

**GENOMIC ORGANIZATION OF CHROMOSOMAL  
CENTROMERES IN CULTIVATED RICE, *Oryza sativa* L., AND ITS WILD  
PROGENITOR, *O. rufipogon* Griff.**

A Dissertation

by

TAESIK UHM

Submitted to the Office of Graduate Studies of  
Texas A&M University  
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2004

Major Subject: Plant Physiology

**GENOMIC ORGANIZATION OF CHROMOSOMAL  
CENTROMERES IN CULTIVATED RICE, *Oryza sativa* L., AND ITS WILD  
PROGENITOR, *O. rufipogon* Griff.**

A Dissertation

by

TAESIK UHM

Submitted to Texas A&M University  
in partial fulfillment of the requirements  
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

---

HongBin Zhang  
(Chair of Committee)

---

H. J. Price  
(Member)

---

J. T. Cothren  
(Member)

---

H. H. Wilkinson  
(Member)

---

M. L. Binzel  
(Chair of Molecular & Environmental  
Plant Sciences Faculty)

---

M. A. Hussey  
(Head of Department)

August 2004

Major Subject: Plant Physiology

**ABSTRACT**

Genomic Organization of Chromosomal Centromeres in Cultivated Rice, *Oryza sativa* L., and Its Wild Progenitor, *O. rufipogon* Griff. (August 2004)

Taesik Uhm, B.S.; M.S., Seoul National University

Chair of Advisory Committee: Dr. HongBin Zhang

Centromeres are responsible for sister-chromatid cohesion, kinetochore formation, and accurate transmission of chromosomes. Rice provides an excellent model for organizational and functional studies of centromeres since several of its chromosomes contain limited amounts of satellite and other repetitive sequences in their centromeres. To facilitate molecular characterization of the centromeres, we screened several BIBAC and BAC libraries of *japonica* and *indica* rice, using several centromere-specific repeat elements as probes. The positive clones were identified, fingerprinted and integrated into our whole genome physical map databases of the two rice subspecies. BAC/BIBAC-based physical maps were constructed for the centromeric regions of the subspecies. To determine whether the genomic organization of the centromeres has changed since the cultivated rice split from its progenitor and to identify the sequences potentially playing an important role in centromere functions, we constructed a large-insert BIBAC library for the wild progenitor of Asian cultivated rice, *O. rufipogon*. The library contains 24,192 clones, has an average insert size of 163 kb, and covers 5 x haploid genome of wild rice. We screened the wild rice library with two centromere 8-specific overgo

probes designed from the sequences flanking centromere 8 of *japonica* rice. A BIBAC-based map was constructed for wild rice centromere 8. Two of the clones, B43P04 and B15E04, were found to span the entire region of the wild rice centromere and thus selected for sequencing the centromere. By sequencing the B43P09 clone, a 95% genomic sequence of the long arm side of wild rice centromere 8 was obtained. Comparative analysis revealed that the centromeric regions of wild rice have a similar gene content to *japonica* rice, but the centromeric regions of *japonica* rice have undergone chromosomal rearrangements at both large scale and nucleotide levels. In addition, although the 155-bp satellite repeats showed dramatic changes at the middle region, they are conserved at the 5' and 3' ends of satellite monomers, suggesting that those regions might have important functional roles for centromeres. These results provide not only new insights into genomic organization and evolution, but also a platform for functional analysis of plant centromeres.

This dissertation is dedicated to my father TaeYoung Uhm and my Mother Sun-Hee Lee.

## ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. HongBin Zhang, for his full support, encouragement, and constructive suggestions to help me complete my dissertation research.

I would like to appreciate the valuable comments and suggestions given me by my other committee members: Dr. H. James Price, Dr. J. Tom Cothren, Dr. Heather H. Wilkinson.

Thank to Dr. Chengcang Wu, Dr. Mi-Kyung Lee, Dr. Chengwei Ren, Dr. Teofila S. Santos, Ms. Keijao Ding, Ms. Chantel Scheuring, Dr. Yan Bo, and Mr. Felipe Santos for all the help and discussions on my research.

Special thanks to Dr. Zhanyou Xu, Ms. Ying Rong, and Ms. Lina Covaleda for all the help and support on my research.

I would like to express my appreciation and respect to my parents, brothers, and sisters-in-law for their love, support, encouragement.

## TABLE OF CONTENTS

	Page
ABSTRACT.....	iii
DEDICATION.....	v
ACKNOWLEDGEMENTS.....	vi
TABLE OF CONTENTS.....	vii
LIST OF FIGURES.....	ix
LIST OF TABLES.....	xi
 CHAPTER	
I INTRODUCTION.....	1
II BAC/BIBAC-BASED PHYSICAL MAPS OF CENTROMERES OF CULTIVATED RICES, <i>Oryza sativa</i> ssp. <i>japonica</i> AND ssp. <i>indica</i> .....	8
Overview.....	8
Introduction .....	9
Materials and Methods .....	10
Results .....	14
Discussion .....	24
III ONE LARGE INSERT PLANT-TRANSFORMATION- COMPETENT BIBAC LIBRARY OF THE WILD RICE, <i>Oryza rufipogon</i> , AND IDENTIFICATION OF THE CLONES DERIVED FROM THE REGION OF ITS CHROMOSOME 8 CENTROMERE.....	27
Overview.....	27
Introduction .....	28
Materials and Methods .....	30
Results .....	35

CHAPTER	Page
Discussion .....	39
IV SEQUENCE ANALYSIS OF THE CHROMOSOME 8 CENTROMERE OF THE WILD RICE, <i>Oryza rufipogon</i> GRIFF .....	43
Overview.....	43
Introduction .....	44
Materials and Methods .....	47
Results .....	50
Discussion .....	72
V SUMMARY .....	77
REFERENCES .....	82
VITA.....	94



## LIST OF FIGURES

FIGURE	Page
2.1 An example of centromeric BAC contigs in <i>japonica</i> rice.....	17
2.2 A BAC contig (ctg4) containing the entire region of a centromere in <i>japonica</i> rice.....	20
2.3 Contig (ctg2) spanning the chromosome 3 centromere of <i>japonica</i> rice determined by the sequences of BAC H023H21 ends.....	21
2.4 Contig (ctg1) of chromosome 8 centromere of <i>japonica</i> rice determined by chromosome-specific overgo hybridization.....	22
2.5 A contig (ctg7) anchored to the chromosome 8 centromere of <i>indica</i> rice by chromosome centromere-specific overgo hybridization.....	23
3.1 PFGE analysis of randomly picked BAC clones of the wild rice <i>Bam</i> HI large insert library.....	37
3.2 Insert size distribution of the clones in the wild rice <i>Bam</i> HI Library.....	38
3.3 Southern analysis of 12 positive clones obtained by library hybridization using the centromere 8-specific overgo probes (RER1 and TGF).....	41
4.1 The BAC contig map of the wild rice centromere 8.....	52
4.2 Minimal overlapping BAC tile path of the centromeric region of the wild rice chromosome 8.....	53
4.3 zPicture analysis for rice <i>RER1B</i> gene.....	57
4.4 zPicture analysis for the sequence contigs of the wild rice.....	58
4.5 zPicture analysis of the sequence contig 12 of the wild rice.....	63
4.6 zPicture and eShadow analyses of the centromeric regions between the wild rice and cultivated <i>japonica</i> rice.....	64

FIGURE	Page
4.7 Sequence similarity between the wild rice CentOr and <i>japonica</i> rice CentO satellite monomers.....	68
4.8 Comparison between alignments of the wild rice 155-bp CentOr and <i>japonica</i> rice 155-bp CentO satellite monomers.....	70

**LIST OF TABLES**

TABLE		Page
2.1	The number of centromeric contigs in the <i>japonica</i> and <i>indica</i> rice database .....	18
2.2	Centromeric BIBAC/BAC contigs of <i>japonica</i> and <i>indica</i> rice.....	19
3.1	BAC clones positive to centromere 8-specific probe RER1 and TGF, and CentO satellite repeats.....	40
4.1	Sequence contigs constituting the sequence map of B43P09 of the wild rice and comparison with the centromere 8 sequence of <i>japonica</i> rice .....	55

## CHAPTER I

### INTRODUCTION

Centromeres are essential for chromosomal sister-chromatid cohesion, kinetochore formation, and faithful transmission in higher organisms. Unlike the well-characterized chromosomal centromeres of the single cell yeast, *Saccharomyces cerevisiae* (Hegemann and Fleig 1993), the chromosomal centromeres of most animals and plants consist of long stretches of short (155 – 340 bp) tandem satellite repeats and various other kinds of repetitive DNA sequences (Henikoff et al. 2001). Because they are flanked by pericentromeric regions, their exact boundaries remain undefined. The pericentromeric regions contain middle repetitive elements, including transposons, retrotransposons, and a few pseudogenes (J Wu et al. 2004; Zhang et al. 2004). For instance, rice core centromeres comprise tandem satellites, occasionally interrupted by other repeated sequences (J Wu et al. 2004; Zhang et al. 2004). Some regions of tandem repeated satellites are thought to be responsible for the function of the centromeres (Henikoff et al. 2001).

The successful assembly of human artificial chromosomes (HAC) using either synthetic or cloned  $\alpha$ -satellite DNA suggested that the satellite sequences have conferred centromere functions in human cell lines (Harrington et al. 1997; Henning et al. 1999;

---

This dissertation follows the style of Theoretical and Applied Genetics.

Ikeno et al. 1998). In maize B chromosomes, the partial deletions of ~500 kb of tandem repeats reduced the chromosome transmission (Kaszas and Birchler 1998). However, the characterization of neocentromeres that completely lack  $\alpha$ -satellite DNA has indicated that under circumstances  $\alpha$ -satellite sequences are not necessary for centromere functions (Barry et al. 2000; DuSart et al. 1997).

Several centromere-specific repetitive DNA elements have been isolated in rice (Dong et al. 1998; Nonomura and Kurata 1999). One of the elements is the CentO repeat. It consists of 155-bp satellite tandem monomers and has been located exclusively in rice chromosomal centromeres (Cheng et al. 2002; Dong et al. 1998). The other centromere-specific elements include RCE1 (Nonomura and Kurata 1999) and pSau3A9-like sequences (Jiang et al. 1996). These elements are derived from the Ty3/gypsy-class retrotransposon family and are localized in the centromeric regions of chromosomes of not only rice, but also other grass species (Langdon et al. 2000; Miller et al. 1998; Presting et al. 1998). Together, they are named as centromeric retrotransposon (CR) elements. The chromatin immunoprecipitation assay with centromeric histone H3 proteins (*CENH3*) in maize demonstrated that some of CR elements and satellites interacted with the *CENH3* proteins, indicating that some of CR elements and satellites are likely to participate in the function of centromeres (Zhong et al. 2002). Most of the CR elements were found as fragmented and truncated retroelements in the rice centromere. In the recently sequenced rice centromere 4, for instance, only one intact retrotransposon was obtained (Zhang et al. 2004).

Cytological analysis by fluorescent *in situ* hybridization (FISH) showed that the rice CentO satellite repeats are located within the spindle fiber-binding regions and the functional domain of the centromeres, by use of the misdivision break points of telocentric rice chromosomes, were mapped to the middle of the CentO arrays. In comparison, the centromere-specific retrotransposon family, CR elements, is widely dispersed in the centromeric regions (Cheng et al. 2002; J Wu et al. 2004; Zhang et al. 2004).

The full length of centromeres in higher eukaryotic organisms is generally at the megabase level. The long stretch of repeated DNA sequences such as centromeric satellites in the centromeres makes it difficult to clone and sequence. Therefore, while the human and Arabidopsis genomes have been sequenced by the clone-by-clone approach (Arabidopsis Genome Initiative 2000; International Human Genome Sequencing Consortium 2001), big gaps still remain to be sequenced in the centromeric regions of the physical maps for each chromosome of the species.

Unlike Arabidopsis and human, rice provides an excellent model for structural, organizational, functional and evolutionary studies of centromeres because several of the rice chromosomes contain a limited amount of satellite repeats in a size range from 65 to 350 kb as detected by intensities of FISH signals (Cheng et al. 2002). Recently, the whole regions of rice chromosome 4 and 8 centromeres have been sequenced (Nagaki et al. 2004; J Wu et al. 2004; Zhang et al. 2004). Compared to centromere 4, centromere 8 contains similar copy numbers of 155-bp CentO satellite monomers in the core region,

but has more genes and fewer CR-related repeated elements in the pericentromeric regions.

Centromeres have been recognized as rapidly evolving regions of chromosomes involving large- and small-scale changes. Although the functions of centromeres are highly conserved, there is low sequence identity among the centromeres of the distantly-related species (Hall et al. 2004). It is little known about the chromosomal rearrangements of the centromeric regions in plants. In *Arabidopsis*, the insertions of some mitochondrial sequences and 5S rDNA into centromeres (Franz et al. 1998; Stupar et al. 2001), and inversions involving centromere 4 (*Arabidopsis* Genome Initiative 2000) have been reported. By contrast, the rapid changes of satellite repeats at the nucleotide level have been characterized (Cheng et al. 2002; Talbert et al. 2002; Heslop-Harrison et al. 2003; Hall et al. 2003). Several hypotheses have been suggested to explain how satellite sequence homogeneity can be maintained while rapid evolution is allowed (Ugarkovic and Plohl 2002; Hall et al. 2004). In the library hypothesis, genomes contain sets of satellite variants in differing abundance. New satellites continuously originate by mutation. These satellites can be homogenized through unequal crossover (Nijman and Lenstra 2001; Hall et al. 2004). Another model suggested that satellite evolution is driven by the selection and co-evolution of satellites and centromeric histone H3 proteins (Henikoff et al. 2001). The centromeric histone H3 proteins (CENP-A) specifically bind to satellite DNA. In such a case, a slight advantage in satellite-CENP-A interactions could lead to rapid genomic fixation of satellite arrays. The rapid

adaptive evolution of CENP-A has been observed in both *Drosophila* and *Arabidopsis* (Malik et al. 2002; Talbert et al. 2002).

The common wild rice, *Oryza rufipogon* Griff., is the ancestor of Asian cultivated rice (*Oryza sativa* L.), including ssp. *japonica* and ssp. *indica*, and thus, is the most important germplasm for cultivated rice genetic improvement (Oka 1988). The discovery of cytoplasmic male-sterility (CMS) gene in *O. rufipogon* has led to the development of high yielding hybrid rice varieties by its introduction to cultivated rice (Yuan et al. 1989). Other agronomically valuable genes, such as tolerance to aluminum and acid sulfate soil, resistance to bacterial blight and tungro virus, and elongation ability found in the wild rice, are of great importance for cultivated rice breeding (Xiao et al. 1996; Brar and Khush 1997; Bellon et al. 1998).

The wild rice has the same genome (AA) as the Asian cultivated rice. There have been numerous reports of genetic differences between cultivated plants and their wild progenitors. Evolution from wild to cultivated rice has led to “domestication syndrome” (Harlan 1975) common to many cereals. Studies of QTLs showed that most of the genetic factors (qualitative or QTLs) controlling the domestication-related traits are concentrated in a few chromosomal blocks of cultivated rice. The domestication of rice is thought to be associated with the clustered genetic factors (Xiong et al. 1999; Cai and Morishima 2003). However, this does not exist in the wild rice because the genetic factors are not clustered on a chromosome (Xiong et al. 1999).

We report here bacterial artificial chromosome (BAC)/plant-transformation-competent binary bacterial artificial chromosome (BIBAC)-based physical maps of



*indica* and *japonica* rice centromeres, constructed by using a high-resolution sequencing gel-based fingerprinting method (Chang et al. 2001; Li et al. 2004). To facilitate the functional analysis of the centromere, we used BIBACs cloned in a binary vector (pCLD04541) that is competent for direct transformation of the cloned centromeric sequences into plants by the *Agrobacterium*-mediated genetic transformation. The fingerprinting method provided higher resolution and used more restriction enzymes than that used in the previously physical mapping of the *japonica* rice genome (Sasaki et al. 2002). Thus, this allowed generation of physical maps for the centromeres of the rice chromosomes efficiently. In addition, using sequence-tagged sites (STSSs) and centromere 8-specific probes, RER1 and TGF, we successfully anchored several centromeric contig maps to rice chromosomes.

The chromosome 8 has the smallest centromere and fewer heterochromatin regions than other chromosomes of rice (Cheng et al. 2002). Recently, the structural organization and DNA sequences of chromosome 8 centromere of the cultivated rice, *O. sativa* ssp. *japonica*, have been reported (Nagaki et al. 2004; J Wu et al. 2004). In this study we have constructed a large-insert plant-transformation-competent BIBAC library of the wild rice, *O. rufipogon*, and identified the large-insert BIBAC clones from the centromeric region of chromosome 8 of the wild rice. The library and clones provide essential resources for studies of centromere function and evolution in the cultivated rice and wild rice, and for genetic improvement by genetic engineering of cultivated rice using wild rice.

Up to now, limited comparisons of centromere DNA from different species have been performed. However, those studies were insufficient to identify the conserved functional elements and elucidate evolutionary significance of structural changes in plant centromeres. For better understanding of the function and evolution of the centromeric elements in closely-related species, we constructed a physical contig map of the wild rice centromere 8 spanning the whole region of the centromere, and have sequenced the long arm side of the centromere in wild rice. Comparative sequence analysis between the cultivated *japonica* rice and wild rice indicated that there were large-scale rearrangements in the pericentromeric regions and rapid small-scale evolution of 155-bp satellites in the core region.

**CHAPTER II**  
**BAC/BIBAC-BASED PHYSICAL MAPS OF CENTROMERES OF**  
**CULTIVATED RICES, *Oryza sativa* ssp. *japonica* AND ssp. *indica***

**Overview**

Physical mapping of centromeres with BAC clones will facilitate not only studies of their origin, organization and evolution, but also their functional analysis and the development of plant artificial chromosomes for crop genetic improvement. Toward these ends we developed the BAC and plant-transformation-competent BIBAC-based physical maps of centromeres of *Oryza sativa* ssp. *japonica* and ssp. *indica*. To construct the physical maps, we screened several BAC and BIBAC libraries of the *indica* and *japonica* rice, using several centromere-specific repeat elements, including two CR (centromeric retrotransposon) elements and the CentO (a 155-bp satellite tandem monomer of *Oryza sativa*) repeat as probes. The positive clones were analyzed by the high-resolution sequencing gel-based fingerprinting method and then integrated into our whole-genome physical map database of the two subspecies. Analysis of the centromere region-specific clones resulted in 17 and 26 contigs for the centromeres of *japonica* rice and *indica* rice, respectively. Among these contigs, four were shown to span the entire region of four centromeres and three were anchored to the corresponding chromosomes by BAC-end sequence alignment and chromosome-specific overgo (overlapping oligonucleotide) hybridization. The results will provide a platform for

molecular characterization and functional analysis of rice centromeres as well as a tool for filling the physical gaps of the rice genome sequence maps in the centromeric regions.

## **Introduction**

Centromeres are essential for the faithful segregation and inheritance of genetic information in higher organisms. In cereal plants, major components of centromeres are satellite tandem repeats and CR elements derived from ty3/gypsy retrotransposon family ( Miller et al. 1998; Presting et al. 1998; Langdon et al. 2000). The recently sequenced rice centromeres 4 and 8 showed different organization of CentO repeats and gene content (J Wu et al. 2004; Zhang et al. 2004). The rice centromere 8 has three large clusters of CentO satellites, containing a total of 442 CentO monomers, in the core region and several putative genes, such as transforming growth factor (TGF)-beta receptor-interacting protein, putative endoplasmic reticulum retrieval protein (RER1), defective chloroplast and leaf (DCL) protein, and retrotransposons, in the pericentromeric regions. In contrast, the rice centromere 4 has 18 smaller clusters of CentO satellites, containing a total of 379 CentO monomers, in the core region and only retrotransposons in the pericentromeric regions.

The international rice genome sequencing project (IRGSP, <http://rgp.dna.affrc.go.jp/IRGSP>) is sequencing the rice genome. However, due to the limitations of the agarose gel-based, single-enzyme fingerprinting method used for the

construction of the sequencing-ready BAC contig maps, the physical contig maps were successfully constructed only for the centromeres of chromosomes 4 and 8 (Nagaki et al. 2004; J Wu et al. 2004; Zhang et al. 2004). No physical contig maps have been reported so far for the centromeres of other chromosomes. The absence of the contig maps for the centromeric regions has limited not only sequencing of the entire rice genome, but also studies of genomic origin, organization, evolution and function of centromeres.

We report here BAC/BIBAC-based physical maps of *indica* and *japonica* rice centromeres, constructed by using a high-resolution and more information-content sequencing gel-based fingerprinting method (Chang et al. 2001; Li et al. 2004). To facilitate the functional analysis of the centromeres, we used BIBACs cloned in a binary vector (pCLD04541) that is competent for direct transformation of the cloned centromeric sequences into plants by the *Agrobacterium*-mediated genetic transformation. Additionally, using sequence-tagged sites (STSs) and centromere 8-specific probes, we successfully anchored several centromeric contig maps to rice chromosomes.

## **Materials and Methods**

### Source BIBAC/BAC libraries and the database of whole physical maps

One BIBAC and three BAC libraries of *O. sativa* L. ssp. *japonica* cv. Nipponbare (Tao et al. 2002) and three BAC libraries of *O. sativa* L. ssp. *indica* cv. Teqing (Zhang et al.

1996; Tao et al. 2001) were used to identify BACs and BIBACs derived from the centromeric regions of *japonica* rice and *indica* rice, respectively. The libraries of *indica* rice were constructed in pBeloBAC11 (Kim et al. 1996) or its derivative pECBAC1 (Frijters et al. 1997) from nuclear DNA partially digested with three restriction enzymes, *Bam* HI, *Hind* III and *Eco* RI, respectively (Zhang et al. 1996; Tao et al. 2001). The libraries of *japonica* rice were constructed in pECBAC1 and pCLD04541 (Tao and Zhang 1998; Tao et al. 2002) from nuclear DNA partially digested with *Bam* HI, *Hind* III and *Eco* RI, respectively. The vector pCLD04541 is a plant transformation-competent binary BIBAC or TAC vector (Tao et al. 1998) and thus its large-insert clones can be directly transformed into plants via *Agrobacterium*.

In our previous studies, whole-genome physical maps of *indica* rice (Tao et al. 2001) and *japonica* rice (Li et al. 2004) were constructed from the BAC and BIBAC libraries. The databases of the whole physical maps (Tao et al. 2001; Li et al. 2004) were integrated with the fingerprinting data of the positive centromeric clones to generate centromeric physical maps.

### Library screening

For the identification of centromeric BAC clones, the four *japonica* and three *indica* rice libraries were robotically double-spotted onto nylon membrane Hybond N+ in a 3 x 3 format and screened with centromere-specific probes, RCE1 (rice centromere element 1) (Nonomura and Kurata 1999), pSau3A9 (Jiang et al. 1996), and CentO (RCS2) (Dong et

al. 1998), respectively. The DNA used for the probes was generated from rice genomic DNA using three pairs of primers specific to each of the centromeric region-specific repeats, respectively, designed using web-based primer3 software ([http://cbr-rbc.nrc-cnrc.gc.ca/cgi-bin/primer3\\_www.cgi](http://cbr-rbc.nrc-cnrc.gc.ca/cgi-bin/primer3_www.cgi)): RCE1 (L) 5'-TGGAATCAAAATGTTCAAAA-3' and (R) 5'-TGGAATCAAAATGTTCAAAA-3'; RCS2 (L) 5'-TTTGATTGGAAGAAACAGGT-3' and (R) 5'-TCTCGACACTCAAGGCTATT-3'; pSau3A9 (L) 5'-GATCTTTGGATTGGAAA-3' and (R) 5'-GATCCATCTAAAAATATA-3'. The library screening hybridization was conducted according to Zhang et al. (1996). The filters were washed in 2x SSC and 0.1% SDS two times at 65°C with gentle shaking, 20 min each wash.

To identify the contig of the chromosome 8 centromere, two pairs of overgo primers were designed from two putative TGF-beta receptor-interacting protein (TGF) and endoplasmic reticulum retrieval protein (RER1) genes in the rice chromosome 8 BAC clone AY360388 (Nagaki et al. 2004): RER1 (F) 5'-GTGAGTTCTAGGAGAGTAGCTTG-3' and (R) 5'-GAAATAATATCCTGCCAAGCTAC-3'; TGF (F) 5'-GAACCGGCGATAAATACCTTGC-3' and (R) 5'-AAATCGAGAAGCCTGCAAGGTA-3'.

Overgo labeling was prepared according to a protocol developed by the California Genome Research Laboratory (CGRL) (<http://informa.bio.caltech.edu/protocols/overgo.html>), using radioactive [<sup>32</sup>P]-dCTP. Hybridization was carried out according to Sambrook et al. (1989).

### BAC fingerprinting

BAC DNA was isolated and fingerprinted according to Chang et al. (2001) and Tao et al. (2001). The DNA was digested with *Hind*III and *Hae*III, and the *Hind*III fragments were end-labeled with [<sup>33</sup>P]-dATP using reverse transcriptase at 37°C for 2 h, and then subjected to 3.5% (w/v) polyacrylamide DNA sequencing gel electrophoresis at 90 W for 100 min. The *Sau*3AI-digested lambda DNA was used as the standard marker. The gel was dried and autoradiographed.

### Contig assembling

Fingerprint editing and contig assembly were conducted using computers equipped with Linux 7.0 platform (Ren et al. 2003a). According to Chang et al. (2001), only the bands ranging from 58 to 773 bp were used for contig assembly. Vector bands were removed manually from the data. The computer program FPC V6.0 was used for contig assembly (Soderlund et al. 2000). For contig assembly and mergence, a fixed tolerance of 2 and a range of cutoffs from 1e-10 to 1e-22 were used. The other parameters used were Diff = 0.3, MinBands = 5, diffbury = 0.10, and Minends = 8.

### BAC end sequencing

BAC clones were inoculated into 96-deep well blocks containing 1.5 ml of Terrific broth



(Invitrogen, USA) per well, and grown overnight at 37°C with shaking at 325 rpm. DNA was extracted by AutoGenprep 960 (Autogen, USA), an automated DNA isolation robotic workstation. The DNA was dissolved in 50 µl distilled water. Sequencing reaction mixture included 2 µl BigDye terminator (ABI, USA), 300 - 500 ng of template DNA, 1 µl 50 µM primer (M13/pUC forward primer: 5'-GTAAAACGACGGCCAGT-3' and M13/pUC reverse primer: 5'-AACAGCTATGACCATG-3'), and 5 µl distilled water added to 10 µl. The PCR cycling reactions were conducted under the following conditions: 95°C for 4 min, then 95°C, 15 s; 51°C, 10 s and 60°C, 4 min for 70 cycles. The reaction mixture was cleaned up by ethanol precipitation, followed by 70% ethanol wash. Sequencing was carried out on the ABI 3100 Genetic Analyzer (Applied Biosystems, USA). Sequence trimming was conducted using the Sequencher v4.1 (Gene Codes Corporation, USA).

## **Results**

### Identification of centromeric clones in *japonica* and *indica* rice

Several centromere region-specific repetitive sequences were previously cloned from rice and other grass species, including the sorghum pSau3A9, *japonica* rice RCE1, and *indica* rice CentO (RCS2) DNA sequences (Jiang et al. 1996; Dong et al. 1998; Nonomura and Kurata, 1999). Using these sequences as probes, we screened the four *japonica* and three *indica* rice BAC and BIBAC libraries. As a result, we identified

1,423 pSau3A9-positive clones, 1,753 RCE1-positive clones, and 309 CentO (RCS2)-positive clones from 20x genome-coverage *japonica* BAC libraries. Using the sequencing gel-based fingerprinting method and two restriction enzymes, *HindIII* and *HaeIII*, we fingerprinted 984 of the pSau3A9-positive clones, 463 of the RCE1-positive clones, and 208 of the CentO-positive clones, and integrated into our database of whole-genome physical map of *japonica* rice (Li et al. 2004).

From 10x genome-coverage *indica* rice BAC libraries, we identified 370 pSau3A9-positive clones, 1097 RCE1-positive clones, and 344 CentO-positive clones. The information of the positive clones has been integrated into our database of whole-genome physical map of *indica* rice (Tao et al. 2001).

#### BAC/BIBAC-based physical maps of centromeres of *japonica* and *indica* rice

To generate the BAC-based contig maps for the centromeric regions of rice, we disassembled the existing whole physical maps of *indica* rice and *japonica* rice (Tao et al. 2001; Li et al. 2004), integrated the fingerprinting data of centromere-specific clones to the database of *indica* and *japonica* rice, respectively, and re-assembled the physical map contigs of *indica* rice and *japonica* rice using the FPC program v 6.0 (Soderlund et al. 1997, 2000). The contigs that contain the centromeric region-specific repeated sequences were selected, manually edited and merged using the FPC program. For the map contig assembly, a fixed tolerance = 2 and Sulston score cutoffs = 1e-11 to 1e-14

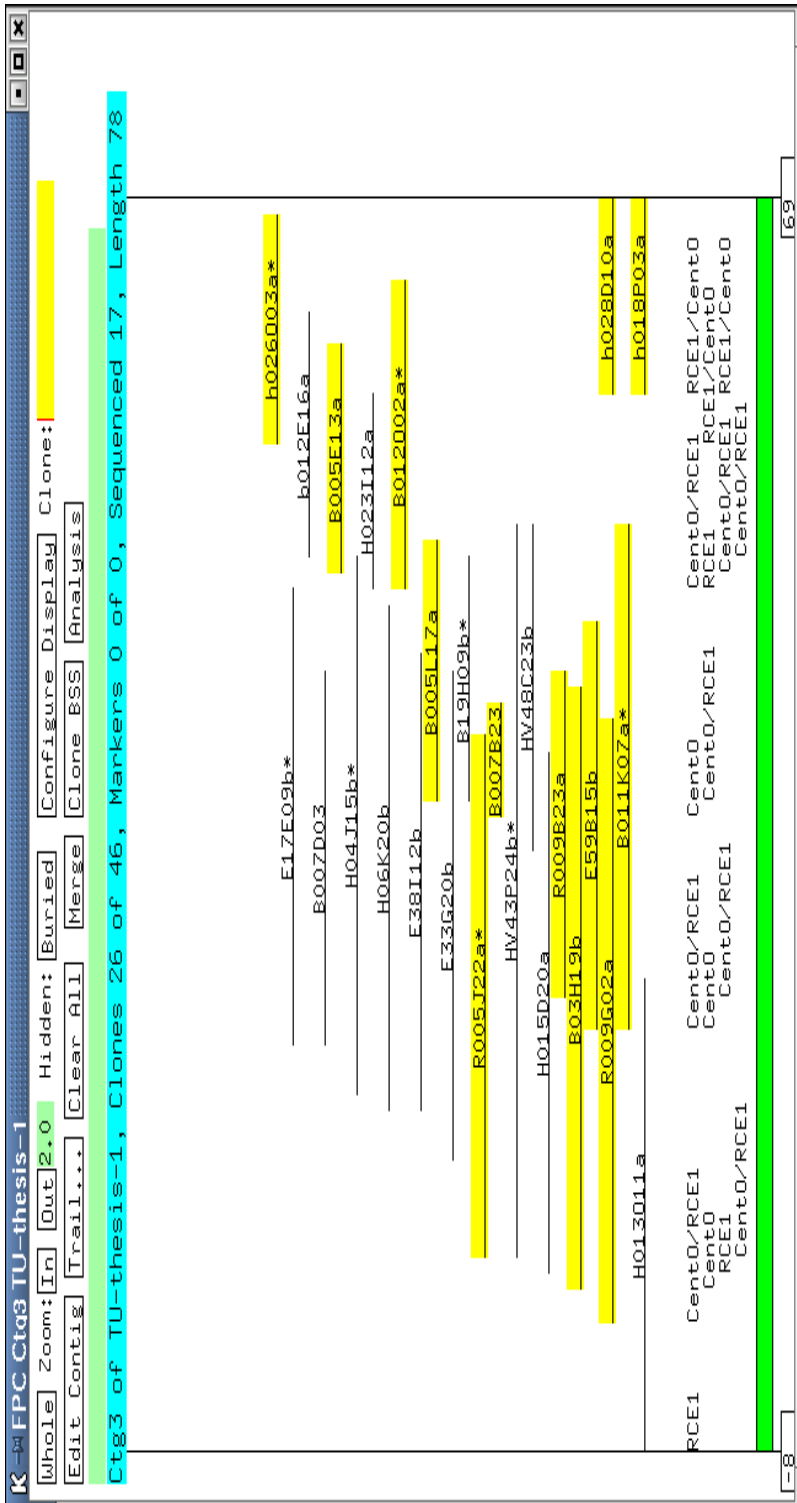
were used, and the mergence of contig-contig was conducted at tolerance = 2 and cutoffs from 1e-10 to 1e-12.

We assembled 17 contigs for the centromeres of the *japonica* rice genome, collectively spanning about 20.9 Mb in physical length, and 26 contigs for the centromeres of the *indica* rice genome, spanning a total of approximately 15.8 Mb in physical length (Fig. 2.1, Table 2.1, and Table 2.2). All of these contigs were observed to contain the CentO satellite repeats. Four of the contigs in both rice subspecies were confirmed to contain the whole regions of the core centromeres (Figs. 2.2, 2.3, 2.4, and 2.5).

#### Chromosome anchoring

To anchor the centromeric contig to its corresponding chromosome, both ends of two flanking BAC clones from each centromeric contig of *japonica* rice were sequenced. Three of the 17 contigs were successfully anchored to their chromosomes, but other BAC end sequences could not be identified in the GenBank due to repetitive sequences of the ends, still ongoing status of rice genome sequencing, or physical gaps in the centromeres. The centromere 3 contig anchored by BAC-end sequencing showed that it contains the core region of the centromere 3, which consists of long arrays of CentO satellite repeats (Fig. 2.3).

The sequence of the entire chromosome 8 centromere of *japonica* rice was recently determined by J Wu et al. (2004). To identify the centromere 8 contig(s) in our



**Fig. 2.1** An example of centromeric BAC contigs in *japonica* rice. The contig 3 consists of 46 clones and 78 unique bands, spanning about 423 kb in physical length according to Li et al. (2004). The clones suffixed with “a” contain RCE1 elements, the clones suffixed with “b” contain pSau3A9, the clones highlighted in yellow contain CentO satellite repeats, and the remaining clones were not studied. The clones with a combination of “a”, “b”, or “yellow” contain two or three repeat elements. The clones prefixed with “R” or “E” were from the *Eco*RI BAC library, with “HV” or “H” from the *Hind*III library, and with “B” from the *Bam*HI library.

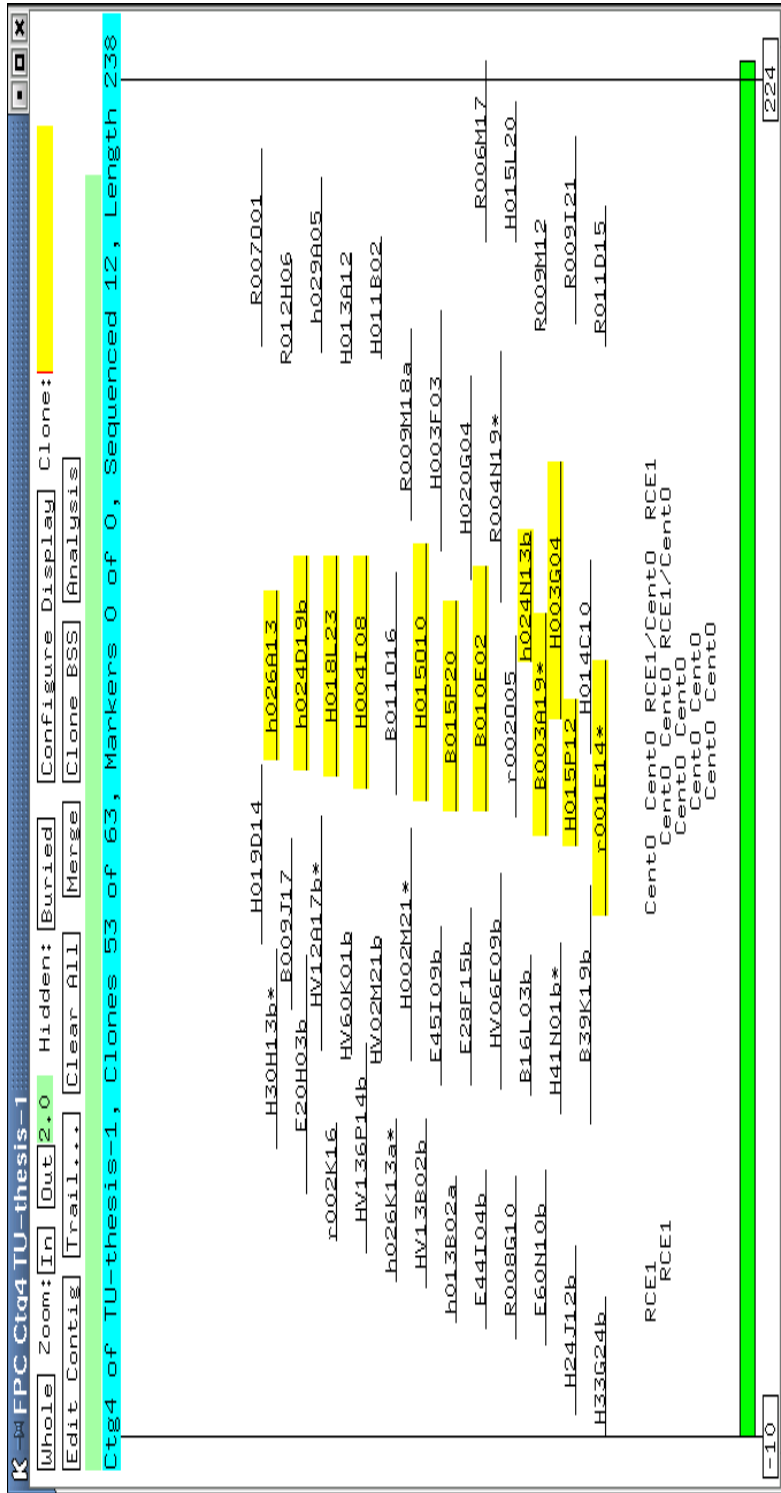
**Table 2.1** The number of centromeric contigs in the *japonica* and *indica* rice database.

	<i>japonica</i>	<i>indica</i>
Centromeric Contigs	17	26
Pericentromeric contigs	86	152
Total	103	178

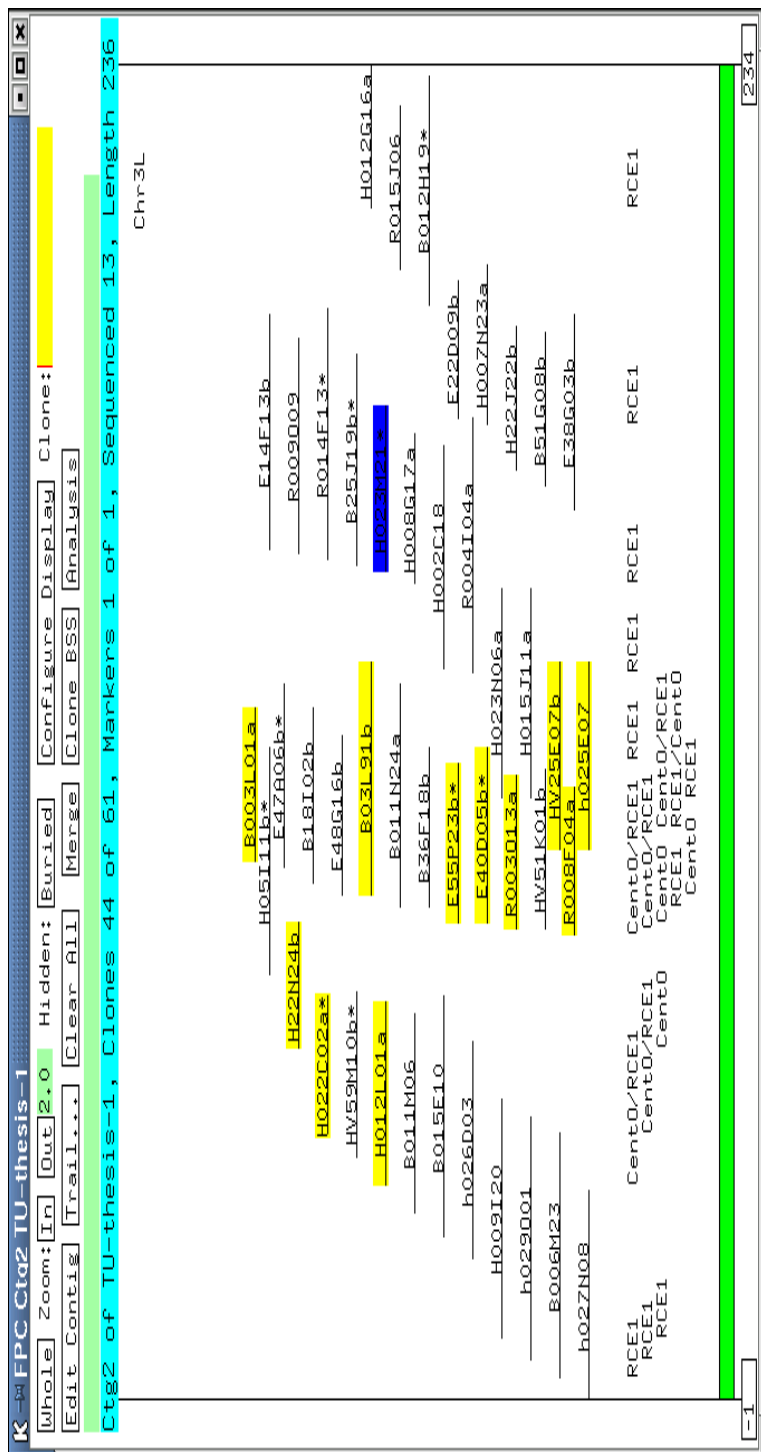
**Table 2.2** Centromeric BIBAC/BAC contigs of *japonica* and *indica* rice.

<i>Japonica</i> rice			<i>Indica</i> rice		
Contig	Clone	kb	Contig	Clone	kb
1*	132	1902	1	43	1064
2*	61	1279	2	38	1008
3	46	422	3	32	951
4	63	1289	4	37	1052
5*	83	2070	5	26	585
6	67	1512	6	43	1121
7	35	856	7*	53	1008
8*	229	5132	8	29	970
9	96	2010	9	5	233
10	19	677	10	28	693
11	11	211	11	22	730
12	17	531	12	7	163
13	24	623	13	25	693
14	6	368	14	10	189
15	10	189	15	5	138
16	39	650	16	4	56
17	35	1252	17	29	970
<b>Total</b>	<b>973</b>	<b>20973</b>	18	7	302
			19	5	56
			20	37	1234
			21	40	1285
			22	7	81
			23	7	390
			24	4	151
			25	13	434
			26	7	308
			<b>Total</b>	<b>563</b>	<b>15865</b>

Asterisk (\*) indicates the anchored BAC/BIBAC contigs.



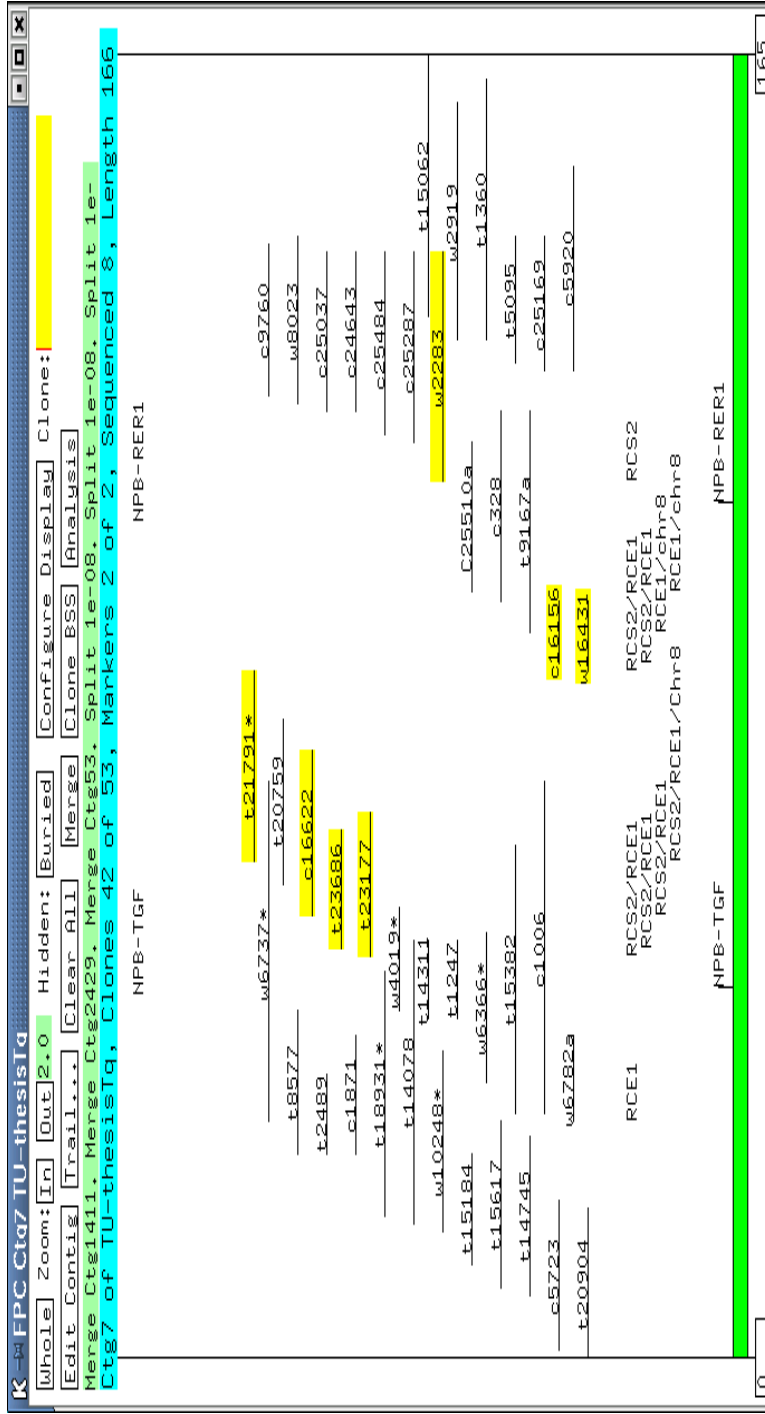
**Fig. 2.2** A BAC contig (ctg4) containing the entire region of a centromere in *japonica* rice. Due to repetitive sequences at the selected BAC ends, the contig was not anchored. The contig has 63 clones and represent *ca.* 1,291 kb in physical length. The clones suffixed with “a” contain RCE1 elements, the clones suffixed with “b” contain pSau3A9, the clones highlighted in yellow contain CentO satellite repeats, and the remaining clones were not studied. The clones with a combination of “a”, “b”, or “yellow” contain two or three repeat elements. The clones prefixed with “R” or “E” were from the *EcoRI* BAC library, with “HV” or “H” from the *HindIII* library, and with “B” from the *BamHI* library.



**Fig. 2.3** Contig (ctg2) spanning the chromosome 3 centromere of *japonica* rice determined by the sequences of BAC H023H21 ends. This contig consists of 61 clones and 236 unique bands, spanning *ca.* 1,281 kb in physical length. The clones suffixed with “a” contain RCE1 elements, the clones suffixed with “b” contain pSau3A9, the clones highlighted in yellow contain CentO satellite repeats, and the remaining clones were not studied. The clones with a combination of “a”, “b”, or “yellow” contain two or three repeat elements. The clone highlighted in blue was sequenced by BAC end sequencing to generate STs for anchoring the contig to its corresponding chromosome. The clones prefixed with “R” or “E” were from the *EcoRI* BAC library, with “HV” or “H” from the *HindIII* library, and with “B” from the *BamHI* library.







**Fig 2.5** A contig (ctg7) anchored to the chromosome 8 centromere of *indica* rice by chromosome centromere-specific overgo hybridization. The contig consists of 53 clones and 166 unique bands, spanning ca. 1008 kb in physical length (Tao et al. 2001). TGF and RER1 are chromosome 8 centromere-specific overgo probes obtained from *japonica* rice. The clones suffixed with “a” contain RCE1 elements, the clones highlighted in yellow contain CentO satellite repeats, and the remaining clones were not studied.

physical maps, two centromere 8-specific overgo probes (RER1 and TGF) were obtained from the flanking regions of the centromere 8. Using the specific overgo probes, we identified the BAC contigs of centromere 8 in *japonica* and *indica* rice (Fig. 2.4 and 2.5). The physical length of the centromere 8 contig in *japonica* rice showed the same physical lengths to the centromere 8 determined by genome sequencing (J Wu et al. 2004).

## **Discussion**

Centromere is one of the most important functional elements of eukaryotic chromosomes, and responsible for proper chromosome division and transmission of genetic materials. Due to large amounts of repetitive sequences and long arrays of satellite tandem repeats in the centromeric regions of higher eukaryotic organisms, there are many physical gaps in the contig maps of the *Arabidopsis* and human genomes (Arabidopsis Genome Initiative 2000; International Human Genome Sequencing Consortium 2001). We explored in this study the feasibility of constructing the physical maps for the centromeric regions of rice using a DNA sequencing gel-based two-enzyme fingerprinting method that has been used in the construction of the genome physical maps of *indica* rice (Tao et al. 2001), *Arabidopsis* (Chang et al. 2001), *japonica* rice (Li et al. 2004), chicken (Ren et al. 2003a) and soybean (Wu et al. 2004a).

We assembled 17 contigs for the centromeres of the *japonica* rice genome, collectively spanning *ca.*20.9 Mb in physical length, and 26 contigs for the centromeres

of the *indica* rice genome, spanning a total of *ca.* 15.8 Mb in physical length (Fig. 2.1, Table 2.1 and Table 2.2). In our physical maps of centromeres, the positive BAC clones to the centromere-specific CR probes, RCE1 and pSau3A9, derived from centromere-specific type3/gypsy retrotransposon family in cereal plants, are distributed mainly on the pericentromeric regions of the chromosomes, but some are observed to locate on the contigs of the other chromosomal regions. The majority of the CR elements flank the CentO arrays in the physical maps of rice centromeres. In the contrast, our physical maps showed that all of the CentO-positive clones were exclusively located to the core regions of the centromeres. The distribution of the CR elements and CentO repeats revealed by contig assembly is consistent with previous fiber-FISH analysis of centromeres (Cheng et al. 2002).

Of the centromere-specific BAC contigs, four were shown to contain the whole core regions of the centromeres in *japonica* rice (Figs. 2.2, 2.3, 2.4, and 2.5). By BAC-end sequencing and overgo hybridization, three contigs were anchored to centromeres 3 and 8, respectively, but one contig still remains undetermined. Other centromeric BAC contigs were shown to contain only one side of the entire centromere and have major gaps in the core regions of the centromeres. In our centromeric physical contig mapping effort, approximately 1/3 of the CentO-positive clones remain as singletons due to limited number of bands for contig assembly. However, these clones were shown to have large inserts on pulse-field gels. Therefore, some CentO-positive clones might not have sufficient numbers of *Hind*III and *Hae*III sites to produce polymorphism for contig

assembly, and a fingerprinting method using different restriction enzymes and/or larger-insert clones are needed to construct the contigs spanning the core centromeric regions

Recently, the whole regions of rice centromeres 4 and 8 were sequenced (Nagaki et al. 2004; J Wu et al. 2004; Zhang et al. 2004). By the overgo hybridization with chromosome 8-specific probes (RER1 and TGF), BAC contigs of centromere 8 were identified in *japonica* and *indica* rice (Figs. 2.4 and 2.5). The contigs show the similar physical size of centromere 8 in the sequencing map of IRGSP.

Our physical contig map of rice will be valuable resource for gap filling in the physical maps of the IRGSP and the functional study of centromeres. Especially, BIBAC clones could facilitate not only the functional study of centromeres, but also the construction of rice artificial chromosome, since BIBAC can be directly transformed into plants, or the pCLD04541 vector already contains a selection maker *nptII* gene resistant to hygromycin for screening transformed plants.

**CHAPTER III**  
**ONE LARGE INSERT PLANT-TRANSFORMATION-COMPETENT BIBAC**  
**LIBRARY OF THE WILD RICE, *Oryza rufipogon*, AND IDENTIFICATION OF**  
**THE CLONES DERIVED FROM THE REGION OF ITS CHROMOSOME 8**  
**CENTROMERE**

**Overview**

Large-insert DNA libraries are crucial to many aspects of advanced genomics research. Here we report one large-insert plant-transformation-competent BIBAC library for the wild rice, *Oryza Rufipogon* Griff. The library was constructed in the *Bam*HI site of a plant-transformation-competent binary vector pCLD04541. It contains 24,192 clones, has an average insert size of 163 kb, and covers 5 x haploid genomes of *O. rufipogon*. For the comparative study of the centromeres between cultivated rice and its wild progenitor, we screened the library with the two probes (RER1 and TGF) flanking the chromosome 8 centromere of cultivated rice, *O. stativa* ssp. *japonica* and verified the positive clones by Southern analysis with a probe derived from the centromere-specific satellite repeat CentO previously cloned from cultivated rice. A total of 12 positive clones were obtained, of which nine were shown to contain the CentO repeats. These positive clones are useful for developing the physical map of the chromosome 8 centromere of *O. rufipogon* and comparatively studying the centromeres between cultivated rice and its wild progenitor. Since *O. rufipogon* has many traits of

importance to agriculture, our library provides a useful resource for research in comparative genomics of the species, and for study and use of its agronomically important traits for cultivated rice improvement through direct genetic transformation of large-insert BIBACs containing genes of agronomic importance.

## **Introduction**

Rice is the staple food crop for half of the world's population and a model species for genome research of cereal crops due to its small genome (430 Mb/1C, Arumuganathan and Earle 1991), established transformation systems (Hiei et al. 1994; Dong et al. 1996), and the availability of a wealth of genomic and genetic resources (Causse et al. 1994; Yamamoto and Sasaki 1997; Harushinma et al. 1998; <http://rgp.dna.affrc.go.jp>; <http://tigr.org/tdb/tgi.html>).

The common wild rice, *Oryza rufipogon*, is an ancestor of Asian cultivated rice (*Oryza sativa*), including ssp. *japonica* and ssp. *indica*, and thus, is the most important germplasm for cultivated rice genetic improvement (Oka 1988). Since wild rice and Asian cultivated rice share the same AA-genome, transfer of agronomically important genes from the wild rice into cultivated rice can be easily carried out through sexual hybridization (Brar and Khush 2002). Nevertheless, the process of this approach often is several years long and leads to the transfer of agronomically deleterious genes while the targeted genes are transferred into the cultivated species. In comparison, the genetic transformation of the DNA fragment carrying the targeted gene(s) has been proven to

effectively accelerate the process of gene transfer from the wild species to cultivated ones (Cui et al. 2000; He et al. 2003).

*Agrobacterium*-mediated transformation has emerged as the method of choice for genetic transformation in rice (Hiei et al. 1994; Dong et al. 1996). Techniques for large-insert BIBAC or TAC (transformation-competent artificial chromosome) transformation have been established in rice (He et al. 2003) and several other plant species, including tobacco (Hamilton et al. 1996), *Arabidopsis* (Liu et al. 1999) and tomato (Hamilton et al. 1999) and the techniques for large DNA fragment transfer in *Brassica* (Wu et al. 2000) and the model legume, *Lotus japonicus* (Men et al. 2002) are under development. Therefore, plant-transformation-competent BIBAC or TAC libraries have been constructed for many species of agricultural importance, for instance, wheat (Moulet et al. 2000), rice (Chauhan et al. 2001; Tao et al. 2002), soybean (Meksem et al. 2000), tomato (Hamilton et al. 1999), sugarbeet (Fang et al. 2004) and *Brassica* (Wu et al. 2000). BIBAC libraries not only streamline the positional cloning of genes and QTLs of agricultural importance (Hamilton et al. 1996; Liu et al. 1999), but also are essential to simultaneously transfer several closely linked genes, such as disease resistance gene clusters, into plants for genetic improvement.

There have been a number of studies that report genetic or phylogenetic relationships between the wild rice and cultivated rice (Xiong et al. 1999; Gao et al. 2002; F Ren et al. 2003; Cai et al. 2004), but quite a few comparative studies of centromeres have shown that centromeric CR (centromeric retrotransposon) elements and CentO (a 155-bp satellite tandem monomer of *Oryza sativa*) satellites were



conserved in the centromeres of five wild rice species, such as *O. rufipogon* (AA genome), *O. officinalis* (CC genome), *O. minuta* (BBCC genome), *O. alta* (CCDD genome), and *O. australiensis* (EE genome) (Uozu et al. 1997; Hass et al. 2003).

Rice is a diploid,  $2n = 24$ , of which chromosome 8 has the smallest centromere and contains fewer heterochromatin regions than other chromosomes (Cheng et al. 2002). Recently, the structural organization and DNA sequences of chromosome 8 centromere of the cultivated rice, *O. sativa* ssp. *japonica*, have been reported (Nagaki et al. 2004; J Wu et al. 2004). However, it is unknown how centromeres originate, function and evolve. In this study we have constructed a large-insert plant-transformation-competent BIBAC library of the wild rice, *O. rufipogon*, and identified the large-insert BIBAC clones from the centromeric region of chromosome 8 of the wild rice. The library and clones have provided essential resources for molecular characterization of the centromere and studies of centromere origin, function and evolution in the cultivated rice and wild rice and for genetic improvement of cultivated rice using wild rice by genetic engineering.

## **Materials and Methods**

### **Plant materials**

The seeds of *Oryza rufipogon* (Accession Number PI590417) were kindly provided by the U.S. National Plant Germplasm System. Plants were grown in a greenhouse for 4-5

weeks, and young leaves were harvested and immediately stored at -80°C. The stored leaves were used for isolation of high-molecular-weight (HMW) DNA for BAC library construction.

#### BAC library construction

The plant-transformation-competent binary vector, pCLD04541 (Tao and Zhang 1998), was prepared according to Zhang (2000) and Wu et al. (2004c). HMW DNA preparation, enzymatic partial digestion, size selection of partially digested HMW DNA, BAC cloning and library construction were essentially carried out as described by our laboratory (Zhang 2000; Ren et al. 2003b; Wu et al. 2004b and Wu et al. 2004c). Nuclei were extracted from about 50 g leaves of the plants, resuspended in 3 ml of 1% low-melting-point (LMP) agarose, and pipetted into plug-form molds to make plugs at 100 µl/plug. The solidified plugs were transferred into 10 x sample volume of lysis buffer (0.5 M EDTA, pH 9.0, 1% lauryl sarcosine, and 1 mg/ml proteinase K), incubated at 50°C with gentle shaking for 24 h, and then stored at 4°C.

Before partial digestion with *Bam*HI, the plugs were washed three times (1 h each time) in 20 x sample volumes of ice-cold TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH8.0) plus 0.1 mM phenylmethyl sulfonyl fluoride, and three times (1 h each time) in 20 volumes of ice-cold TE. The plugs for partial digestion were cut into small pieces approximately equal in size, equilibrated on ice for 1 h in 1x React 3 buffer (Invitrogen, USA) plus 2 mM spermidine, 1 mM DTT, and 0.2 mg/ml BSA. The

equilibrated plug slices were transferred into the fresh ice-cold reaction mixture containing 1.6 U/tube of *Bam*HI, incubated on ice for 1 h, and then incubated in a 37°C water bath for 8 min. The reactions were stopped by adding 1/10 volume of 0.5 M EDTA (pH 8.0). Partially digested DNA was fractionated by pulse-field electrophoresis on a 1 % agarose gel using the CHEF DRIII (BioRad, USA) in 0.5x TBE buffer (45 mM Trizma base, 45 mM boric acid, 1.0 mM EDTA, pH 8.3) for 20 h. CHEF apparatus was set at 6 V/cm, 12.5°C, 120° angle, 90 s switch time for 18 h, and at 4 V/cm, 12.5°C, 120° angle, 4 s switch time for 2 h. After electrophoresis, DNA fragments ranging from 100 to 300 kb were excised from the gel. The DNA was recovered from the agarose-gel slices by electroelution in dialysis tubing (molecular-weight exclusion limit = 12,000-14,000 Daltons) (Invitrogen, USA) using the CHEF apparatus at 6 V/cm, 12.5°C, 120° angle, and 30 s switch time in 0.5x TBE for 4 hour, followed by reversing the polarity of the current for 60 s. The recovered DNA was subjected to a second size selection on a 1 % agarose gel using the conditions that compress the DNA fragments larger than 100 kb into a thin band (4 V/cm, 5 s switch time, 10 h at 12.5°C in 0.5x TBE). The compressed DNA band was cut out of the gel and recovered from the gel slices by electroelution as above. The recovered DNA solution was dialyzed against 1,000 volumes of ice-cold 0.5x TE twice at 4°C, 1 h each dialysis.

The dialyzed DNA was collected, quantified and then ligated into the *Bam*HI-digested and dephosphorylated pCLD04541 vector at a molar ratio of vector: insert DNA of 8:1 in the presence of 2.0 U T4 DNA ligase (Invitrogen, USA) per 100 µl ligation mixture. The ligation reaction was incubated for 10 h at 16 °C. The ligation

mixture was directly used to transform *E. coli* DH10B competent cells (Invitrogen, USA) by electroporation. One microliter of ligation mixture was mixed with 20  $\mu$ l of the competent cells. The device (Cell Porator and Voltage Booster System, Gibco BRL) for electroporation was set at 350 V, 330  $\mu$ F capacitance, low ohm impedance and fast charge rate with Voltage Booster at 4 k $\Omega$  resistance. Transformed cells were transferred into 1 ml of SOC medium (Sambrook et al. 1989), incubated at 37°C with gentle shaking for 1 h, and then plated on selective LB medium (Sambrook et al. 1989) containing 15  $\mu$ g/ml tetracycline, 0.55 mM IPTG, and 80  $\mu$ g/ml X-gal. The plates were incubated at 37°C for 24 h. White colonies were arrayed into 384-well microtiter plates containing 60  $\mu$ l of LB plus freeze medium (Zhang et al. 1996; Zhang 2000; Ren et al. 2004b; Wu et al. 2004b) in each well, incubated overnight at 37°C, and stored at -80°C.

#### BAC analysis

BAC clones were grown overnight at 37°C in 5 ml LB broth (Sambrook et al. 1989) containing 15  $\mu$ g/ml tetracycline with shaking at 250 rpm. BAC DNA was prepared according to Zhang (2000). To estimate the BAC insert sizes, 5  $\mu$ l of each BAC DNA miniprep was digested with 1.0 U *NotI* (BioLabs, USA) in a 20  $\mu$ l reaction. The digested DNAs were separated by pulse-field gel electrophoresis (PFGE) on a 1% agarose gel. The PFGE condition was 6 V/cm, 120° angle, 12.5°C, a 5 s initial switch time and a 15 s final switch time in 0.5xTBE for 16 h. The gel was stained with ethidium bromide and photographed.

## BIBAC library screening and Southern analysis

High-density colony filters were made from the BIBAC library using the GeneTAC Robotic Workstation (Genomic Solutions, USA). Each BAC clone was double-spotted onto 8 x 12 cm Hybond-N+ membranes (Amersham-Pharmacia Biotech, USA) in a 3 x 3 format. The clones were grown on the filters placed LB agar plates containing 15 µg/ml tetracycline at 37°C for 16 h, and the filters were processed according to Zhang (2000), followed by baking at 80°C for 2 h.

For library screening, two pairs of overgo (overlapping oligonucleotide) primers were designed from the rice BAC clone AY360388 (Nagaki et al. 2004) spanning the entire centromere of cultivated rice chromosome 8, using a web-based primer3 program ([http://cbr-rbc.nrc-cnrc.gc.ca/cgi-bin/primer3\\_www.cgi](http://cbr-rbc.nrc-cnrc.gc.ca/cgi-bin/primer3_www.cgi)): RER1 (F) 5'-GTGAGTTCT-AGGAGAGTAGCTTG-3' and (R) 5'-GAAATAATATCCTGCCAAGCTAC-3'; TGF (F) 5'-GAACCGGCGATAAATACCTTGC-3' and (R) 5'-AAATCGAGAAGCCTGC-AAGGTA-3'. Overgo labeling was prepared according to a protocol developed in the California Genome Research Laboratory (CGRL) (<http://informa.bio.caltech.edu/protocols/overgo.html>), using radioactive [<sup>32</sup>P]-dCTP. Each set of 16 filters was pre-hybridized in 150 ml hybridization buffer (5x SSC, 5x Denhardt's solution, 100 µg/ml salmon sperm DNA, 0.5% SDS) at 65°C for 1 h. Hybridization was performed overnight at 65°C in 30 ml fresh hybridization buffer plus the denatured radioactive

probes (RER1 or TGF). The filters were washed twice at 65°C in 2x SSC, 0.1% SDS for 20 min.

For Southern blot hybridization, probe DNA was amplified by PCR using the primers designed from tandem repeat RCS2 (AF058902) (Dong et al. 1998). The PCR product was electrophoresed on a 1% agarose gel, excised, and purified using the QuiaQuick gel extraction kit (Qiagen Inc., USA). Southern blot hybridization was performed as described (Sambrook et al. 1989). BAC DNAs were prepared by an alkaline lysis method (Sambrook et al. 1989), digested with *NotI* (BioLabs, USA), run out on an agarose gel, and blotted onto Hybond-N+ membrane (Amersham-Pharmacia Biotech, USA). Prehybridization and hybridization were done at 65°C in hybridization buffer. After hybridization, the membrane was washed in 0.1x SSC, 0.1% SDS at 65°C, 20 min each wash, and then was exposed to x-ray films.

## **Results**

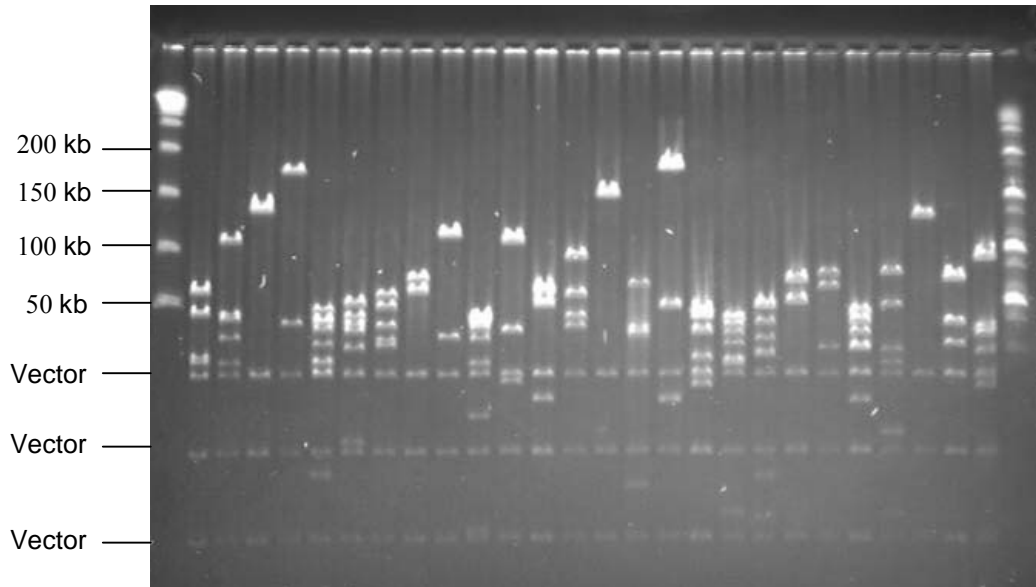
### BIBAC library construction and characterization

A BIBAC library of *O. rufipogon* was constructed in the *Bam*HI site of a plant-transformation-competent binary vector (pCLD04541). The library consists of 24,192 clones arrayed in 63 384-well microplates. To estimate the insert size of the library, 59 clones were randomly selected and grown in LB medium containing 15 µg/ml tetracycline. DNAs were isolated, digested with *NotI*, and separated on pulse-field gels.

The gels were stained with ethidium bromide and photographed (*e.g.*, Fig. 3.1). The insert size of each clone was estimated using the lambda DNA ladder and the cloning vector fragment(s) as the molecular-weight standards. The library has an average insert size of 163 kb with a range of 118 - 245 kb (Fig. 3.2). Less than 0.1 % clones had no inserts. Based on the haploid genome size of 760 Mb (Bennett et al. 1982) of *O. rufipogon*, the coverage of the library is about 5 equivalents of the haploid genome, providing a greater than 99% probability of obtaining a single-copy clone from the library. Since the binary vector pCLD04541 has two other *NotI* sites besides the double sites flanking the multiple cloning site, the vector was digested into three DNA fragments. Most BAC clones had several insert fragments (*e.g.*, Fig. 3.1), which is a typical pattern observed in the large-insert BACs of monocotyledon plants that contain one or more *NotI* sites in a 100-kb BAC insert (Tao et al. 2002), in contrast to those of dicotyledonous plants, such as *Arabidopsis* (Chang et al. 2003) and soybean (Wu et al. 2004b), in whose insert of 100 kb there is often no *Not I* site.

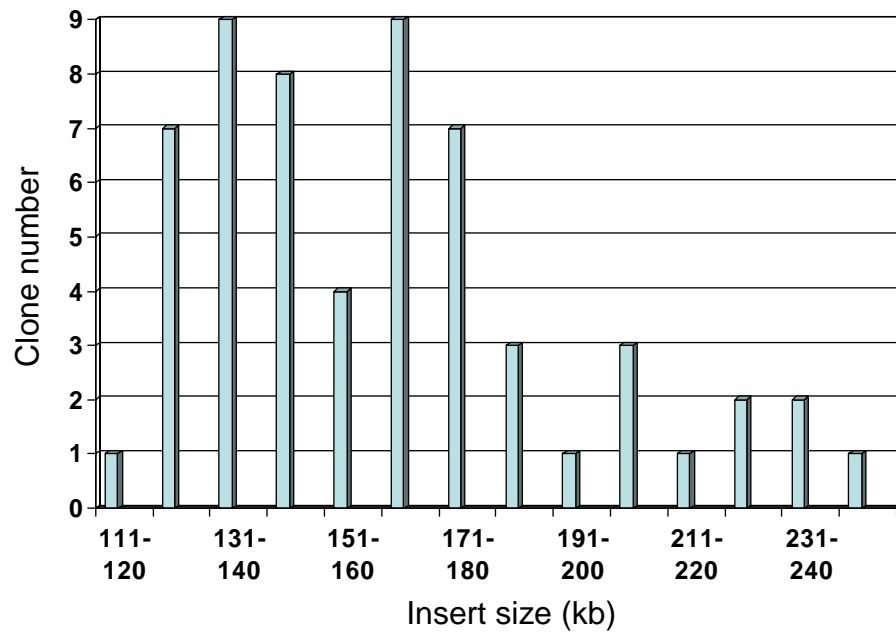
Library screening with rice centromere 8-specific DNA sequences and Southern analysis with centromeric CentO satellite repeat sequence of rice

Rice chromosome 8 has the smallest centromere among its 12 chromosomes. To take advantage of its relatively smaller amount of repeated elements and availability of genomic sequences in the GenBank, chromosome 8 centromere was selected for the comparative study of centromeres in rice and its progenitor, *O. rufipogon*. To isolate the



**Fig. 3.1** PFGE analysis of randomly picked BAC clones of the wild rice *Bam*HI large insert library. The two outside lanes are molecular size markers (lambda concatemers from Sigma Chemical, St. Lois, Mo)





**Fig. 3.2** Insert size distribution of the clones in the wild rice *Bam*HI library. BIBAC inserts were released after *Not*I digestion and PFGE separation, and estimated for sizes by comparison with lambda concatemer size markers.

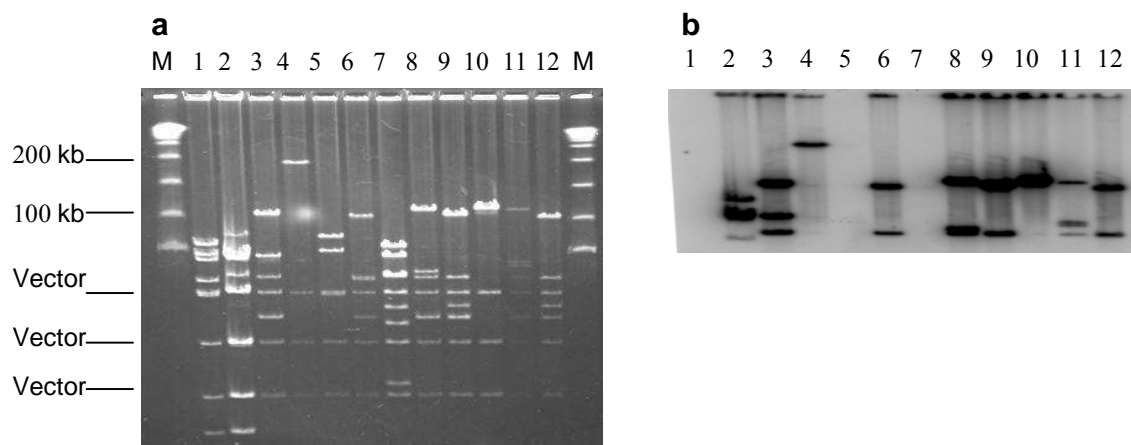
BIBACs derived from the chromosome 8 centromere of *O. rufipogon*, two pairs of Overgos (RER1 and TGF) were designed from the unique sequences flanking the core centromere consisting of the CentO satellite tandem repeats (Nagaki et al. 2004; J Wu et al. 2004). Using the centromere 8-specific probes (RER1 and TGF) prepared by overgo labeling reactions, respectively, a total of 12 clones were obtained (Table 3.1). To further verify the positive clones and determine which clones contain CentO repeats, DNAs were isolated from the positive clones, digested with *NotI*, Southern blotted, and hybridized with probe RCS2 containing the CentO repeats of *indica* rice. Nine of the 12 positive clones were shown to contain the CentO satellite repeats (Fig. 3.3 and Table 3.1).

## **Discussion**

We constructed a plant-transformation-competent BIBAC library for the wild rice, *O. rufipogon*. The library was estimated to have an average insert size of 163 kb and to be equivalent to 5-fold haploid genomes of the wild rice, providing a probability greater than 99% of recovering any single-copy DNA sequence from the library. The large insert size of this library will be helpful for constructing a physical map covering the entire chromosome 8 centromere and facilitate genomic research in the wild rice. Furthermore, since the library was cloned into an *Agrobacterium*-mediated, plant-transformation-competent binary vector pCLD04541 (Jones et al. 1992; Tao and Zhang 1998), it can be directly transformed into plants via *Agrobacterium*. This will help the

**Table 3.1** BAC clones positive to centromere 8-specific probe RER1 and TGF, and CentO satellite repeats.

Sample	BAC clone	Insert size	Probe	CentO bands (kb)
1	B05F22	169	RER1	Negative
2	B43P09	76	RER1	40, 35
3	B15E04	172	TGF	105, 40, 20
4	B14K02	145	TGF	145
5	B28E02	120	TGF	Negative
6	B41C03	82	TGF	55, 20
7	B46J11	138	TGF	Negative
8	B53N06	147	TGF	115, 25, 20
9	B55I08	140	TGF	105,20
10	B56M10	115	TGF	115
11	B58J19	210	TGF	115, 45, 30, 20
12	B58N22	140	TGF	105, 20



**Fig. 3.3** Southern analysis of 12 positive clones obtained by library hybridization using the centromere 8-specific overgo probes (RER1 and TGF). The centromere 8-specific Agarose gel of the positive BACs to centromere 8-specific probes (RER1: lanes 1-2, and TGF: lane 3-12) digested with *NotI* (**a**). Southern blot hybridization of the BACs with CentO probe (**b**). M indicates lambda molecular size marker and lanes 1-12 show the selected positive BAC clones.

utilization of the library in functional analysis of centromeres, as well as the genetic improvement of rice by introducing agronomically valuable genes of the wild rice.

The wild rice *O. rufipogon* is the progenitor of the cultivated Asian rice (Oka 1988). The genome of the wild rice (760 Mb/1C, Bennett et al. 1982) is 1.8-fold larger than that of the cultivated rice (430 Mb/1C, Arumuganathan and Earle 1991). The wild rice has the same genome (AA) as Asian cultivated rice. There are abundant genetic differences between cultivated and wild progenitors. The differences include morphological characters (Harlan 1975), allozymes (Kahler and Allard 1981; Second 1982), and DNA restriction fragments and sequences (Khairallah et al. 1992; Saghai Maroof et al. 1995; Liu et al. 1996). Evolution from wild to cultivated rice led to “domestication syndrome” (Harlan 1975) common to many cereals. Studies of QTLs showed that most of the genetic factors (qualitative or QTLs) controlling the domestication-related traits are concentrated in a few chromosomal blocks of cultivated rice. The domestication of rice is associated with linked genetic factors that are not linked in the wild rice (Xiong et al. 1999; Cai and Morishima 2003).

Unlike other regions of genomes, there is little known about organization, function and evolution in plant centromeres. To study structure and evolution of wild and cultivated rice centromeres, we isolated 12 BAC clones potentially spanning the chromosome 8 centromere of the wild rice. Nine of the clones have been proven to contain rice centromere-specific satellite repeats CentO. These clones will help the construction of a physical map of the wild rice chromosome 8 centromere and determine the genomic sequences of the centromere.

**CHAPTER IV**  
**SEQUENCE ANALYSIS OF THE CHROMOSOME 8 CENTROMERE OF THE**  
**WILD RICE, *Oryza rufipogon* GRIFF.**

**Overview**

A complete sequence of a chromosome centromere in the closely related species is necessary for better understanding of centromere organization, origin, function and evolution. Using a capillary electrophoresis-based fingerprinting method, we constructed a BIBAC-based contig map for centromere 8 of the wild rice, *Oryza rufipogon*. Approximately 95% genomic sequence of the long arm side of the wild rice centromere 8 was obtained through shotgun sequencing of a BIBAC clone (B43P09) spanning the major portion of the centromere. Comparative sequence analysis of the clone showed that the centromere has a gene content largely similar to that of the cultivated rice, *O. sativa* ssp. *japonica*, centromere 8, including the putative *RERI* (endoplasmic reticulum retrieval protein) gene. However, the pericentromeric regions of the centromere have undergone large-scale rearrangements, such as deletions and additions. The core centromeric regions consisting of the 155-bp satellite repeats have evolved rapidly at the nucleotide level since the cultivated rice evolved from the wild rice. We confirmed that although the conserved regions of the 155-bp satellites underwent dramatic changes, the CentOr of the wild rice and the CentO of the cultivated

rice maintain conserved nucleotides at the 5' and 3' ends. These results provide novel insights into the evolution and function of rice centromeres.

## **Introduction**

Centromeres are essential for faithful segregation and transmission of chromosomes in both mitotic and meiotic cell division. In higher eukaryotes, a centromere consists of two major regions, pericentromeric and core centromeric, but the exact boundaries of each region remain largely undefined. The core centromere comprises tandem repeated satellite sequences, the amount of which varies from several hundred kilobases to megabases. Some regions of the core centromere are thought to be responsible for the function of the centromere (Schueler et al. 2001). The pericentromeric regions flanking the core centromere contain moderate repetitive elements, including transposons, retrotransposons, and pseudogenes (Copernhaver et al. 1999; Cheng et al. 2002; J Wu et al. 2004; Zhang et al. 2004).

In cereal plants, centromere-specific repetitive sequences, CR (centromeric retrotransposon) elements derived from a type3/gypsy-class retrotransposon family, are widely dispersed in the pericentromeric regions (Dong et al. 1998; Nonomura and Kurata 1999; Cheng et al. 2002). The chromatin immunoprecipitation assay with centromeric histone H3 proteins (CENH3) demonstrated some of CR elements and satellites interact with CENH3 proteins, suggesting that some CR elements could participate in the functions of centromeres (Zhong et al. 2002).

In general, full length of centromeres in higher eukaryotic organisms is at the megabase level. The long range of repeated sequences, such as centromeric satellites, in the centromeres makes it difficult to clone and sequence the regions. The human and *Arabidopsis* genomes have been sequenced by the clone-by-clone approach (*Arabidopsis* Genome Initiative 2000; International Human Genome Sequencing Consortium 2001). However, due to the long stretches of satellite repeats and other repetitive sequences, big gaps remain to be sequenced in the centromeric regions of the physical maps for each chromosome of the species.

Unlike *Arabidopsis* and human, rice has provided an important model for organizational study of centromeres because the centromeres of several of its chromosomes, such as 3, 4 5, 8 and 12, contain a limited amount of satellite DNA and other repetitive sequences, as detected by the fluorescence intensities of the fluorescence *in situ* hybridization signals (Cheng et al. 2002). Recently, the entire centromeres of the cultivated rice ssp. *japonica* chromosomes 4 and 8 have been sequenced and characterized (Nagaki et al. 2004; J Wu et al 2004; Zhang et. al 2004). In comparison, centromere 8 contains a similar copy number of 155-bp CentO satellite monomers to centromere 4 in the core region, but has more genes and less amounts of CR-related repeated sequences in the pericentromeric regions. Centromere 8 consists of three large clusters of CentO satellites whereas centromere 4 consists of 18 small clusters in the core centromeres (J Wu et al. 2004; Zhang et al. 2004).

Centromeres have been recognized as rapidly evolving regions of chromosomes through large- and small-scale nucleotide sequence changes. Although the functions of



centromeres are highly conserved, there is only a limited sequence similarity among the centromeres of the distantly-related species (Hall et al. 2004). In *Arabidopsis*, the insertions of some mitochondrial sequences and 5S rDNA into centromeres (Stupar et al. 2001; Franz et al. 1998) and inversions involving centromere 4 (*Arabidopsis* Sequencing Consortium 2000) have been reported. The rapid changes of satellite repeats at the nucleotide level have also been observed in several plant species (Cheng et al. 2002; Talbert et al. 2002; Heslop-Harrison et al. 2003; Hall et al. 2003). However, it is little known about the large-scale rearrangements of the centromeric regions between related plant species.

The common wild rice, *Oryza rufipogon* Griff., is a progenitor of Asian cultivated rice, *Oryza sativa* L. (Oka 1988), including ssp. *indica* and ssp. *japonica*. The genome of the wild rice (760 Mb/1C, Bennett et al. 1982) is 1.8-fold larger in size than those of the cultivated rice (430 Mb/1C, Arumuganathan and Earle 1991). Although the wild rice has the same genome (AA) as the cultivated rice, genetic differences between them has been reported (Harlan 1975; Kahler and Allard 1981; Second 1982; Khairallah et al. 1992; Saghai Maroof et al. 1995; Liu et al. 1996). Especially, genetic studies of QTLs showed that the processes of natural selection and domestication from the wild rice to the cultivated rice have led to clustering of genetic factors controlling the domestication-related traits in a few chromosomal blocks of the cultivated rice (Xiong et al. 1999; Cai and Morishima 2002).

To date, limited comparisons of centromere DNA from different species have been revealed that some DNA elements are conserved. However, those studies were

insufficient to identify the conserved functional elements and elucidate evolutionary significance of structural changes in plant centromeres. For better understanding of the function and evolution of the centromeric elements in closely-related species, we constructed a physical contig map of wild rice centromere 8 spanning the whole region of the centromere, and have sequenced the long arm side of the centromere 8 in the wild rice. Comparative sequence analysis between the cultivated *japonica* rice and the wild rice indicated that large-scale rearrangements in the pericentromeric regions and rapid small-scale evolution of 155-bp satellites in the core region have been occurred since they have split.

## **Materials and Methods**

### BAC fingerprinting by capillary electrophoresis

The DNA of BACs derived from the chromosome 8 centromere of the wild rice was isolated using Qiaprep Spin Miniprep Kit (Qiagen Inc., USA). BAC DNA fingerprinting was carried out in wells of a 96-well PCR plate according to Xu et al. (2002, 2004). Approximately 200 – 500 ng of the BAC DNA was digested with 2 units of each *HindIII*, *XbaI*, and *XhoI* in the reaction buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM magnesium acetate, 1 mM DDT, 100 ng/μl BSA) at 37°C for 2 h. The reaction was then terminated at 65°C for 15 min. Fifteen microliters of a mixture containing 2.0 μl of SNaPshot Multiplex Ready Reaction Mix containing ddATP-R6G, ddGTP-R110,

ddCTP-TAMRA, ddT(U)TP-ROX and Taq polymerase FS (Applied Biosystems, USA), 2.0 units of *Bam*HI, 2.5 units of *Hae*III, 25 mM Tris-acetate (pH 7.8), 100 mM potassium acetate, 10 mM magnesium acetate, and 1 mM DTT was added. The reaction was continued to incubate at 37°C for 2 h, followed by incubating at 65°C for 1.0 h.

The DNA sample was precipitated using 100% ethanol, and then dissolved in a mixture containing 9.8 µl of Hi-Di formamide (Applied Biosystems, USA) and 0.2 µl of ABI internal size marker LIZ-500 ranging from 35 to 500 bp (Applied Biosystems, USA). The DNA was denatured at 95°C for 8 min and then put on ice for 8 min. Capillary electrophoresis was carried out on the ABI 3100 Genetic Analyzer with 36-cm capillary using the ABI default GeneScan module (Applied Biosystems, USA). The fingerprint data was collected by ABI GeneScan version 3.70 and ABI data collection version 1.01 software (Applied Biosystems, USA).

#### FPC contig assembly

The band data was transferred into a SUN computer workstation from the fingerprinting raw data in the ABI 3100 Genetic Analyzer and edited by use of the MultiColor editor software (<http://hbz.tamu.edu>). The edited data were converted into the FPC format data by the SizeToFPC software (<http://hbz.tamu.edu>). To construct physical map contigs, FingerPrinted Contig (FPC) program v.6.0 (Soderlund et al. 1997; Soderlund et al. 2000) was used. The automatic contigs were assembled at tolerance = 2, cutoff =  $10^{-15}$  to  $10^{-25}$ , DiffBury 0.1, and MinBands 8.

## Nucleotide sequencing and sequence assembly

The shotgun sequencing approach was used to sequence the BACs spanning the centromeric region of the wild rice chromosome 8. The BACs were selected from the contig map of the centromere and DNA was isolated using Qiaprep spin miniprep kit (Qiagen Inc., USA). BAC DNA was sheared into fragments with an average of 1.5 kb by the HydroShear Machine (GenomicSolutions, USA). After being polished at the ends with Klenow and T4 DNA polymerases, the sheared DNA was ligated into the pBluescript vector and transformed into *E. coli* DH10B (Invitrogen, USA) by electroporation. The shotgun subclones were sequenced from the forward end by the dideoxy chain termination method using the BigDye Terminator Cycle Sequencing v1.0 and v2.0 ready reaction kits (Applied Biosystems, USA). The sequencing reactions were analyzed on the ABI 3100 Genetic Analyzer (Applied Biosystems, USA). The sequences were trimmed, refined and assembled by using the Sequencher V4.1 (Gene Codes Corporation, USA). The primary assembled results were refined by carefully manual checking to correct the misalignments caused by repeat sequences.

## Sequence analysis

The sequences were searched against the GenBank database using the BLASTN and BLASTX homology search software and compared using the BLAST 2 software.

Sequence alignments were performed and refined manually using the Sequencher V4.1 and GeneDoc software (<http://www.psc.edu/biomed/genedoc>).

Sequence comparative analysis of the centromeric regions between the wild rice and cultivated rice ssp. *japonica* was performed using zPicture and eShadow web-based software (Ovcharenko et al. 2004a; Ovcharenko et al. 2004b). Wild/*japonica* rice sequence alignments were generated and visualized by the zPicture program (<http://zpicture.dcode.org>), using standard parameters ( $\geq 100$  bp,  $\geq 70\%$ ). The conserved coding exons in the pairwise alignments were detected using the HMMI method of eShadow software (<http://eshadow.dcode.org/>).

## **Results**

### Physical map of the wild rice chromosome 8 centromere

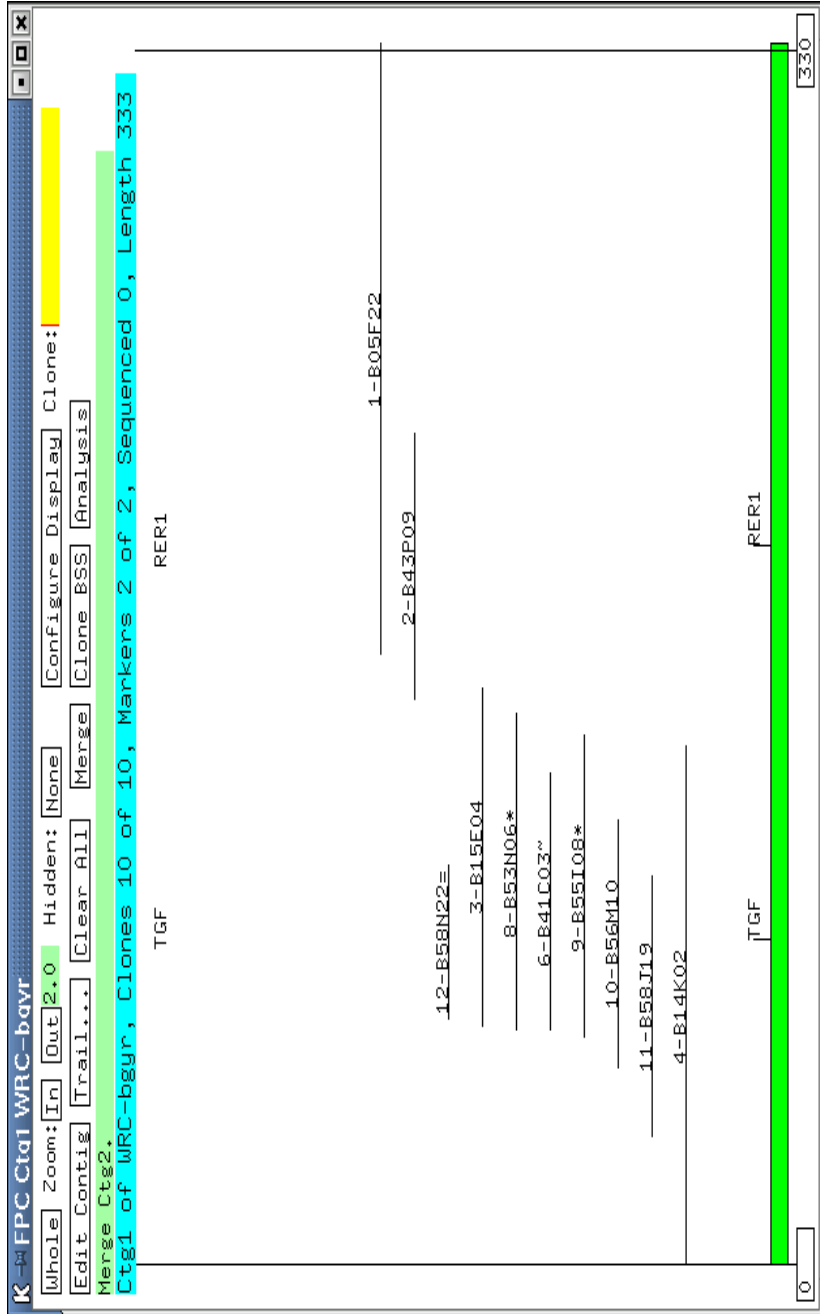
Using the novel capillary sequencing-based fingerprinting method developed by Xu et al. (2002, 2004), we fingerprinted the 12 BAC clones positive to the cultivated rice ssp. *japonica* centromere 8-specific probes, RER1 and TGF, with five restriction enzymes (*HindIII*, *XbaI*, *XhoI*, *BamHI*, and *HaeIII*) to generate enough bands for contig assembly. The fingerprint raw data were edited by the MultiColor editor software (<http://hbz.tamu.edu>).

Two automated contigs were assembled using the FPC program v.6.0 (Soderlund et al. 1997a; 2000b) and then merged into one contig because the contig end clones

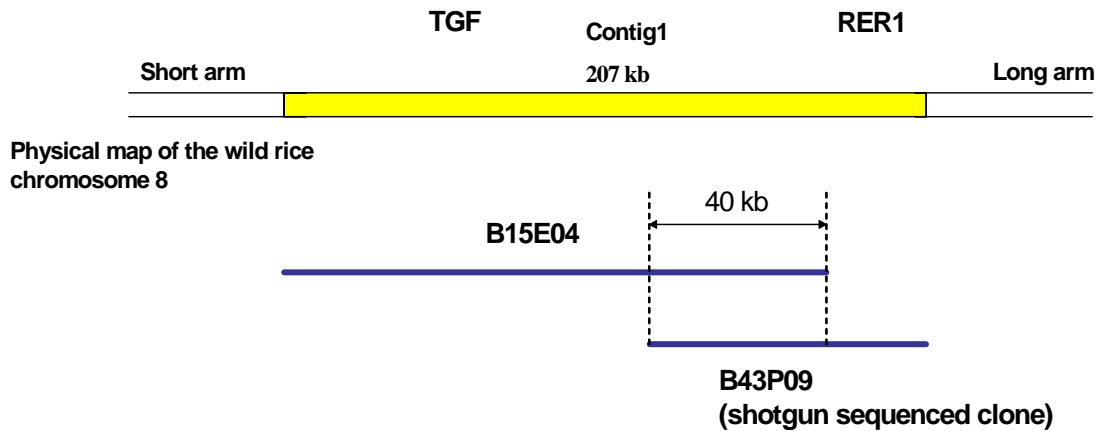
(B43P09 and B15E04) share some fingerprint bands and a 40-kb *NotI* fragment containing CentO repeats (Fig. 3.3 and Table 3.1). The contig physical map consisted of 10 of the 12 BAC clones and spans approximately 364 kb in physical length (Fig. 4.1). By Southern hybridization of the clones with the cultivated rice centromere 8-specific probes that flank the core region of the centromere and that of the key centromere component CentO satellites, we found that the region covered by two clones, B43P09 and B15E04, contained the core region of the centromere, based on comparison to the *japonica* centromere 8 (Fig. 4.2).

#### Sequencing of the wild rice centromere 8

Two BAC clones, B43P04 and B15E04, of the centromere 8 contig map of the wild rice were selected for shotgun sequencing. The region covered by the two clones represents 207 kb in physical length and was proven to contain the whole region of the centromere 8 (Fig 4.2). Shotgun libraries were constructed from the two clones, respectively and analysis of random clones from the libraries indicated that the libraries had average insert sizes of 1.5 kb, suggesting that they are suited for shotgun sequencing the centromeric region. Due to the smaller size of its insert, B43P09 was first sequenced and a 7-fold coverage of sequences was obtained. The sequences were assembled by use of the Sequencher V4.1, and the assembly result was refined by carefully manual checking to correct the misalignments caused by repeats.



**Fig. 4.1** The BAC contig map of the wild rice centromere 8. TGF and RER1 are centromere 8-specific overgo probes designed from the sequences flanking the centromere 8 of *japonica* rice.



**Fig. 4.2** Minimal overlapping BAC tile path of the centromeric region of the wild rice chromosome 8. Contig 1, which covers the entire region of the centromere 8 (yellow), is indicated in blue lines. The BIBAC clones, B43P09 and B15E04, were shown to share a 40-kb fragment containing CentOr satellites and span the entire centromere. TGF and RER1 are cultivated rice centromere 8-specific overgo probes flanking its centromere 8 (Nagaki et al. 2004).



The sequence map of B43P09 consists of 12 contigs and 42 singletons, spanning approximately 78 kb in physical length. We estimated the sequenced region to represent at least 95% for the 75-kb insert of the clone (Table 4.1). By sequencing the subclones constructed with *NotI* and a 10-kb shotgun library of B43P09, we are filling the gaps of the sequence map.

#### Comparative sequence analysis with the cultivated *japonica* rice centromere 8

On the basis of the sequence analysis with BLASTN and BLASTX programs in the GenBank, 91% of the 78-kb centromeric sequences from B43P09 were repetitive, most of which had high similarities to the type3/gypsy-class retrotransposon family. Only one putative gene RER1 was found in the sequence map. In the comparison with the region of centromere 8 in *japonica* rice, 69% of the retroelements and other sequences were shown to be highly conserved, but some sites were observed to have nucleotide substitutions, and the insertions and deletions of several nucleotides in *japonica* rice.

For the further analysis based on the region-to-region comparison in the centromere 8 between the wild and *japonica* rice, the zPicture program (<http://zpicture.dcode.org>) was used to perform alignments and visualization of the conserved sequences (Ovcharenko et al. 2004b), and the eShadow (<http://eshadow.dcode.org>) programs were used to detect conserved coding regions (Ovcharenko et al. 2004a). The compared region of *japonica* rice contains one 6928-bp putative endoplasmic reticulum retrieval protein gene (*RER1*) (J Wu et al. 2004). In the

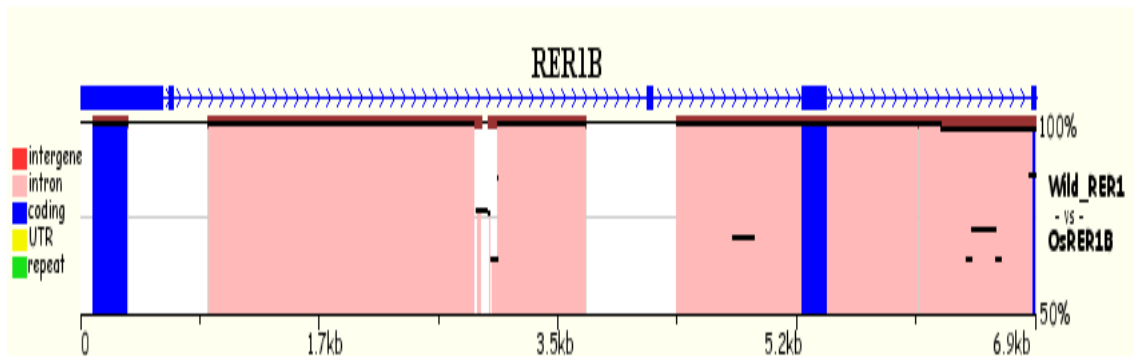
**Table 4.1** Sequence contigs constituting the sequence map of B43P09 of the wild rice and comparison with the centromere 8 sequence of *japonica* rice.

Sequence Contig	Size (kb)	Sequences not found in the <i>japonica</i> rice centromere 8 (kb)
1	4.1	1.4
2	13	2.3
3	3.7	-
4	3.5	0.8
5	5.4	1.6
6	4.7	2.5
7	2.3	-
8	1.1	0.5
9	0.8	-
10	14.7	2.9
11	1.9	-
12	11.8	11.8
Total	67	23.8

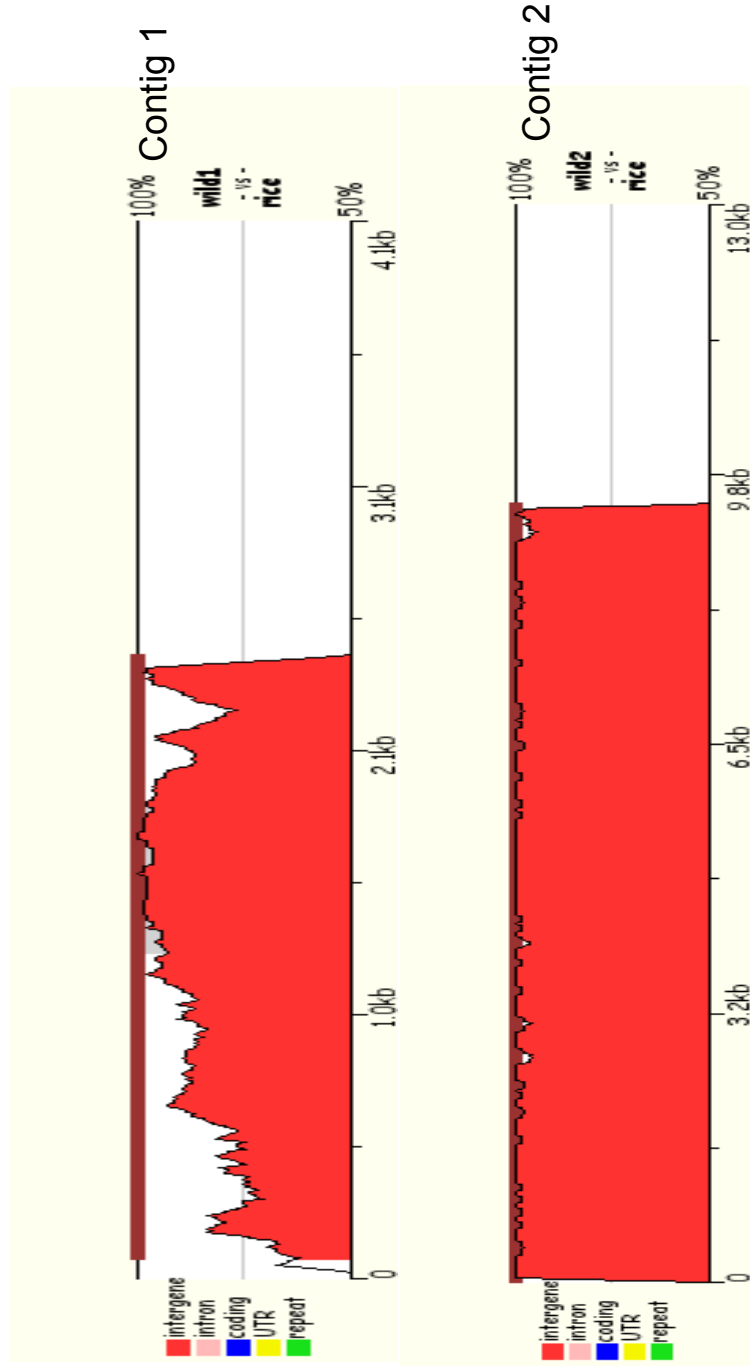
zPicture analysis, the sequences of two introns and two partial sequences of the intron 2 in RER1 of *japonica* rice were shown to be highly conserved except the deletion of 15 nucleotides and addition of 7 nucleotides. However, the two exons of the wild rice *RER1* are identical to those of *japonica* rice, except for several base substitutions (Fig. 4.3). The zPicture analysis of the remaining sequences revealed that a total 46.5-kb sequence of 6 contigs contains large deletions of several kilobases or addition of other centromeric sequences in the centromere 8 of *japonica* rice (Fig 4.4 and Table 4.1). Furthermore, the 11.8-kb sequence of the wild rice contig 12 was not found in the comparable region of centromere 8 in *japonica* rice, but was conserved in other regions of centromere 8 and other chromosomes. In the comparison of the 11.8-kb contig 12 with the whole BAC sequence (192 kb) of *japonica* AP006480, quite a few small similar fragments were found in the wild rice contig (Fig. 4.5). Most conserved elements were proven to be partial coding regions of the polyproteins or gag proteins (Fig. 4.6) by the HMMI prediction method of eShowdown program and the BLASTX search in the GeneBank. These results provided the evidence that the processes of the natural selection and the domestication might lead to the significant structural rearrangements of the proximal centromeric regions from the long arm of the rice chromosome 8.

#### Comparison of the wild rice CentOr and the cultivated rice CentO satellites

The sequence map of B43P09 in the wild rice contains more than 29 of monomers of the wild rice 155-bp satellite repeats (CentOr), with a total of 6.02 kb in length. We were



**Fig. 4.3** zPicture analysis of rice *RER1B* gene. Three sequence contigs of the wild rice were aligned with the *japonica* rice *RER1B* gene. Wild/*japonica* alignments were generated and visualized by the zPicture program using the standard parameters ( $\geq 100$  bp,  $\geq 70\%$ ). The conserved elements corresponding to exons (blue) and introns (pink) are indicated. The y-axis corresponds to percent identity of the sequences and x-axis corresponds to size in base pairs. The alignment has three gaps in the zPicture plot. The locations of the compared regions of wild rice (contig 3 and 11) are depicted as violet bars. The addition and deletions of nucleotides are shown as black bars.



**Fig. 4.4** zPicture analysis of the sequence contigs of the wild rice. Wild/*japonica* rice alignments were generated and visualized by the zPicture program using standard parameters ( $\geq 100$  bp,  $\geq 70\%$ ). The conserved elements corresponding to intergenic elements are indicated in red. The y-axis indicates the percent identity of nucleotide sequences between the wild rice and *japonica* rice, and the x-axis indicates the size of the compared region in base pairs.

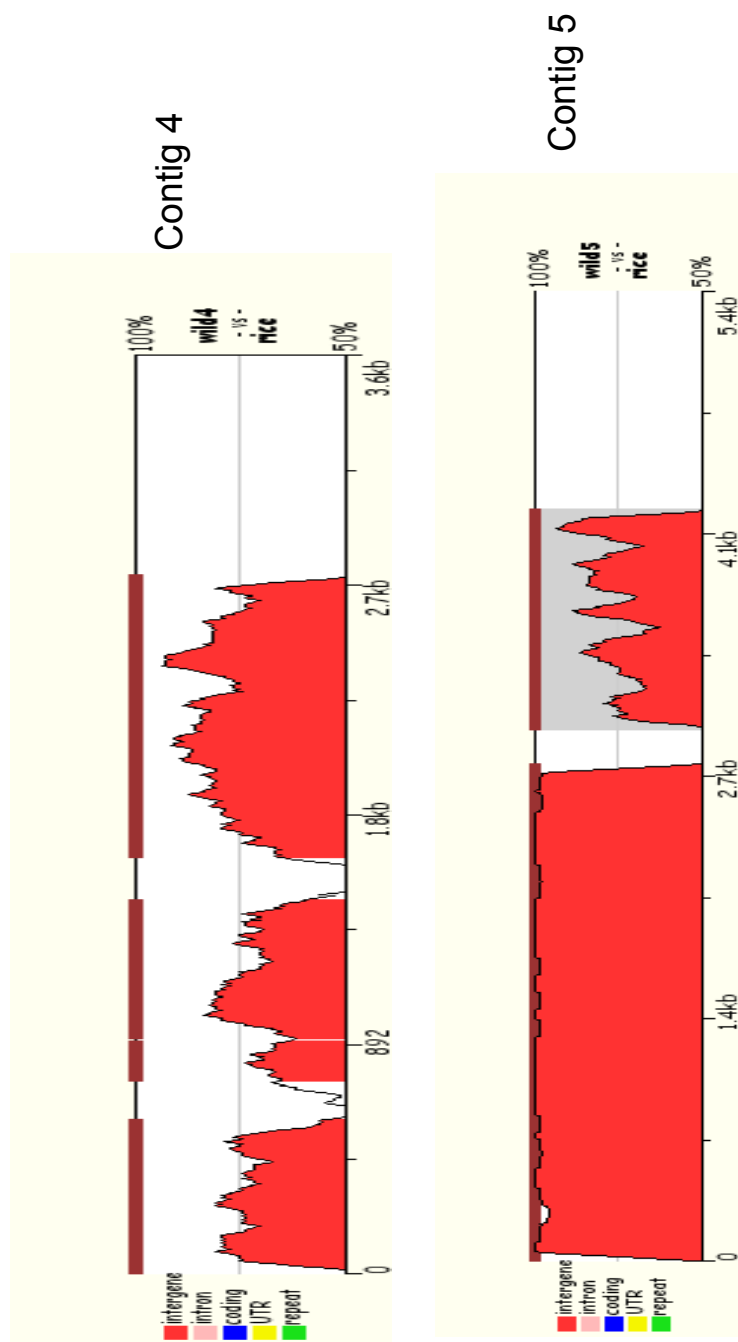
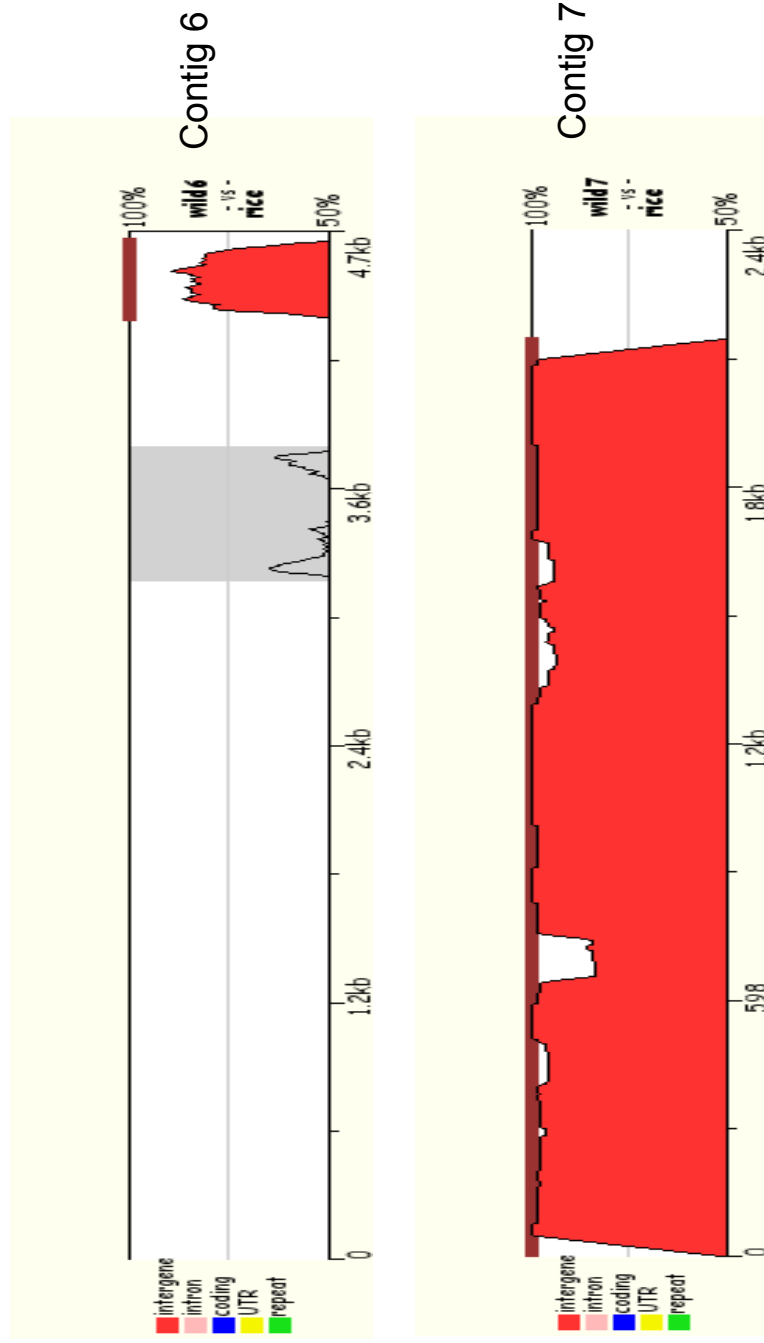


Fig. 4.4 (continued)



**Fig. 4.4** (continued)

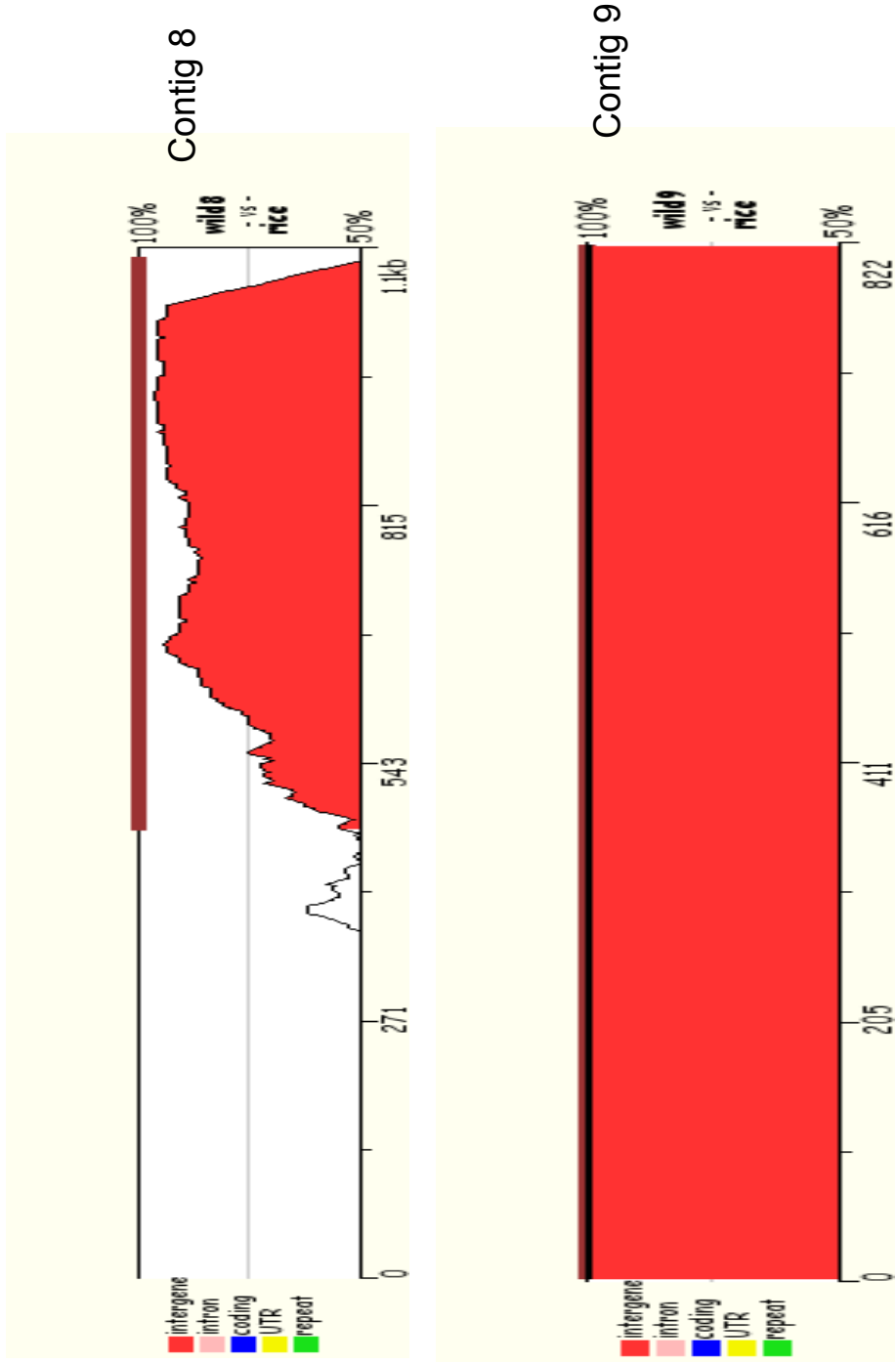


Fig. 4.4 (continued)



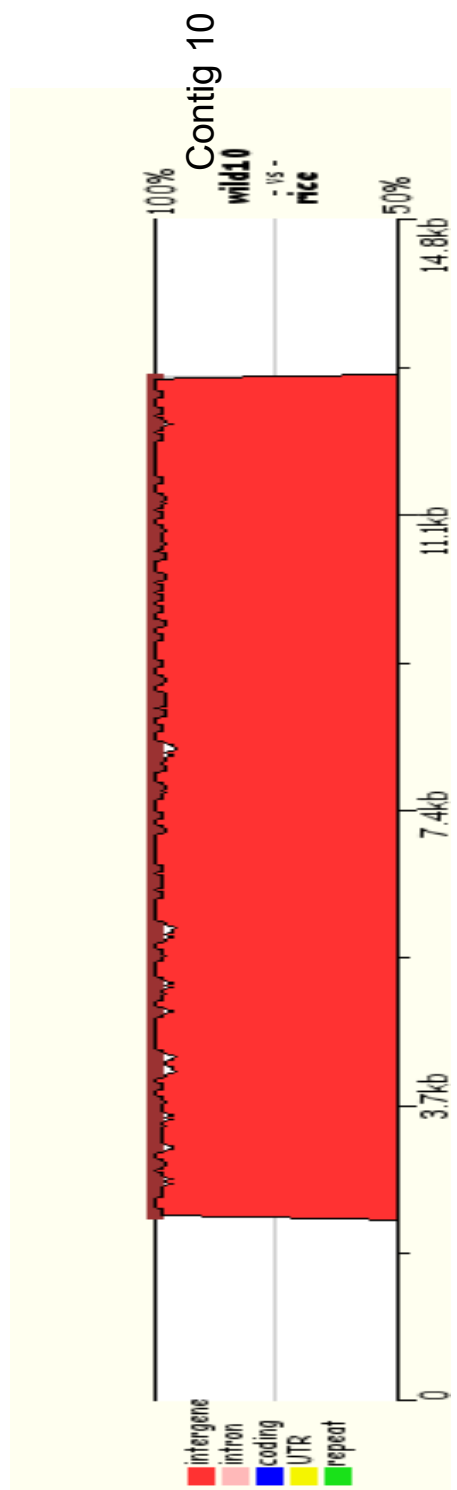
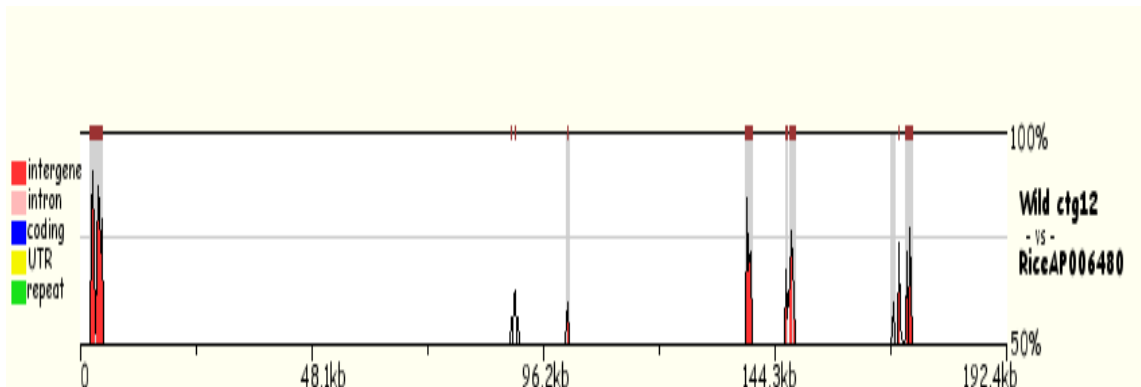
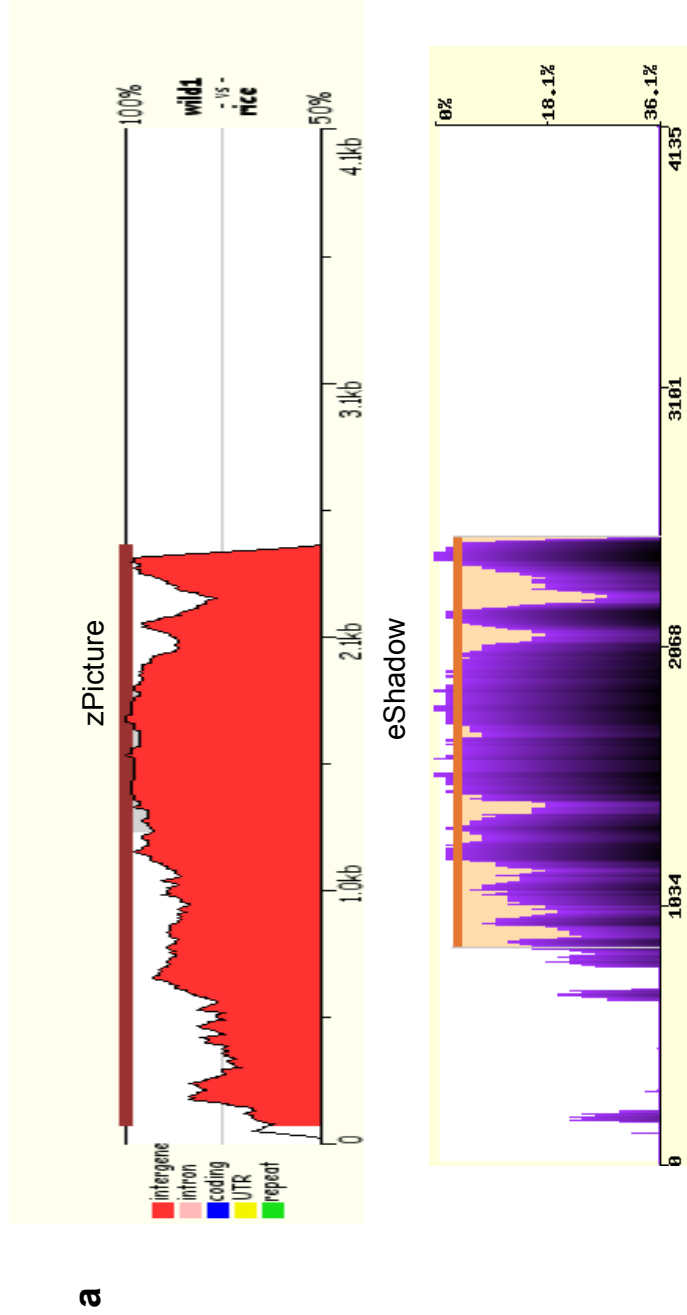


Fig. 4.4 (continued)



**Fig. 4.5** zPicture analysis of the sequence contig 12 of the wild rice. The wild rice contig 12 sequence (11.8 kb) was aligned to the whole BAC sequence of *japonica* rice AP006480, and the conservation ( $\geq 100$  bp,  $\geq 70\%$ ) between the sequences was analyzed using the zPicture program. The contig 12 sequence has only small fragments observed in the compared *japonica* BAC clone AP006480. The conserved sequence corresponding to intergenic elements are indicated in red. The y-axis indicates the percent identity of sequences between the wild rice and *japonica* rice, and the x-axis indicates the size of the compared region in base pairs.



**Fig 4.6** zPicture and eShadow analyses of the centromeric regions between the wild rice and cultivated *japonica* rice. The zPicture conservation plot between the wild rice and *japonica* rice was compared with the eShadow conservation plot between the two species. For these analyses, sequence contigs 1, 2 and 6 of the wild rice were compared with the corresponding regions of *japonica* BAC clone AP006480. The Wild/*japonica* rice alignments were generated and visualized by the zPicture program using standard parameters ( $\geq 100$  bp,  $\geq 70\%$ ) and the conserved elements corresponding to intergenic elements are indicated in red. Using the HMMI method, eShadow tool identified predicting exons in the wild/*japonica* pairwise alignments. The HMMI exon predictions are in beige, and the locations of known exons are depicted as red bars in eShadow plots. Percent identity: y-axis and size in bp: x-axis in zPicture plots. Percent variation: y-axis and size in bp: x-axis in eShadow plots.

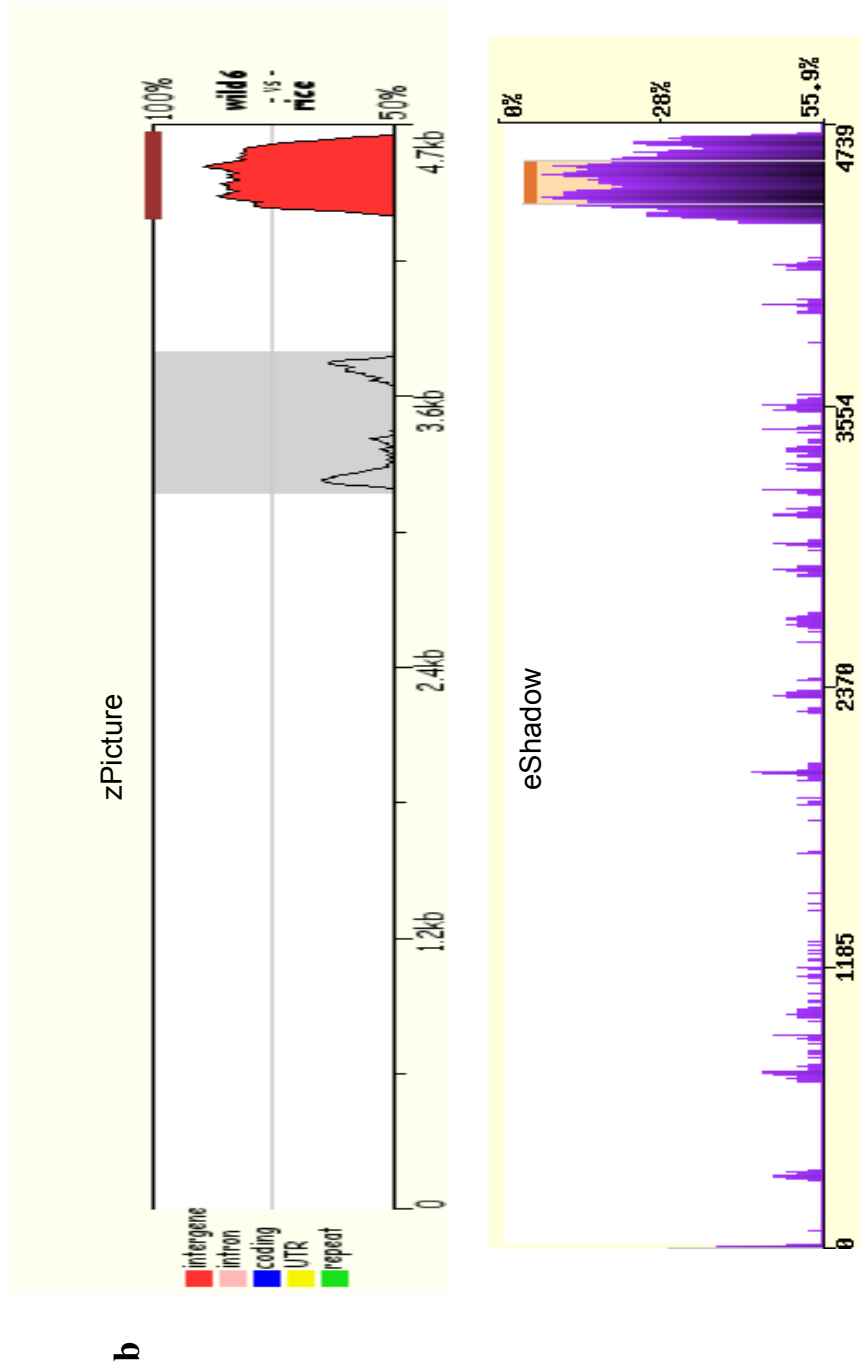
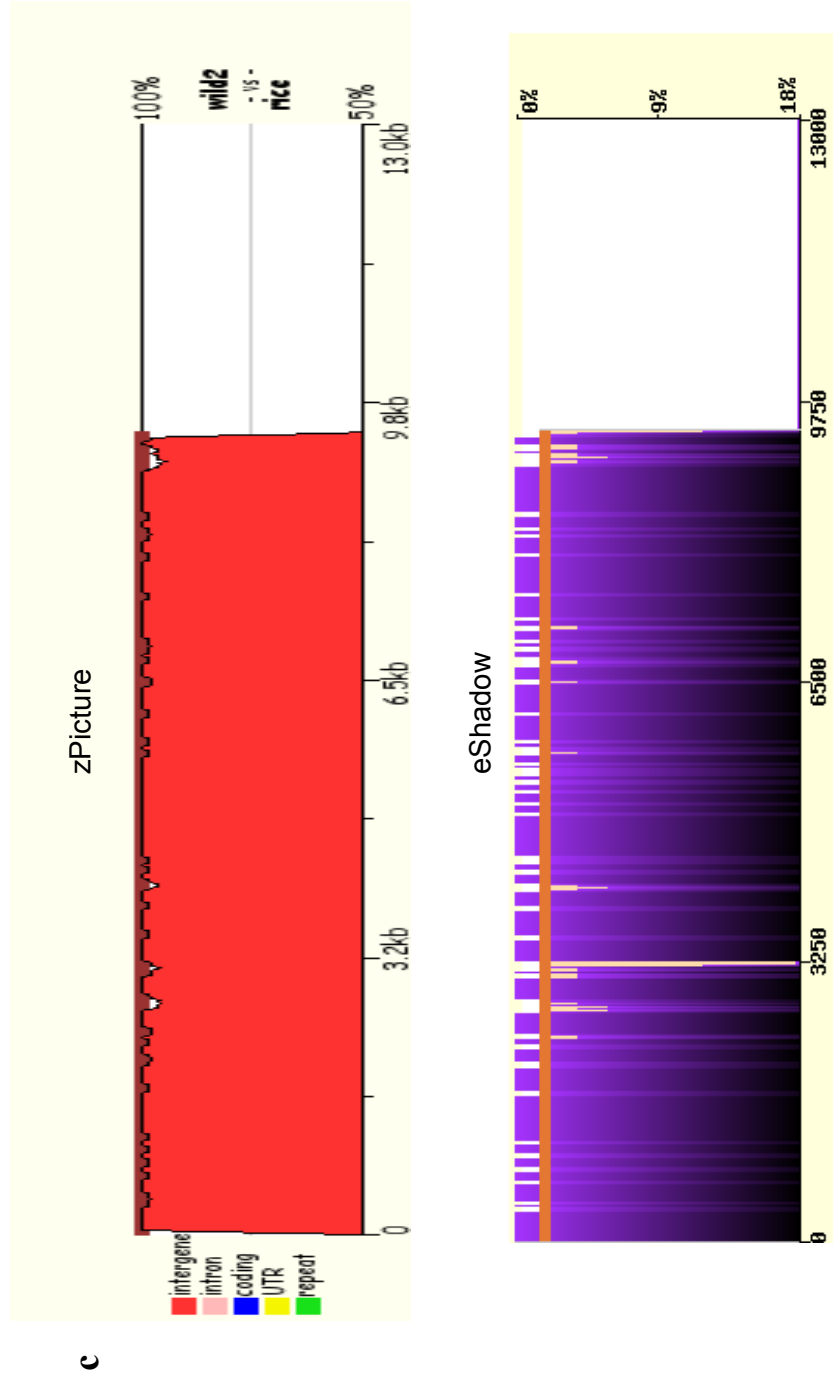


Fig. 4.6 (continued)



**Fig. 4.6** (continued)

unable to determine whether the CentOr satellites are interrupted by other sequences due to the several major gaps in the sequence contigs. According to the length and structure of CentOr unit, the monomers could be classified into three subgroups: 155-bp CentOr (18 copies), 156-bp CentOr (3 copies), and incomplete CentOr (8 copies). The 156-bp CentO had the addition of one nucleotide at the same position. In the incomplete CentOrs, 5 copies had the deletion of 10 nucleotides at the 5'-end, and 3 copies showed the internal deletions of 1-3 nucleotide(s) at different positions. Although all CentOr monomers were conserved in length, the polymorphisms were dispersed along the whole 155-bp consensus satellite sequences. The identities between monomers varied from 88% to 98%.

In the core centromere of the cultivated rice, there were three large clusters of CentO satellite repeats, which comprised 442 units of the 155-bp satellite repeat (J Wu et al. 2004). The alignment of the wild rice CentOr satellites with the cultivated *japonica* rice CentO satellites confirmed that both 5' and 3' ends of the satellite unit were highly conserved (Fig. 4.7), and the central region was relatively divergent (Fig. 4.8). These results were consistent with the previous comparison of *japonica* rice CentO satellites with the maize CentC satellites (Cheng et al. 2002). Interestingly, the CentO monomers in *japonica* centromere 8 do not have a deletion of 10 nucleotides at the 5'-end found in the wild rice CentOr monomers. Instead, *japonica* satellites have 9 copies of 72-bp deletion at 5' end (J Wu et al. 2004). The 83-bp incomplete CentO satellite monomers might be derived from 145-bp incomplete CentOr satellite by further deletion of 5'-end CentOr satellite during evolution.

5' conserved domain

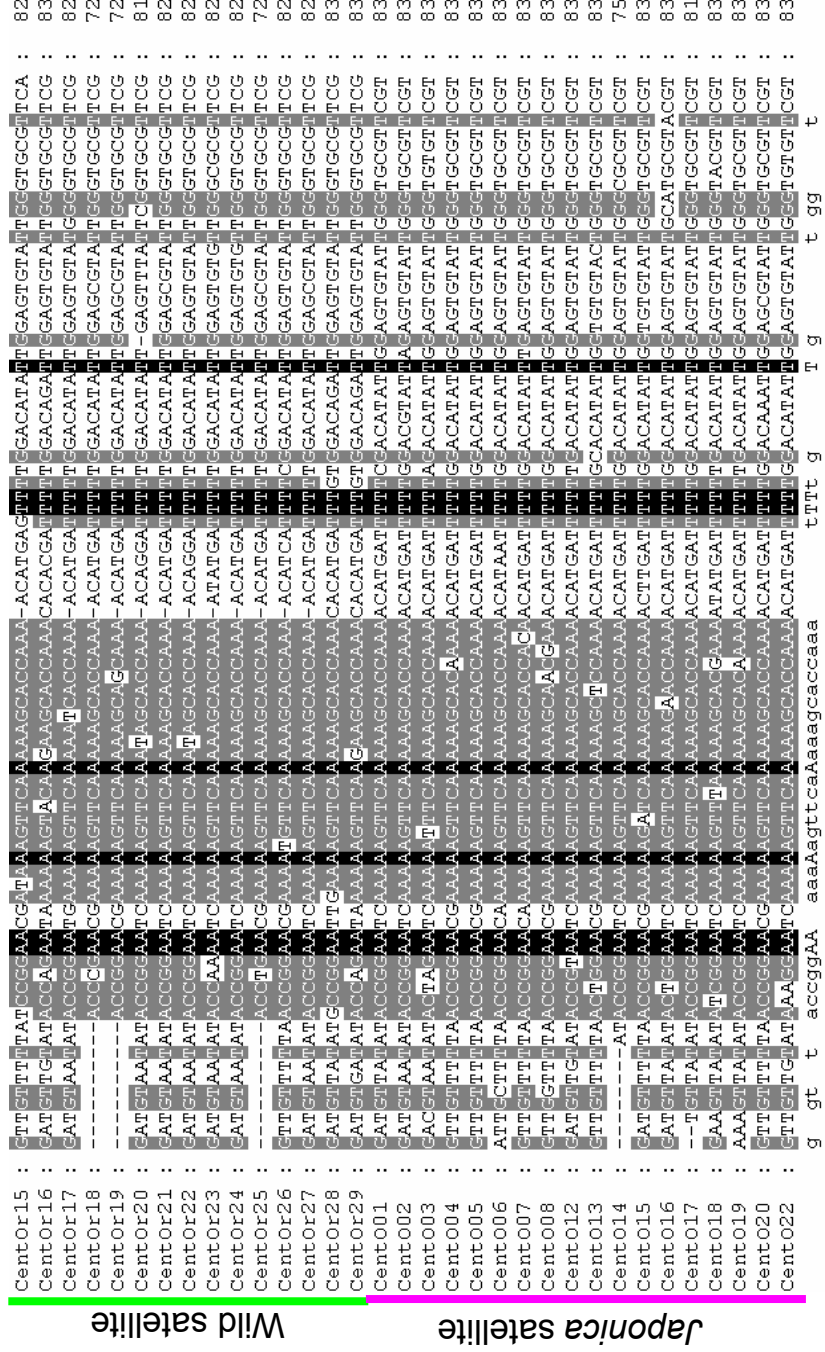


Fig 4.7 Sequence similarity between the wild rice CentOr and japonica rice CentO satellite monomers. Sequence conservation between the CentO repeats is indicated by shaded background, with black indicating 100% conservation and gray indicating 80% or higher conservation.

3' conserved domain

CentOr15	: TGGCAAA	ACTCAC	TCCAGACTCGCGG	CGGAC	TTTTTTGTTAAATGTCGATAT	TGCAC	ACCTGGGTGC	: 155
CentOr16	: TGGCAAA	GCTCAC	TCGGTACTCGGC	GCGAC	TTTTTTCATTAATGCCGATAT	TGAACA	ACCTGGGTGC	: 156
CentOr17	: TGGCAAA	ACTCAC	TCGGGACTCACGT	GCGAC	TTTTTTCATTAATGCCGATAT	TGCAC	CTGTGGGTGC	: 155
CentOr18	: TGGCAAA	ACTCAC	TCGGGACTCGCGC	GCGAC	TTTTTTCATTAATGCCGATAT	TGAACA	ACCTGGGTGC	: 145
CentOr19	: AGGCAAA	ACTCAA	TCGGGCTCGCGC	GCGAC	TTTTTTCATTAATGCCGATAT	TGCAC	ACCTGGGTGC	: 145
CentOr20	: TGGCAAA	AGTCAC	TCGGGACTACGCA	GCGAC	TTTTTTCATTAATGCCGATAT	TGAACA	ACCTGGGTGC	: 154
CentOr21	: TGGCAAA	ACTCAC	TCGGGACTCGCGC	GCGAC	TTTTTTCATTAATGCCGATAT	TGAACA	ACCTGGGTGC	: 155
CentOr22	: TGGCAAA	ACTCCC	TCGGGACTCGAGC	GCGAC	TTTTTTCATTAATGCCGATAT	TGCAC	CTGTGGGTGC	: 155
CentOr23	: TGGCAAA	ACTCAC	TCGGCACTCGCGT	GCGAC	TTTTTTCATTAATGCCGATAT	TGCAC	CTGTGGGTGC	: 155
CentOr24	: TGGCAAA	ACTCAC	---GGACTCGGT	GCGAC	TTTTTTCATTAATGCCGATAT	TGCAC	CTGTGGGTGC	: 152
CentOr25	: AGGCAAA	ACTCAC	TCGGGACTCGCGC	GCGAC	TTTTTTCATTAATGCCGATAT	TGAACA	ACCTGGGTGC	: 145
CentOr26	: AGGCAAA	ACATAC	TCACGACTCGAC	GCGAC	TTTTTTCATTAATGCCGATAT	TGCAC	ACCTGGGTGC	: 155
CentOr27	: TGGCAAA	ACTCAC	TCGGGACTCGCGC	GCGAC	TTTTTTCATTAATGCCGATAT	TGCAC	ACCTGGGTGC	: 155
CentOr28	: TGGCAAA	GCTCAC	TCGGTACTCGCGC	GCGAC	TTTTTTCATTAATGCCGATAT	TGCAC	ACCTGGGTGC	: 156
CentOr29	: TGGCAAA	GCTCAC	TCGGTACTCGCGC	GCGAC	TTTTTTCATTAATGCCGATAT	TGCAC	ACCTGGGTGC	: 155
Cent001	: GCAAAA	CTCACT	TCGGACTCGCGG	TGAAC	TTTTTTCATTAATGCCGATAT	TGCAC	ACCTGGGTGC	: 155
Cent002	: GCAAAA	CTCACT	TCGGACTCGCGG	TGAAC	TTTTTTCATTAATGCCGATAT	TGCAC	ACCTGGGTGC	: 155
Cent003	: GCAAAA	CTTACT	TCGGACTCGCGG	TGAAC	TTTTTTCATTAATGCCGATAT	TGCAC	ACCTGGGTGC	: 155
Cent004	: GCAAAA	CTTACT	TCGGACTCGCGG	TGAAC	TTTTTTCATTAATGCCGATAT	TGCAC	ACCTGGGTGC	: 155
Cent005	: GCAAAA	CTCACT	TCGGACTCGCGG	TGAAC	TTTTTTCATTAATGCCGATAT	TGCAC	CTGTGGGTGC	: 155
Cent006	: GCAAAA	CTCACT	TCGGACTCGCGG	TGAAC	TTTTTTCATTAATGCCGATAT	TGCAC	CTGTGGGTGC	: 155
Cent007	: GCAAAA	CTCACT	TCGGACTCGCGG	TGAAC	TTTTTTCATTAATGCCGATAT	TGCAC	CTGTGGGTGC	: 155
Cent008	: GCAAAA	CTCACT	TCGTGACTCGCGG	TGAAC	TTTTTTCATTAATGCCGATAT	TGCAC	CTGTGGGTGC	: 154
Cent012	: GCAAAA	CTCAT	TCGGACTCGCGG	CGATC	TTTTTTCATTAATGCCGATAT	TGCAC	CTGTGGGTGC	: 155
Cent013	: GCGAAA	CTCAGT	TCGGACTCGCGG	GGAAC	TTTTTTCATTAATGCCGATAT	TGCAC	ACCTGGGTGC	: 155
Cent014	: GCAAAA	CTCACT	TCGGACTCGCGG	GGAAC	TTTTTTCATTAATGCCGATAT	TGCAC	ACCTGGGTGC	: 147
Cent015	: GCAAAA	CTCACT	TCGGACTCGCGG	GGAAC	TTTTTTCATTAATGCCGATAT	TGCAC	CTGTGGGTGC	: 155
Cent016	: GCAAAA	CTCACT	TCGGACTCGCGG	GGAAC	TTTTTTCATTAATGCCGATAT	TGCAC	CTGTGGGTGC	: 154
Cent017	: GCAAAA	CTCACT	TCGCAACTCGCGG	CAAA	TTTTTTCATTAATGCCGATAT	TGCAC	ACCTGGGTGC	: 153
Cent018	: GCAAAA	CTCACT	TCGGACTCGTGC	GGAAC	TTTTTTCATTAATGCCGATAT	TGCAC	ACCTGGGTGC	: 155
Cent019	: GCAAAA	CTCACT	TCGGACTCGCGG	GGAAC	TTTTTTCATTAATGCCGATAT	TGCAC	ACCTGGGTGC	: 155
Cent020	: GCAAAA	CTCACT	TCGCTACTCGCGG	GGAAC	TTTTTTCATTAATGCCGATAT	TGCAC	ACCTGGGTGC	: 155
Cent022	: GGTATA	CTCACT	TCGGGCTCGCGG	CGGGC	TTTTTTCATTAATGCCGATAT	TGCAC	ACCTGGGTGC	: 155

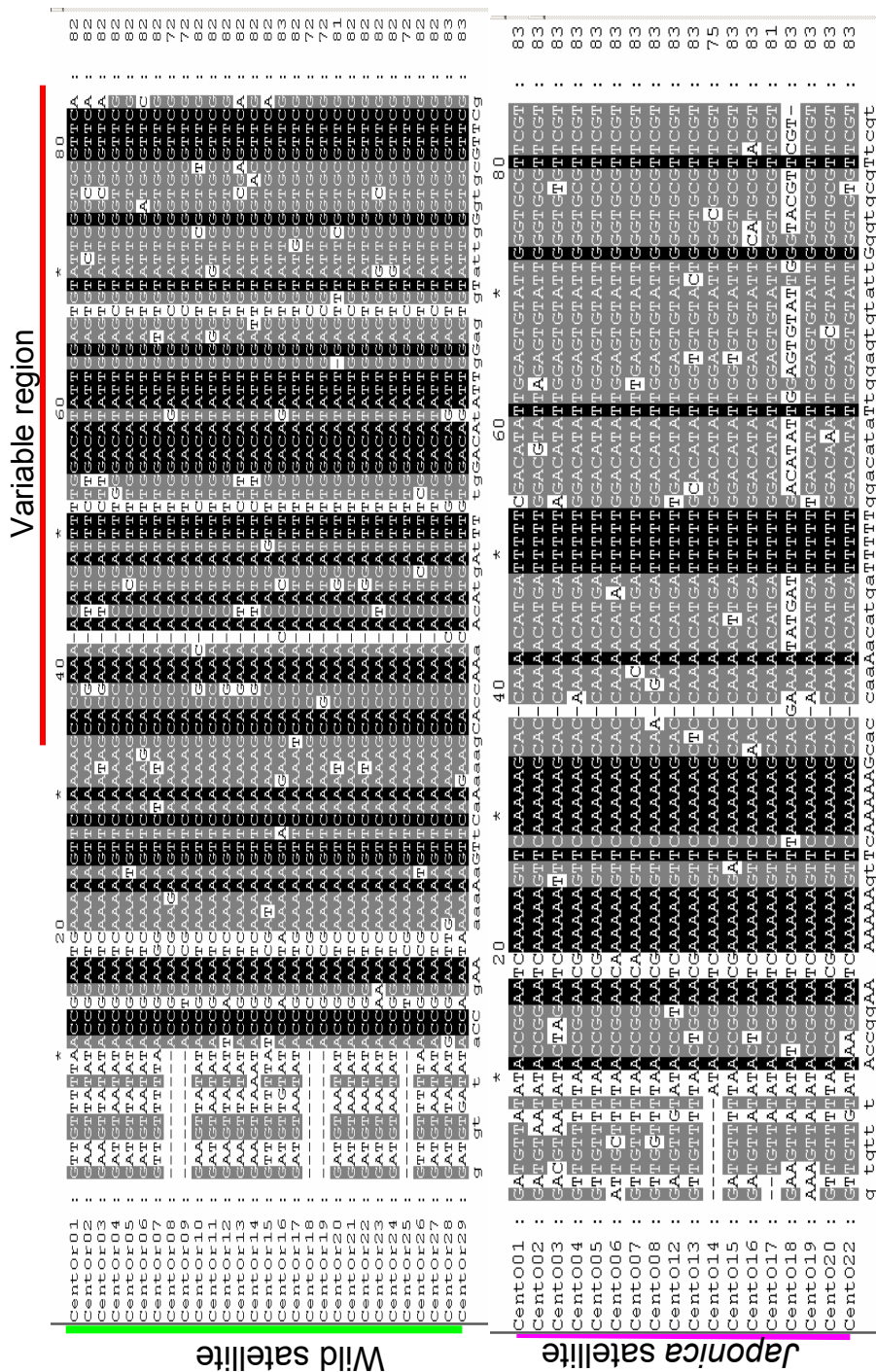
G aaAa t g a TttTgtcaattAatgccgatatTgccAc Cgtgggtgc

Wild satellite

Japonica satellite

Fig 4.7 (continued)





**Fig. 4.8** Comparison between alignments of the wild rice 155-bp CentOr and japonica rice 155-bp CentO satellite monomers. Sequence conservation between the repeats is indicated by shaded background, with black representing 100% conservation and gray 80% or higher conservation.



## Discussion

Putative TGF (TGF-beta receptor-interacting protein) and RER1 (endoplasmic reticulum retrieval protein) genes were identified at the pericentromeric regions flanking the core centromeric region of chromosome 8 in *japonica* rice (Nagaki et al. 2004). To identify the clones covering the whole region of centromere 8 in the wild rice, we designed two overgo (overlapping oligonucleotide) probes from the unique sequences in the putative RER1 and TGF genes. However, there are two partial sequences of the RER1 gene in chromosomes 1 and 4, and two whole sequences of the TGF gene in chromosomes 5 and 11, but no CentO satellites in those regions. Interestingly, BLASTN search showed that there is a partial sequence of RER1 in the centromere 4 and a whole sequence of TGF in the centromere 5. Using the centromere 8-specific probes, the positive BIBAC clones were successfully identified from the wild rice BIBAC library. Southern blot hybridization with CentO satellite sequences as the probes confirmed that the positive clones were from the centromere 8 (Fig. 3.3). The positive clones were further examined by the capillary electrophoresis-based five-enzyme fingerprinting method and assembled into a contig (Fig. 4.1). We confirmed that two of the positive clones that were excluded from the centromere 8 contig were from other chromosomes by Southern and fingerprint analysis.

Two BAC clones, B43P04 and B15E04, in the contig map of the wild rice centromere 8 were selected for sequencing, because they both contain CentOr satellite repeats and the TGF and RER1 locus-specific overgo probes, and span the entire region

of the centromere (Fig 3.2). For the first step of shotgun sequencing, BAC43P04 representing the centromeric region of long arm was selected due to its small insert size (76 kb).

The sequence map comprised 12 sequence contigs and 42 singletons, spanning about 78 kb in length (Table 4.1). The BLASTX search showed that the sequence region contains genes, including one RER1 gene, which is consistent with the sequence analysis of *japonica* rice centromere 8 (J Wu et al. 2004). The whole region-to-region structural comparison of the wild rice B43P09 clone with *japonica* centromere 8 could not be done until completing the final step of filling gaps. However, we found the evidence that there were chromosomal rearrangements at both large scale and nucleotide level in the pericentromeric regions, and rapid changes of satellites in the core centromere. These changes may result from the processes of evolution and/or domestication. Centromeres have been recognized as rapidly evolving regions of chromosomes. Comparative studies of closely related primate species have revealed blocks of recently duplicated sequences located in the pericentromeric regions, and showed dramatic differences in the distribution and organization of these duplications (Eicher and Sankoff 2003). In *Arabidopsis*, significant structural changes such as the integration of the mitochondrial sequences into CEN2 and insertion of a tract of 5S rDNA into CEN3 have been reported (Franz et al. 1998; Stupar et al. 2001). For the large-scale changes in the pericentromeric regions, Hall et al. (2004) suggested that ectopic recombination might play a role in the evolution of these regions. The evidence from human and *Arabidopsis* centromeric regions showed that ectopic recombination occurs at a relatively high

frequency (Eicher and Sankoff 2003). In addition, the process of domestication from the wild rice to cultivated rice has led to significant chromosomal rearrangement in rice genomes (Xiong et al. 1999; Cai and Morishima 2003). The effects of this strong pressure from its domestication on centromeric regions are still unclear. Our comparative analysis of wild rice and *japonica* rice showed that a deletion of 11.8-kb pericentromeric sequence in the wild rice occurred, while a total of 46.5-kb sequence underwent large deletion of several kilobases or insertion by other sequences. A completion of the sequencing project will provide clearer examples of structural rearrangement of pericentromeric regions during evolution.

In addition to the large-scale rearrangements in the pericentromeric regions, satellites in the core centromere underwent rapid changes at the nucleotide level even within a species. In *Arabidopsis*, ecotype-specific satellite variants have been identified (Hall et al. 2003). The comparison between alignments of CentOr and CentO satellites revealed how rapidly the conserved nucleotides in the satellites changed in the middle region of the repeat units (Fig. 4.8). However, two conserved regions believed to be of functional significance are maintained at 5' and 3' ends (Fig. 4.7), consistent with the analysis of sequence similarity between CentO and maize CentC satellites (Cheng et al. 2002).

Several hypotheses have been suggested to explain how satellite sequence homogeneity can be maintained while still allowing for rapid evolution (Ugarkovic et al. 2002; Hall et al. 2004). In the library hypothesis, genomes contain sets of satellite variants in differing abundance. New satellites can occur by mutation. These satellites

can be homogenized through unequal crossover (Nijman and Lenstra 2001; Hall et al 2004). The older satellite(s) can be outcompeted by newer homogeneous satellites. Another model suggested that satellite evolution is driven by the selection and co-evolution of satellites and centromeric histone H3 proteins (Henikoff et al. 2001). The centromeric histone H3 proteins (CENP-A) specifically bind to satellite DNA. In such a case, a slight advantage in satellite-CENP-A interactions could lead to rapid genomic fixation of satellite arrays. The rapid adaptive evolution of CENP-A has been observed in both *Drosophila* and *Arabidopsis* (Malik et al. 2002; Talbert et al. 2002). In the comparison of the wild rice CentOr and cultivated rice CentO monomers, the library hypothesis could explain that 5'-end 10-bp deleted CentOr satellites of the wild rice would undergo further deletion of 5'end by the process of degenerating inactive satellites, and then become more degenerated 72-bp deleted CentO satellites in *japonica* rice.

For the first time at the sequence level, we performed the region-to-region comparative analysis of centromere 8 between wild rice and the cultivated *japonica* rice. Our comparative analysis provided distinct evidence for the structural changes of centromeric regions at both large scale and nucleotide levels, although our sequencing project is still on-going. Since comparative analyses of closely related species are required to discern the functional significance of structural changes in the centromeres, the completion of our sequencing project will provide better understanding of structural organization and evolution of centromeres. Furthermore, both BAC contig and sequence maps of wild rice will be valuable resource for launching the functional analysis of the

centromere using artificial chromosomes. Especially, the binary BAC clones will facilitate the construction of rice artificial chromosomes for mapping the functional centromeric regions because the cloning vector (pCLD04541) can be directly used for screening transformed plants.

## CHAPTER V

### SUMMARY

Centromeres are the major structurally and functionally specialized features of chromosomes in higher organisms. To decipher the genomic origin, organization, function and evolution of plant centromeres, we developed the BAC and plant-transformation-competent BIBAC-based physical maps for the centromeres of cultivated rice, *Oryza sativa* ssp. *japonica* and *indica*. We screened several BIBAC and BAC libraries of *indica* and *japonica* rice using several cereal centromere-specific repeat elements previously isolated in rice and sorghum, RCE1, pSau3A9, and CentO satellites, as probes. We identified 1,423 pSau3A9-, 1,753 RCE1-, and 309 CentO-positive clones from the *japonica* rice BAC and BIBAC libraries, and 370 pSau3A9-, 1097 RCE1-, and 344 CentO-positive clones from the *indica* rice BAC libraries. Using the high-resolution sequencing gel-based fingerprinting method and two restriction enzymes, *Hind*III and *Hae*III, we fingerprinted the positive clones, integrated the fingerprinted data into the fingerprint databases of our whole-genome physical maps of the subspecies, and assembled the contig maps for the centromeric regions of the two rice subspecies, respectively.

Seventeen contigs were assembled for the core centromeres of the *japonica* rice genome, spanning about 20.97 Mb in physical length, and 26 contigs for the core centromeres of the *indica* rice genome, spanning around 15.86 Mb in physical length. Of the contig maps, three of *japonica* rice and one of *indica* rice were shown to cover



the entire regions of centromeres. Selected BAC end sequence analysis and library hybridization using two centromere 8-specific overgo probes (RER1 and TGF) anchored three of the contigs each to chromosome 3 of *japonica* rice, and chromosome 8 of both *indica* and *japonica* rice.

Analysis of the three cereal centromere-specific repeat elements, RCE1, pSau3A9, and CentO, in the positions of the contig maps suggests that the CentO repeats solely reside at the core centromeric regions whereas the RCE1 and pSau3A9 repeats are preferentially distributed in the pericentromeric regions flanking the core centromeres. The physical maps of the core centromeres of *japonica* and *indica* rice will provide valuable resources for gap filling in the physical maps of the International Rice Genome Sequencing Project (IRGSP) and the functional mapping of centromeres. Furthermore, BIBAC clones in our contigs could facilitate not only the functional study of centromeres, but also the construction of rice artificial chromosomes for practical applications.

In the second part of the study, we constructed one large-insert plant-transformation-competent BIBAC library of the wild progenitor of the cultivated rice, *Oryza rufipogon*, to develop the physical map of the centromere 8 in the wild rice for comparative study of rice centromeres. The library was constructed in the *Bam*HI site of a binary vector pCLD04541, contained 24,192 clones arrayed in 63 384-well microplates, had an average insert size of 163 kb, and covered 5x haploid genome of the wild rice. To determine the genomic evolution of centromeres since the cultivated *japonica* rice evolved and domesticated from its wild progenitor and to identify the sequences

potentially important for centromere functions, we screened the wild rice library with two overgo probes (RER1 and TGF) flanking the chromosome 8 centromere of *japonica* rice, and then verified the positive clones by Southern hybridization with a probe derived from the centromere-specific CentO satellites cloned from the cultivated rice. A total of 12 clones were identified by the library hybridization, nine of which were confirmed to contain the CentO satellites.

Our library is a useful resource for comparative genomics of the rice species, including our comparative analysis of centromeres, and practical use of the agronomically important traits in the wild rice for the cultivated rice improvement through direct genetic transformation of large-insert BIBACs containing genes of agronomic importance.

To date, only very limited comparative studies of centromeres have been performed in higher organisms. Importantly, comparative analysis of centromeres from closely-related species could elucidate conserved functional elements and lead to a better understanding of the genomic organization and evolution of plant centromeres because underlying evolutionary mechanisms differ between euchromatic regions and centromeres.

The last part of the study represents the first report on the comparative analysis of centromeres in closely-related plant species, the Asian cultivated rice (*O. sativa* ssp. *japonica*) and its wild progenitor (*O. rufipogon*). By fingerprinting the positive BIBAC clones isolated from the wild rice library with the centromere 8-specific probes (RER1 and TGF) designed from the sequences flanking the centromere 8 of *japonica* rice on a

capillary sequencer, we constructed the BIBAC-based physical contig map of the centromere 8 of the wild rice. Since the clones, B43P04 and B15E04, constituted the minimal tiling path of the contig map, they were selected for shotgun sequencing. By shotgun sequencing of B43P09, 95% genomic sequence of the long arm side of the wild rice centromere 8 was obtained. Comparison with the centromere 8 of *japonica* rice showed that 69% of the sequences were highly conserved between the two species, but some sites were observed to have nucleotide substitution, insertions, and deletions in *japonica* rice. Large-scale rearrangements were observed in the pericentromeric regions. A total of 46.5-kb sequence underwent large deletion of several kilobases or insertion by other sequences between the cultivated and wild species. An 11.8-kb sequence contig of the wild rice was totally deleted from the compared region in the cultivated rice, but were highly conserved in other regions of centromere 8 and other chromosomes of the cultivated rice. In addition to the large-scale rearrangements in the pericentromeres, the 155-bp satellites in the core centromere were shown to undergo rapid changes at the nucleotide level. Conserved nucleotides of the 155-bp satellites in the internal regions were significantly different between the wild and cultivated rice, indicating dramatic changes in the region during evolution. The wild rice CentOr and cultivated rice CentO satellites maintain conserved nucleotides at 5' and 3' ends, which is consistent with the previous sequence analysis between rice CentO and maize CentC satellites. This result suggests that the 5'- and 3'-end domains of the CentO satellites may play an important role in centromere function.

A region-to-region comparative analysis of the centromere 8 was done between the cultivated rice and its wild progenitor. Our analysis provided a line of distinct evidence for the evolutionary structural changes in both large and small scales in the centromeres of related species during evolution and domestication. Since comparative studies of closely-related species are valuable to recognize the functional significance of structural changes in the centromeres, the completion of our sequencing project will provide a better understanding of the structural organization and evolution of centromeres. Furthermore, our physical BAC contig and sequence map is an important platform to launch the functional analysis of the centromeres using rice artificial chromosomes.

## REFERENCES

- Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796-815
- Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol rep* 9:208-219
- Barry AE, Bateman M, Howman EV, Cancilla MR, Tainton KM, Irvine DV, Saffery R, Choo KH (2000) The 10q25 neocentromere and its inactive progenitor have identical primary nucleotide sequence: further evidence for epigenetic modification. *Chromosoma* 100:156-161
- Bellon MR, Brar DS, Lu BR, Pham JL (1998) Rice genetic resources. In: Dwoling NG, Greenfield SM, Fisher KS (eds) Sustainability of rice in the global food system. Davis, California. Pacific Basin Study Center and IRRI, Manila, pp251-283
- Brar DS, Khush GS (1997) Alien introgression in rice. *Plant Mol Biol* 35:35-47
- Brar DS, Khush GS (2002) Quantitative genetics, genomics and plant breeding. In: Kang MS (ed). Transferring genes from wild species into rice. CABI Publishing, Oxon, UK, pp197-217
- Cai HW, Morishima H (2002) QTL clusters reflect character associations in wild and cultivated rice. *Theor Appl Genet* 104:1217-1228
- Cai HW, Wang XK, Morishima H (2004) Comparison of population genetic structures of common wild rice (*Oryza rufipogon* Griff.), as revealed by analyses of quantitative traits, allozymes, and RFLPs. *Heredity* 92:409-417

- Causse MA, Fulton TM, Choi YG, Ahn SN, Chunwongse J et al. (1994) Saturated molecular map of the rice genome based on an interspecific backcross population. *Genetics* 138:1251-1274
- Chang YL, Henriquez X, Preuss D, Copenhaver GP, Zhang HB (2003) A plant transformation-competent binary BAC library of the *Arabidopsis thaliana* Landsberg ecotype for functional and comparative genomics. *Theor Appl Genet* 106:269-276
- Chang YL, Tao Q, Scheuring C, Ding K, Meksem K, Zhang HB (2001). An integrated map of *Arabidopsis thaliana* for functional analysis of its genome sequence. *Genetics* 159:1231-1242
- Chauhan RS, Farman ML, Zhang H-B, Leong SA (2002) Genetic and physical mapping of a rice blast resistance locus, *Pi-CO39(t)*, that corresponds to the avirulence gene *AVRI-CO39* of *Magnaporthe grisea*. *Mol Genet Genomics* 267: 603-612
- Cheng Z, Dong F, Langdon T, Ouyang S, Buell CB, Gu M, Blattner FR, Jiang J (2002) Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell* 14: 1691-1704
- Copenhaver GP, Nickel K, Kuromori T, Benito MI, Kaul S, Lin X, Bevan M, Murphy G, Harris B, Parnell LD et al. (1999) Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* 286:2468-2474
- Cui Y, Bi YM, Brugiere N, Arnoldo M, Rothstein SJ, (2000) The S locus glycoprotein and the S receptor kinase are sufficient for self-pollen rejection in *Brassica*. *Proc Natl Acad Sci USA* 97: 3713-3717

- Dong F, Miller JT, Jackson SA, Wang GL, Ronald PC, Jiang J (1998) Rice (*Oryza sativa*) centromeric regions consist of complex DNA. Proc Natl Acad Sci USA 95: 8135-8140.
- Dong J, Teng W, Buchholz WG, Hall TC (1996) *Agrobacterium*-mediated transformation of *japonica* rice. Mol Breed 2:267-276
- du Sart D, Cancilla MR, Earle E, Mao JI, Saffery R, Tainton KM, Kalitsis P, Martyn J, Barry AE, Choo KHA (1997) A functional neo-centromere formed through activation of a latent human centromere and consisting of non- $\alpha$ -satellite DNA. Nat Genet 16:144-153
- Eichler EE, Sankoff D (2003) Structural dynamics of eukaryotic chromosome evolution. Science 301:793-797
- Fang X, Gu S, Xu Z, Chen F, Guo D, Zhang H-B, Wu N (2004) Construction of a binary BAC library for an apomictic monosomic addition line of *Beta corolliflora* in sugar beet and identification of the clones derived from the alien chromosome. Theor Appl Genet 108:1420-1425
- Feng Q, Zhang Y, Hao P, Wang S, Fu G, Huang Y et al. (2002) Sequence and analysis of rice chromosome 4. Nature 420: 316-320
- Franz P, Armstrong S, Alonso-Blanco C, Fischer TC, Torres-Rulz RA, Jones G (1998) Cytogenetics for the model system *Arabidopsis thaliana*. Plant J 13:867-876
- Gao L, McCarthy EM, Ganko EW, McDonald JF (2003) Evolutionary history of *Oryza sativa* LTR retrotransposons: a preliminary survey of the rice genome sequences. BMC Geomics 5:18

- Hall AE, Keith KC, Hall SE, Copenhaver GP, Preuss D (2004) The rapidly evolving field of plant centromeres. *Curr Opin Plant Biol* 7:108-114
- Hall SE, Kettler G, Preuss D (2003) Centromere satellites from *Arabidopsis* populations: maintenance of conserved and variable domains. *Genome Res* 13:195-205
- Hamilton CM, Frary A, Lewis C, Tanksley SD (1996) Stable transfer of intact high molecular weight DNA into plant chromosomes. *Proc Natl Acad Sci USA*. 93: 9975-9979
- Harlan JR (1975) *Crops and man*. American Society of Agronomy, Madison, WI.
- Harrington JJ, Bokkelen GV, Mays RW, Gustashaw K, Willard HF (1997) Formation of *de novo* centromeres and construction of first-generation human artificial microchromosomes. *Nat Genet* 15:345-355
- Harushima Y et al. (1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics* 148:479-494
- Hass BL, Pires JC, Porter P, Phillips RL, Jackson SA (2003) Comparative genetics at the gene and chromosome levels between rice (*Oryza sativa*) and wildrice (*Zizania palustris*). *Theor Appl Genet* 107:773-782
- He RF, Wang Y, Shi Z, Ren X, Zhu L, Weng Q, He GC (2003) Construction of a genomic library of wild rice and *Agrobacterium*-mediated transformation of large insert DNA linked to BPH resistance loci. *Gene* 321:113-121
- Hegemann JH, Fleig UN (1993) The centromere of budding yeast. *Bioassays* 15:451-460.
- Henikoff S, Ahmad K, and Malik HS. 2001. The centromere complex: stable inheritance with rapidly evolving DNA. *Science* 293:1098-1102



- Henning KA, Novotny EA, Compton ST, Guan XY, Liu PP, Aslock MA (1999) Human artificial chromosomes generated by modification of a yeast artificial chromosome containing both human alpha satellite and single-copy DNA sequences. *Proc Natl Acad Sci USA* 96:592-597
- Heslop-Harrison JS, Brandes A, Schwarzacher T (2003) Tandemly repeated DNA sequences and centromeric chromosomal regions of *Arabidopsis* species. *Chromosome Res* 11:241-253
- Hiei Y, Ohta S, Komari T, Kumashiro T (1994) Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J* 6:271-282
- Ikeno M, Grimes B, Okazaki T, Nakano M, Saitoh K, Hoshino H, McGill NI, Cooke H, Masumoto H (1998) Construction of YAC-based mammalian chromosomes. *Nat Biotechnol* 16:431-469
- International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860-921
- Jiang J, Nasuda S, Dong F, Scherrer CW, Woo SS, Wing RA, Gill BS, Ward DC (1996) A conserved repetitive DNA element located in the centromeres of cereal chromosomes. *Proc Natl Acad. Sci USA* 93:14210-14213
- Jones JG, Shlumukov L, Carland F, English J, Scofield SR, Bishop GJ, Harrison K (1992) Effective vectors for transformation, expression of heterologous genes, and assaying transposon excision in transgenic plants. *Transgenic Res* 1:285-297

- Kahler AL, Allard RW (1981) Worldwide patterns of genetic variation among four esterase loci in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 59:101-111
- Kaszas E, and Birchler JA. 1998. Meiotic transmission rates correlate with physical features of rearranged centromeres in maize. *Genetics* 150:1683-1692
- Khairallah MM, Sears BB, Adams MM (1992) Mitochondrial restriction fragment length polymorphisms in wild *Phaseolus vulgaris* L: insights on the domestication of the common bean. *Theor Appl Genet* 84:915-922
- Langdon T, Seago C, Mende M, Leggett M, Thomas H, Forster JW, Thomas H, Hones RN, and Jenkins G. 2000. Retrotransposon evolution in diverse plant genomes. *Genetics* 150: 1683-1692
- Li Y, Uhm T, Ren C, Wu C, Santos TS, Lee MK, Santos F, Zhang A, Xu Z, Liu D, Zhang HB (2004) A plant-transformation-competent BIBAC-based physical map of *japonica* rice for functional analysis of the rice genome. *Genetics* (submitted)
- Liu KD, Yang GP, Zhu SH, Zhang Q, Wang XM, Saghai Maroof MA (1996) Extraordinarily polymorphic ribosomal DNA in wild and cultivated rice. *Genome* 39:1109-1116
- Liu YG, Liu H, Chen L, Qiu W, Zhang Q, Wu H, Yang C, Su J, Wang Z, Tian D, Mei M (2002) Development of new transformation-competent artificial chromosome vectors and rice genomic libraries for efficient gene cloning. *Gene* 282:247-255
- Malik HS, Vermaak D, Henikoff S (2002) Recurrent evolution of DNA-binding motif in the *Drosophila* centromeric histone. *Proc Natl Acad Sci USA* 99:1449-1454

- Meksem K, Ruben E, Zobrist K, Hyten D, Tao Q, Zhang HB, Lightfoot AD (2000) Two large-insert soybean genomic libraries constructed in a binary vector: applications in chromosome walking and genome-wide physical mapping. *Theor Appl Genet* 101:747-755
- Men AE, Meksem K, Kassem MA, Lohar D, Stiller J, Lightfoot D, Gresshoff PM (2001) A bacterial artificial chromosome library of *Lotus japonicus* constructed in an *Agrobacterium tumefaciens*-transforable vector. *Mol Plant Microbe Interact* 14:422-425
- Miller JT, Dong F, Jackson SA, Song J, Jiang J (1998) Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. *Genetics* 150:1615-1623
- Moulet, O, Zhang H-B, Lagudah ES (1999) Construction and characterisation of a large DNA insert library from the D genome of wheat. *Theor Appl Genet* 99:305-313
- Nagaki K, Cheng Z, Ouyang S, Talbert P, Kim M, Jones KM, Henikoff S, Buell CR, Jiang J (2004) Sequencing of a rice centromere uncovers active genes. *Nat Genet* 36:138-145
- Nijman IJ, Lenstra JA (2001) Mutation and recombination in cattle satellite DNA: a feedback model for the evolution of satellite DNA repeats. *J Mol Evol* 52:361-371
- Nonomura KI, Kurata N (1999) Organization of the 1.9-kb repeat unit RCE1 in the centromeric region of rice chromosomes. *Mol Gen Genet* 261:1-10
- Oka HI (1988) Origin of cultivated rice. Japanese Scientific Society Press/Elsevier, Amsterdam

- Ovcharenko I, Boffelli D, Loots GG (2004a) eShadow: a tool for comparing closely related sequences. *Genome Res* 14:1191-1198
- Ovcharenko I, Loots GG, Hardison RC, Miller W, Stubbs L (2004b) zPicture: dynamic alignment and visualization tool for analyzing conservation profiles. *Genome Res* 14:472-477
- Presting GG, Malysheva L, Fuchs J, Schubert I (1998) A *Ty3/gypsy* retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. *Plant J* 16: 721-728
- Ren C, Lee MK, Yan B, Ding K, Cox B, Romanov MN, Price JA, Dodgson JB, Zhang HB (2003a) A BAC-based physical map of the chicken genome. *Genome Res* 13:2754-2758
- Ren C, Xu ZY, Sun S, Lee M-K, Wu C, Zhang HB (2003b) Genomic libraries for physical mapping. In: Meksem K and Kahl G (eds.) *Handbook of plant genome mapping*. Wiley-VCH Verlag GmbH, Weinheim, Germany (in press)
- Ren F, Lu BR, Li S, Huang J, Zhu Y (2003) A comparative study of genetic relationships among the AA-genome *Oryza* species using RAPD and SSR markers. *Theor Appl Genet* 108:113-120
- Saghai Maroof MA, Zhang Q, Biyashev R (1995) Comparison of restriction fragment length polymorphism in wild and cultivated barley. *Genome* 38:298-306
- Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T et al. (2002) The genome sequence and structure of rice chromosome 1. *Nature* 420:312-316

- Schueler MG, Higgins AW, Rudd MK, Gustashaw K, Willard HF (2001) Genomic and genetic definition of a functional human centromere. *Science* 294:109-115
- Second G (1982) Origin of the genetic diversity of cultivated rice (*Oryza spp.*): study of the polymorphism scored at 40 isozyme loci. *Jpn J Genet* 57:25-57
- Soderlund C, Humphray S, Dunham A, French L (2000) Contigs built with fingerprints, markers, and FPC V4.7. *Genome Res* 10:1772-1787
- Soderlund C, Longden I, Mott R (1997) FPC: a system for building contigs from restriction fingerprinted clones. *CABIOS* 13:523-535
- Stupar RM, Lilly JW, Town CD, Cheng Z, Kaul S, Buell CR, Jiang J (2001) Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: implication of potential sequencing errors caused by large-unit repeats. *Proc Natl Acad Sci USA* 98:5099-5103
- Talbert PB, Masuelli R, Tyagi AP, Comai L, Henikoff S (2002) Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. *Plant Cell* 14:1449-1454
- Tao Q, Wang A, Zhang HB (2002) One large-insert plant-transformation-competent BIBAC library and three BAC libraries of *japonica* rice for genome research in rice and other grasses. *Theor Appl Genet* 105:1058-1066
- Tao Q, Chang YL, Wang J, Chen H, Islam-Faridi MN, Scheuring C, Wang B, Stelly DM, Zhang HB (2001) Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprinting analysis. *Genetics* 158:1711-1724.

- Ugarkovic D, Plohl M (2002) Variation in satellite DNA profiles causes and effects. EMBO J 21:5955-5959
- Uozu S, Ikehashi H, Ohmido N, Ohtsubo H, Ohtsubo E, Fukui Kiichi (1997) Repetitive sequences: cause for variation in genome size and chromosome morphology in the genus *Oryza*. Plant Mol Biol 35:791-799
- Wu C, Sun S, Nimmakayala P, Santos FA, Meksem K, Springman R, Ding K, Lightfoot DA, Zhang HB (2004a) A BAC- and BIBAC-based physical map of the soybean genome. Genome Res 14:319-326
- Wu C, Nimmakayala P, Santos FA, Springman R, Scheruing C, Meksem K, Lightfoot DA, Zhang HB (2004b) Construction and characterization of a soybean bacterial artificial chromosome library and use of multiple complementary libraries for genome physical mapping. Theor Appl Genet (in press)
- Wu C, Xu Z, Zhang H-B (2004c) DNA Libraries. In: Meyers RA (ed.) Encyclopedia of Molecular Cell Biology and Molecular Medicine. Volume 3 (2<sup>nd</sup> Edition), Wiley-VCH Verlag GmbH, Weinheim, Germany. pp385-425
- Wu J, Yamagata H, Hayashi-Tsugane M, Hijishita S, Fujisawa M, Shibata M, Ito Y, Nakamura M, Sakaguchi R, Yohihara H, Ito K, Karasawa W, Yamamoto M, Saji S, Katagiri S, Kanamori H, Namiki N, Katayose Y, Matsumoto T, Sasaki T (2004) Composition and structure of the centromeric region of rice chromosome 8. The Plant Cell 16: 967-976
- Wu Y, Tulsieram L, Tao Q, Zhang H-B, Rothstein SJ (2000) A binary vector-based large insert library for *Brassica napus* and identification of clones linked to a

- fertility restorer locus for *Ogura* cytoplasmic male sterility (CMS). *Genome* 43: 102-109
- Xiao J, Grandillo S, Ahn SN, McCouch SR, Tanksley SD (1996) Genes from wild rice improve yield. *Nature* 384:223-224
- Xiong LZ, Liu KD, Dai XK, Xu CG, Zhang Q (1999) Identification of genetic factors controlling domestication-related traits of rice using F2 population of a cross between *Oryza sativa* and *O. rufipogon*. *Theor Appl Genet* 98:243-251
- Xu Z, Sun S, Ding K, Covalada L, Zhang HB (2002) Accuracy of physical map assembly and fingerprinting methods. Plant, animal and Microbe Genomes X Conference. Jan 12-16, San Diego, California, Scherago International, Inc., pp149
- Yamamoto K, Sasaki T (1997) Large-scale EST sequencing in rice. *Plant Mol Biol* 35:135-144
- Yuan LP, Virmani SS, Mao CX (1989) Hybrid rice: achievements and further outlook. In: Progress in irrigated rice research. International Rice Research Institute, Manila, The Philippines, pp219-223
- Zhang HB, Wing RA (1997) Physical mapping of the rice genome with BACs. *Plant Mol Biol* 35:115-127
- Zhang Y, Huang Y, Zhang L, Li Y, Lu T, Lu Y, Feng Q, Zhao Q, Cheng Z, Xue Y, Wing RA, Han B (2004) Structural features of the rice chromosome 4 centromere. *Nucleic Acids Res* 32:2023-2030
- Zhao Q, Zhang Y, Cheng Z, Chen M, Wang S, Feng Q, et al. (2003) A fine physical map of the rice chromosome 4. *Genome Res* 12:817-823

Zhong CX, Marshall JB, Topp C, Mroczek R, Kato A, Nagaki K, Birchler JA, Jiang J, Dawe RK (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. *Plant Cell* 14:2825-2836



## VITA

Name: **TAESIK UHM**  
 Address: Department of Soil & Crop Sciences, and Institute for Plant Genomics & Biotechnology, Rm 148, 2123 TAMUS, Texas A&M University, College Station, TX 77843, USA  
 Phone: (O) 979-862-4800 & (H) 972-964-7584  
 Fax: 979-862-4790  
 E-mail: tuhm@tamu.edu or tasik47@yahoo.co.kr

## EDUCATION

2004 Philosophy of Doctor, Plant Physiology, Texas A&M University, College Station, TX  
 1993 Master of Science, Botany, Seoul National University, Republic of Korea  
 1991 Bachelor of Science, Botany, Seoul National University, Republic of Korea

## PUBLICATIONS

1. **Uhm T**, Ying R, Xu Z, Covalada L, Scheuring C, Zhang HB. A large-insert plant transformation-competent BIBAC library of the wild rice, *Oryza rufipogon* L. and a BIBAC-based physical map of the wild rice chromosome 8 centromere. Manuscript in preparation
2. **Uhm T**, Li Y, Xu Z, Zhang HB. BAC/BIBAC-based physical maps and genomic organization of centromeres of cultivated rice, *Oryza sativa* L. ssp. *japonica* and ssp. *indica*. Manuscript in preparation
3. Li Y, **Uhm T**, Ren C, Wu C, Santos TS, Lee MK, Santos F, Zhang A, Xu Z., Liu D, Zhang HB (2004) A plant-transformation-competent BIBAC-based integrated physical map of japonica rice for functional analysis and genetic engineering of the rice genome. Genetics (submitted)
4. Choi HK, Kim DJ, **Uhm T**, Limpens E., Lim H., Mun JH, Kalo P, Penmetza R.V, Seres A, Kulikova O, Roe BA, Bisseling T, Kiss GB, Cook DR (2004) A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. Genetics 166:1463-1502
5. Cohn JR, **Uhm T**, Ramu S, Nam YW, Kim DJ, Penmetza RV, Wood TC, Denny RL, Young ND, Cook DR, Stacey G (2001) Differential regulation of a family of apyrase genes from *Medicago truncatula*. Plant Physiol 125:2104-2119

## THESES

1993, Feb Nucleotide Sequences of cDNA clones in Garlic Mosaic Virus genome. **Master's Thesis**. Seoul National University, Republic of Korea  
 1991, Feb Characterization of 5'-flanking region in leghemoglobin gene of kidney bean. **Bachelor's Thesis**. Seoul National University, Republic of Korea