S., Solis, R., Kamijima, O. and Ishii, T. RAPD, RFLP and SSLP analyses of phylogenetic relationships between cultivated and wild species of rice. **Genes Genet. Syst.** 76 (2001) 71-79.
20. Cheng, C.Y., Motohashi, R., Tsuchimoto, S., Fukuta, Y., Ohtsubo, H. and Ohtsubo, E. Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. **Mol. Biol. Evol.** 20 (2003) 67-75.
21. Yi, Q.M., Deng, W.G., Xia, Z.P. and Pang, H.H. Polymorphism and genetic relatedness among wild and cultivated rice species determined by AP-PCR analysis. **Hereditas** 122 (1995) 135-141.
22. Sun, Q., Wang, K., Yoshimura, A. and Doi, K. Genetic differentiation for nuclear, mitochondrial and chloroplast genomes in common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*Oryza sativa* L.). **Theor. Appl. Genet.** 104 (2002) 1335-1345.
23. Park, K.C., Kim, N.H., Cho, Y.S., Kang, K.H., Lee, J.K. and Kim, N.S. Genetic variations of AA genome *Oryza* species measured by MITE-AFLP. **Theor. Appl. Genet.** 107 (2003) 203-209.
24. Ishii, T., Xu, Y. and McCouch, S.R. Nuclear- and chloroplast-microsatellite variation in A-genome species of rice. **Genome** 44 (2001) 658-666.
25. Yang, Z.N., Ye, X.R., Choi, S., Molina, J., Moonan, F., Wing, R.A., Roose, M.L. and Mirkov, T.E. Construction of a 1.2-Mb contig including the citrus tristeza virus resistance gene locus using a bacterial artificial chromosome library of *Poncirus trifoliata* (L.) Raf. **Genome** 44 (2001) 382-393.