DNA FINGERPRINTING AND MOLECULAR CHARACTERIZATION OF MICROSATELLITES AND MINISATELLITES IN RICE

A THESIS SUBMITTED

TO THE **UNIVERSITY OF POONA**

FOR THE DEGREE OF

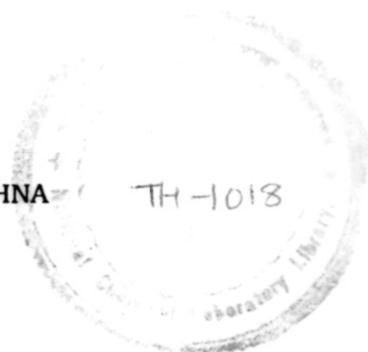
DOCTOR OF PHILOSOPHY

(IN CHEMISTRY)

BY

WUSIRIKA RAMAKRISHNA

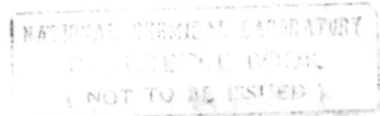
M.Phil.



DIVISION OF BIOCHEMICAL SCIENCES

NATIONAL CHEMICAL LABORATORY,

PUNE - 411 008, (INDIA)



1994

RR
577.2:633.18(043)
RAM

CONTENTS

ACKNOWLEDGEMENTS

DECLARATION

LIST OF ABBREVIATIONS

LIST OF FIGURES

LIST OF TABLES

SYNOPSIS 1

CHAPTER 1 : MINI AND MICROSATELLITES - A NEW ERA OF MOLECULAR MARKERS 11

CHAPTER 2 : MATERIALS AND METHODS 29

CHAPTER 3 : RESULTS 55

CHAPTER 4 : DISCUSSION 99

COMPUTERISED

ACKNOWLEDGEMENTS

I consider it a great privilege to work under the guidance of Dr.P.K. Ranjekar. He has been a constant source of inspiration and encouragement. He has given me full freedom at work. I recall many instances when he has gone out of way to help me in difficult situations.

I would like to express my sincere thanks to Dr. Vidya Gupta for her keen interest in the work and her extra ordinary ability to convert complicated interpretations to simple and understandable language. She has been kind enough to go through the minute details of the thesis.

I do not find any words in dictionary to express my feelings for the invaluable support and help rendered by Venkat, Shastry and Anjali. I consider myself extremely lucky to have a friend in Milind who has stood by me in trying times. My special thanks are due for Meena for her enthusiastic help in my work. I would also like to express my gratitude to Shubhada and B.G. Patil for their untiring efforts with DNA sequencing. During the course of my Ph.D. work, I have enjoyed the companionship of Mohini, Hendre, Nirmala, Lalitha, Usha, Deepak, Sami, Manisha and Armaity. I can recall the lively environment in the initial days of my work attributed to my ex-lab colleagues - Kalpana, Vaiju, Vivek, Prafulla, Radha, Madhur, Suman, Renuka, Urmila, Savita, Sujata & Vishvas.

I am greatly indebted to Dr. R.A. Mashelkar, Director, NCL, Prof John Barnabas and Dr. M.C. Srinivasan, Heads of Biochemical Sciences for providing the laboratory facilities during the tenure of my work. I acknowledge the Council of Scientific and Industrial Research (CSIR) for providing financial assistance through the junior and senior research fellowships. A substantial part of my work was supported through funds from the Rockefeller Foundation and Department of Biotechnology and I express my gratitude to these organizations. The published literature received from Dr.O'Toole was also extremely useful in writing the thesis.

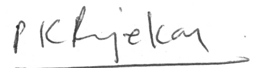
I thank Prof. Jossi Hillel for providing the cattle probe R18.1 and Prof. Jonathan Longmire for probe pV47. I am grateful to Dr. K.N. Ganesh, Head, Organic Chemistry and his group at NCL for synthesizing oligonucleotides and use of PCR machine.

I have enjoyed the company of Ajit while typing my thesis. I greatly appreciate his enthusiasm and timely help. I am thankful to Chittari and Tandel for their help in computer work. I would once again like to express my deep sense of gratitude to all colleagues in Biochemistry & Plant Tissue Culture and other divisions of NCL who have made my stay at NCL a memorable one.

W. Ramakrishna

DECLARATION

Certified that the work incorporated in the thesis entitled "DNA fingerprinting & molecular characterization of micro & minisatellites in rice", submitted by Mr. Wusirika Ramakrishna was carried out by the candidate under my supervision. The material obtained from other sources has been duly acknowledged in the thesis.



(P.K.Ranjekar)

Research Guide

LIST OF ABBREVIATIONS

bp	:	base pairs
kbp	:	kilobase pairs
gm	:	grams
μg	:	micrograms
ng	:	nanograms
ml	:	millilitre
μl	:	microlitre
U	:	units of enzyme
nm	:	nanometer
Ci	:	curies
mCi	:	millicuries
μCi	:	microcuries
rpm	:	revolutions per minute
dCTP	:	deoxycytidine 5'- triphosphate
dATP	:	deoxyadenosine 5'- triphosphate
dGTP	:	deoxyguanosine 5'- triphosphate
dTTP	:	deoxythymidine 5'- triphosphate
Tris	:	Tris-hydroxymethyl amino methane
SDS	:	Sodium dodecyl sulphate
EDTA	:	Ethylene diamine tetra acetic acid
TAE	:	Tris-acetate - EDTA buffer
TBE	:	Tris-borate - EDTA buffer
TPE	:	Tris-phosphate - EDTA buffer
Tm	:	melting temperature
MOPS	:	3 [N-morpholino] propane sulfonic acid buffer

LIST OF FIGURES

FIG.NO.	TITLE
1	Hybridization patterns detected by probe $(CAC)_5$ in *RsaI* digests of genomic DNA of rice.
2	Hybridization patterns detected by probe $(GATA)_4$ in *TaqI* digests of genomic DNA of rice.
3A-F	DNA fingerprints of rice genotypes using oligonucleotide probes a $(GATA)_4$, b $(CAC)_5$, c $(GAA)_6$, d $(GGAT)_4$, e $(GACA)_4$ and f $(TG)_{10}$.
4	Hybridization patterns detected by probe $(GATA)_4$ in *HindIII* digests of DNA from several rice genotypes.
5	DNA fingerprints of a F_2 population of a cross between Basmati-370 (1) and Taichung-65 (2) with $(GATA)_4$ as probe.
6A	Hybridization patterns detected by a telomeric repeat probe in *DraI* digested rice DNAs.
6B	Hybridization patterns detected by a telomeric repeat probe in *HindIII* digested rice DNAs.
7	Hybridization patterns detected by mouse rDNA probe in *DraI* digested rice DNAs.
8A	DNA fingerprints of *DraI* digested rice DNAs probed with R18.1.
8B	DNA fingerprints of *HindIII* digested rice DNAs probed with R18.1.
9	DNA fingerprints of *DraI* digested DNA from rice genotypes probed with pV47.

10 DNA fingerprints of *DraI* digested DNA from rice genotypes probed with M13 repeat probe.

11 DNA fingerprints of Basmati-370 and Taichung-65.

11B DNA fingerprints of a F_2 population of a cross between Basmati-370 (1) and Taichung-65 (2) with R18.1 as probe.

12A Dot blot hybridization of $(GATA)_4$ to total RNA and DNA from different tissues of rice.

12B The same blot was reprobed with 5' end labelled $(CAC)_5$.

13A Northern blot hybridization of rice RNA from different tissues probed with end labelled $(CAC)_5$.

13B Northern blot hybridization of rice poly$(A)^+$ RNA with end labelled $(CAC)_5$.

14A DNA methylation patterns detected in cultivars-Adt-27 & Basmati-370 digested with *MspI, HpaII, MboI, Sau3AI, DpnI* & probed with R18.1.

14B DNA methylation patterns detected in cultivars-Indrayani and Taichung-65 digested with *MspI, HpaII, MboI, Sau3AI, DpnI* and probed with R18.1.

15A Colony hybridization showing two GATA positive clones.

15B Plasmid DNAs from $(GATA)_4$, $(TG)_{10}$, M13, pV47 and R18.1 positive clones.

16A Restriction enzyme digestions of clone $1E_6$.

16B Double digestions of clone $1E_6$.

16C The gel shown in 16B was hybridized with $(GATA)_4$.

17 Restriction map of the clone $1E_6$.

18A	Subclone of $1E_6$ digested with *PvuII* and *HindIII*.
18B	Plasmid DNA of $1E_6$ subclone digested with *PvuII* and *HindIII* hybridized to $(GATA)_4$.
19	Partial DNA sequence of clone $1E_6$ showing the presence of simple repeats.
20	PCR amplified products from rice genotypes using primers flanking GATA and TG simple sequence repeats at $1E_6$ locus.
21	Rice genotypes were PCR amplified as described in figure 20 and digested with *Sau3AI*.

LIST OF TABLES

Table No.	Title
1	Rice genotypes used in the present study
2	Number of bands hybridizing to different oligonucleotide probes in cultivar Basmati-370 digested with different restriction enzymes
3	Analysis of DNA Fingerprints using different oligonucleotide probes
4	Similarity index values for different rice genotypes with probe $(GATA)_4$ & enzyme *HindIII*
5	Analysis of DNA fingerprints of rice using M13, pV47 and R18.1
6	DNA methylation patterns in different rice cultivars

SYNOPSIS

DNA fingerprinting makes use of the presence of hypervariable repetitive DNA sequences which account for a significant proportion of the genome. Short repeat motifs of 2-10bp core sequence repeated in tandem and dispersed in the genome constitute "simple sequences" or "microsatellites". Simple repetitive DNA sequences are a major source of genetic variation (1) and are regarded as the most informative genetic markers (2). GATA/GACA family of simple repeats was originally detected in female specific snake satellite DNA and subsequently found to be conserved throughout the eukaryotes (3). Therefore, probes with simple motifs such as $(GATA)_4$, $(GACA)_4$ and $(CAC)_5$ which detect multiple loci were successfully used for DNA fingerprinting in humans (4).

Minisatellites comprise 10-40bp long core sequences which are also repeated in tandem and dispersed in the genome of most organisms and are hypervariable in nature. Individual specific DNA fingerprints were developed with several highly informative minisatellite probes. Repeats in the human myoglobin gene (5), the wild type M13 phage (6), *chi* sequence of *E.coli*, "per" gene of *Drosophila*, 3' hypervariable region of apolipoprotein B and telomeres were shown to be highly polymorphic in humans. DNA fingerprints in cattle were generated by a human minisatellite probe pV47 (7). Highly polymorphic DNA fingerprint patterns were generated in farm animals and ornamental plants with a bovine genomic clone, R18.1 and in a variety of vertebrates and horses with a simple $(TG)_n$ probe (8). The poly (TG) family has been shown to be the most abundant microsatellite in the human genome.

Since their discovery, mini and microsatellites have been attributed to be involved in transcriptional regulation of adjacent genes, gene conversion, recombination and sex determination. Expression of simple repeats is critical to the overall significance of the conservation of these repeats in eukaryotes.

There are a few reports of poly $(A)^+$ transcripts that include sequences homologous to simple tandem repeats. Thus $(GATA)_n$ and $(GACA)_n$ homologous sequences in mouse and $(CAC)_n$ homologous sequences in dogs have been reported amongst the transcripts. Studies with respect to the methylation status of the mini & microsatellites have been scanty. C-methylation has been reported in humans at the minisatellite loci and in Brassica at the microsatellite loci. Methylation has been implicated in a number of biological phenomena including gene regulation. Study of methylation at minisatellites and microsatellites is especially an interesting aspect as these sequences have been suggested to be hot spots of recombination. On the other hand, methylated sequences have also been showed to be involved in recombination.

In plants, there is a lack of systematic study with reference to the inter and intraspecific variability using minisatellite and microsatellite probes. Rice is the world's single most important food crop and a primary food for more than a third of world's population. *Oryza* is an agronomically important genus with many wild species containing genes for resistance to diseases and insect pests. Wild species of *oryza* provide a rich source of alien gene transfer to cultivated rice. Prior to application of DNA fingerprinting in rice breeding programs, it is necessary to show that mini and microsatellites can detect intra and interspecific variability. Moreover, in rice, 50% of the genome consists of repetitive DNA sequences, hence characterization of repeat families will be of interest in understanding the genome evolution. The polymorphisms as revealed by the repeat sequences in various rice genomes/cultivars will be of immense use to generate probes for varietal identification, seed certification and patent rights.

SUMMARY OF WORK

In the present work, DNA fingerprints were developed for various rice genotypes using micro and minisatellites and several repeat probes. A few such sequences were characterized with respect to their transcriptional activity, methylation status and DNA sequence analysis of the complementary regions in the rice genome. The details of these aspects are summarized as follows

[1] DNA fingerprinting in rice with oligonucleotides specific for simple repetitive DNA sequences (microsatellites)

Six oligonucleotide probes viz. $(GATA)_4$, $(GACA)_4$, $(GGAT)_4$, $(CAC)_5$, $(GAA)_6$ and $(TG)_{10}$ were explored for their ability to generate DNA fingerprints in rice. All these simple sequence motifs were present in rice genome although the level of informativeness and fingerprint profiles revealed by different probes and enzymes varied. Presence of long uninterrupted stretches of simple repeats with lack of sites for tetracutting restriction enzyme was not observed in rice which was an interesting finding. Amongst several restriction enzymes used, *DraI* was found to be the most informative enzyme. Hybridization with $(GATA)_4$, $(CAC)_5$, $(GACA)_4$, $(GAA)_6$ and $(GGAT)_4$ resulted in highly polymorphic strong bands whereas hybridization with $(TG)_{10}$ resulted in a heavy background smear indicating large number of loci containing TG repeats. Rice cultivars belonging to both indica and japonica subspecies and accessions of wild rice belonging to *O.rufipogon* (AA genome) and *O.officinalis* (CC genome) were analyzed. Probability that two different rice genotypes have identical fingerprints was calculated to be 2.5×10^{-9} for $(GATA)_4$, 5.5×10^{-9} for $(GAA)_6$, 1.1×10^{-11} for $(GGAT)_4$, 7.5×10^{-7} for $(CAC)_5$ and 2.1×10^{-5} for $(GACA)_4$.

Since $(GATA)_4$ gave the maximum band resolution, the DNA fingerprinting was extended to more number of cultivated and wild rice genotypes using this probe. With reference to GATA containing loci, indica and japonica subspecies appear to form two distinct groups. Secondly, there is a clear distinction between the wild and cultivated rice genotypes. In addition the level of variation observed between the cultivated and wild rice species belonging to AA genome especially *O.rufipogon* was less as compared to other genomes of rice. The latter supports the view that *O.rufipogon* is the ancestor of *O.sativa* as suggested by Kochko *et al* (7).

[2] DNA fingerprinting in rice using minisatellites and a cattle probe R18.1

In these studies, the presence and polymorphic behavior of minisatellites and several other hypervariable DNA sequences capable of detecting multiple loci in rice were investigated. Oligonucleotide probes corresponding to the consensus core repeat of *chi* sequence of *E.coli*, "Per" gene of *Drosophila* and 3' hypervariable region of apolipoprotein B did not show clear-cut band patterns whereas the telomeric repeat, was found to be remarkably polymorphic. Repeats in the M13 bacteriophage, pV47 - a human minisatellite probe and R18.1 - a cattle probe containing GT repeats could decipher both intra and interspecific variability in seven rice cultivars as well as six wild rice genotypes. Probability that two different rice genoytpes have identical fingerprints was calculated to be 2.0×10^{-10} for M13, 1.5×10^{-11} for pV47 and 2.5×10^{-20} for R18.1. Inheritance of DNA fingerprints detected by R18.1, the most polymorphic probe, was analyzed in a F_2 population of a cross between Basmati-370 and Taichung-65. Here, the bands in DNA fingerprints showed a Mendelian pattern of inheritance.

[3] $(CAC)_5$ detects sequences homologous to gene transcripts in rice

To investigate the presence of transcribed simple repeats in rice, oligonucleotide probes such as $(GATA)_4$, $(GACA)_4$, $(CAC)_5$ and $(TG)_{10}$ were hybridized to dot blot of RNAs isolated from two different tissues, seed and leaf of rice (variety Basmati-370). Interestingly, only $(CAC)_5$ hybridized to dot blot of total RNAs at a concentration as low as 0.5ug RNA suggesting the presence of complementary sequences which were transcriptionally active. Two major bands of 1.9kb and 2.3kb on a background of heterogeneously sized RNA were observed in seed and leaf, respectively indicating presence of tissue specific transcripts in addition to a common band at 1.2kb.

On hybridization of poly$(A)^+$ RNAs with $(CAC)_5$, strong signals corresponding to 1.2kb RNA were obtained in case of seed and leaf. However, hybridization to heterogeneously sized class RNA below 1.2kb was observed in seed and was totally absent in leaf. The hybridization patterns in total RNA and poly$(A)^+$ RNA indicate sequences homologous to $(CAC)_5$ are present in poly$(A)^-$ and poly$(A)^+$ RNA species. Since the use of computer software program DNASIS did not show significant homology of $(CAC)_5$ to ribosomal DNA and histone gene sequences of rice, it is possible that other repetitive sequences may have resulted in a population of RNAs of sizes 2.3 and 1.9kbp giving hybridization to $(CAC)_5$. Analogous to the role proposed for other repeated sequences in gene regulation and evolution (9), similar role can be envisaged for $(CAC)_5$ in rice.

[4] Methylation status at the micro and minisatellite loci in rice

Several rice cultivars such as Indrayani, Taichung-65, Adt-27 and Basmati-370 were digested with methylation sensitive/insensitive isoschizomers such as *MspI*, *HpaII*, *MboI*, *Sau3AI* and *DpnI*. $(GATA)_4$, R18.1 and pV47 were used to investigate the methylation status at the loci detected by these probes in

rice. Digestion with *HpaII* is inhibited when the internal cytosine residue in 5'-CCGG-3' is methylated whereas *MspI* is insensitive to this methylation. Similarly digestion with *Sau3AI* is inhibited when the cytosine residue in 5'-GATC-3' is methylated whereas *MboI* is insensitive to C-methylation but inhibited by adenine methylation. *DpnI* digests only when adenine in 5'-GATC-3' is methylated.

With $(GATA)_4$, some hybridizing fragments in *HpaII* digested DNA were in higher molecular weight region which were absent in MspI digested DNA suggesting a high frequency of methylation at CpG in the sequence CCGG. Several bands were common indicating that these cutting sites are not methylated. Similarly C-methylation and A-methylation were observed in the sequence GATC in the vicinity of GATA containing loci. However, the degree of methylation varied among different rice cultivars with (GATA)4 as well as R18.1 and pV47. With R18.1 and pV47, adenine and cytosine methylation was limited at GATC whereas cytosine methylation was more pronounced at CCGG.

Mini and microsatellites have been suggested to be hotspots of recombination in humans while methylated cytosines have also been shown to serve as recombinational hotspots. Thus a high level of methylation at mini and microsatellites in rice further supports the concept of involvement of these sequences in recombination.

[5] Identification & characterization of a microsatellite from rice

In an attempt to identify a hypervariable probe for DNA fingerprinting from homologous system, *PstI* partial genomic library of rice was screened with $(GATA)_4$, $(TG)_{10}$, R18.1 and pV47, the most informative probes in rice and 40 positive clones were identified. On hybridization of plasmid DNAs of these clones with the above probes, IE_6, a GATA positive clone gave a very strong signal as

compared to others and also hybridized with (TG)10 and R18.1, suggesting a clustering of both GATA and TG repeats in this clone. Since IE_6 could decipher polymorphisms with a set of rice genotypes, its restriction map was constructed which showed sites for enzymes *HindIII*, and *PvuII*. Further, based on hybridization with $(GATA)_4$, it was concluded that GATA repeats in IE_6 were located in a 0.8kbp *PstI-HindIII* fragment which was subcloned for sequencing. Partial sequencing of IE_6 revealed clustering of $(GA)_n/(CT)_n$ and $(TG)_n/(AC)_n$ dinucleotide repeats along with $(GATA)_n/(CTAT)_n$ repeats in the clone. Locus specific amplifications were performed by polymerase chain reaction using primers flanking the GATA and TG repeats. To detect polymorphisms on a fine scale, the 215bp fragment amplified by PCR was digested with *Sau3AI* and analyzed on a sequencing gel. Distinct polymorphisms could be detected amongst different rice genotypes by creation or deletion of *Sau3AI* sites at the IE_6 locus.

Overall conclusions

The use of simple repeats and minisatellites has been shown to generate DNA fingerprints and detect intra and interspecific variability in rice revealing their potential in a wide range of applications in rice breeding programs. These repetitive sequences were further characterized with reference to their transcriptional activity and methylation at cytosine and adenine residues. A homologous probe from rice containing GATA and TG repeats was identified which was used to perform locus specific amplifications in rice. The high level of polymorphism and methylation of these loci in rice strongly suggest a possible role of these sequences in recombination as shown in humans.

REFERENCES

[1] Tautz D, Trick M, Dover GA (1986) Nature 322 : 652

[2] Tautz D (1989) Nucleic Acids Res 17 : 6463

[3] Epplen JT (1988) J Hered 79 : 409

[4] Schafer R, Zischler H, Birsner U, Becker A, Epplen JT (1988) Electrophoresis 9 : 369

[5] Vassart G, Georges M, Monsieur R, Brocas H, Lequarre AS and Christophe D (1987) Science 235 : 683

[6] Jeffreys AJ, Wilson V, Thein SL (1985) Nature 314 : 67

[7] Longmire JL, Kraemer PM, Brown NC, Harde Kopf LC, and Deaven LL (1990) Nucleic Acids Res 18 : 1658

[8] Kochko AD, Kiefer MC, Cordesse F, Reddy AS, Delseny M (1991) Theor Appl Genet 82 : 57

[9] Davidson EH, Posakony JW (1982) Nature 297 : 633

MANUSCRIPTS

[1] DNA fingerprinting in rice using oligonucleotide probes specific for simple repetitive DNA sequences. **W. Ramakrishna**, M.D. Lagu, V.S. Gupta and P.K. Ranjekar. Theoretical and Applied Genetics, 1994 (In Press)

[2] $(CAC)_5$, detects DNA fingerprints and sequences homologous to gene transcripts in rice. V.S. Gupta **W. Ramakrishna**, S.R. Rawat and P.K.Ranjekar. Biochemical Genetics, 1993 (In Press)

[3] DNA fingerprinting in rice to detect genetic variation using hypervariable DNA sequences. **W. Ramakrishna**, K.V Chowdari, M.D Lagu, V.S. Gupta and P.K. Ranjekar. Communicated

[4] Identification and molecular characterization of a microsatellite from rice. **W. Ramakrishna**, K.V. Choudhari, V.S. Gupta and P.K. Ranjekar (Manuscript under preparation)

[5] DNA fingerprinting of *Xanthomonas oryzae* pv *Oryzae* using mini, micro satellite and avirulence gene probes. **W. Ramakrishna**, K.V. Choudhari, S.A. Tamhankar, S. Gnanamanickam, V.S. Gupta and P.K. Ranjekar (Manuscript under preparation)

PRESENTATIONS AT INTERNATIONAL SYMPOSIA

[1] P.K.Ranjekar, **W.Ramakrishna**, K.V.Choudhari, M.D.Lagu, V.S.Gupta. DNA fingerprinting in rice using simple repetitive DNA sequences and their characterization in rice. Plant Genome II, the second Intl conference on the plant genome, Jan 24-27,1994, Sandiego, California, USA, P168 pp56.

[2] V.S.Gupta, **W.Ramakrishna**, P.K.Ranjekar. DNA fingerprinting and detection of gene transcripts in rice using simple repetitive DNA sequences. Second Asia Pacific conference on agricultural biotechnology, Madras, INDIA, March 6-10, 1994.

[3] **W.Ramakrishna**, V.S.Gupta, P.K.Ranjekar. DNA fingerprinting in rice using microsatellites and minisatellites. Abstract, Intl program on rice biotechnology - 7th annual meeting, May 16-21, 1994, Bali, INDONESIA

CHAPTER 1

MINI AND MICROSATELLITES :

A NEW ERA OF MOLECULAR MARKERS

1.1	Definition
1.2	Occurrence
1.3	Some widely used mini and microsatellites
1.3.1	(GATA/GACA)n
1.3.2	(CA/GT)n
1.3.3	(CAC)5
1.3.4	M13 probe
1.3.5	pV47
1.3.6	Chi sequence
1.3.7	Telomeric repeats
1.3.8	Other minisatellite related probes
1.4	PCR mediated microsatellite
1.5	Micro and minisatellites amplifications in plants
1.6	Mechanism for generation of variability
1.7	Transcriptional activity
1.8	Methylation studies
1.9	Applications
1.10	References

1.1 Definition

Repetitive DNA sequences constitute a large fraction of the eukaryotic genome. "Minisatellites" or "VNTR" (variable number of tandem repeats) or "HVR" (highly variable regions) are one class of repeat sequences which comprise 10-40bp long core sequences repeated in tandem and dispersed in the genome of most organisms. When these core sequences are hybridized to genomic DNAs digested with restriction enzymes which cut outside the repeat unit leaving the internal repeats intact at a relatively low stringency, several hypervariable loci are detected simultaneously (1,2). Short motifs of 2-10bp long core sequences repeated in tandem and dispersed in the genome constitute simple repetitive DNA sequences (microsatellites) which can serve as a major source of genetic variation (3,4). These simple sequence repeats (SSR) or microsatellites such as $(TG)_n$, $(CAC)_n$ and $(GATA)_n$ are found to be abundant, highly polymorphic in nature and are flanked by conserved restriction endonuclease sites. The restriction fragment length depends on the number of repeats and thus polymorphism is generated between two individuals/genotypes on hybridization with a probe that detects many hypervariable loci.

1.2 Occurrence

Initially, Wyman and White (5), identified a hypervariable locus with 8 alleles in humans. Several other highly variable regions have been identified near the human insulin gene, alpha-globin gene and c-Ha-ras-1 oncogene (6,7,8). However, the isolation of human minisatellites 33.6 and 33.15 from a human genomic library with a 767bp repeat probe consisting of 23 repeats of a 33 base pair sequence in the intron of myoglobin gene and their use in DNA fingerprinting was a landmark achieved by Jeffreys *et al.* (1). The presence of simple repeats has been demonstrated in a variety of eukaryotic genomes by slot blot

hybridizations (9). This finding was especially promising since unlimited number of these simple repeats could be used for DNA fingerprinting in genomes other humans. For example, complex and highly variable hybridization patterns have been obtained with $(GATA)_4$, $(GACA)_4$, $(CAC)_5$ and $(TG)_n$. Ryskov *et al.* (10) used M13 probe to identify polymorphic regions in barley and subsequently this probe was used to detect minisatellites in a few gymnosperms and angiosperms (11,12). This probe was further used for identification and assessment of genetic variability in Rosaceae (13) and ornamental plants (14). In recent years, oligonucleotide probes specific for simple repeats have been successfully used for DNA fingerprinting in higher plants (15), especially in chickpea, banana, tomato, Brassica, sugarbeet and grapevine (16,17,18,19,20,21).

1.3 Some widely used mini and microsatellites

1.3.1 (GATA/GACA)n

GATA/GACA repeats were first identified as a major component of snake satellite DNA (22) where GATA containing probe such as Bkm probe was isolated (23) and used for DNA fingerprinting. Using synthetic oligonucleotide probes such as $(GATA)_4$, hybridization conditions can be adjusted in order to eliminate even single base mismatch (24). Compared to the satellite DNAs, GATA/GACA sequences are not clustered as long uninterrupted stretches several kilobases long but are interspersed with all classes of single copy and repetitive sequences. $(GATA/GACA)_n$ repeats are found to cross hybridize with diverse eukaryotes and are highly polymorphic.

1.3.2 $(CA/GT)_n$

$(CA/GT)_n$ appear to be the most predominant sample repeats among animals including humans (25,26) and are shown to be highly polymorphic in diverse eukaryotes from yeasts to humans (27,28,29) including plants (30). For

example, R18.1, a bovine genomic clone with poly GT stretches, could generate highly polymorphic DNA fingerprint patterns in farm animals & ornamental plants (14). Simple sequence repeat markers especially the dinucleotide repeats have contributed immensely in the construction of physical and genetic maps in humans (31,32,33,34). Based on the distribution of $(TG)_n$ repeats, there has been much speculation about the functional role of these sequences. $(CA/TG)_n$ sequences have been shown to favour left handed Z-DNA formation under physiological conditions (35,36,37) apart from a role in regulation of gene expression and recombination (37,38).

1.3.3 $(CAC)_5$

Compared to other simple repeats, $(CAC)_5$ probe has revealed the most informative DNA fingerprints in man. Later it has been used extensively for generation of locus specific probes (39), screening of hybridomas and cell lines (40), detection of mutations in tumors (41) and clonal diversity in fishes (42).

1.3.4 M13 probe

It was a very interesting finding that two clusters of 15 base pair core minisatellite sequence in the protein III gene of wild type bacteriophage M13 could generate DNA fingerprints in human and animal DNA (43,10). These DNA fingerprints could be generated only in the absence of herring sperm DNA in the hybridization mixture, since the minisatellites in the fish DNA compete with M13 probe. M13 probe has been considered as a universal DNA fingerprinting probe and was used to generate population specific DNA fingerprints in a neotropical pseudoscorpion (44) and to assess genetic variability in the solitary bee Megachile rotundata (45), fruitfly, Ceratitis capitata (46) and Koala (47).

1.3.5 pV47

Many other minisatellite related sequences which were highly informative have been reported. A novel human minisatelite probe pV47 isolated from a human chromosome-16 specific library could generate individual specific DNA fingerprints in humans and cattle and was found to be more informative than M13 probe (48,49).

1.3.6 *CHI* sequence

Since, most of the minisatellites discovered so far have homology with the *chi* sequence, a signal for rec BCD dependent recombination in *E.coli*, an oligonucleotide with homology to the *chi* sequence could generate informative DNA fingerprints in humans (50). Further, a Drosophila minisatellite was shown to contain multiple chi sequences useful for DNA fingerprinting (51).

1.3.7 Telomeric repeats

Eukaryotic chromosomes begin and end with a stretch of repetitive DNA. The telomeric repeat probe from yeast was found to hybridize to a wide range of eukaryotes (52). Recently, telomeric repeats have been shown to detect polymorphisms in tomato and maize (53,54) and used to define the ends of chromosomes to confirm the presence of RFLP markers at the telomeres.

1.3.8 Other minisatellite related probes

Repeats in the 'Per' gene of *Drosophila* involved in the expression of biological rhythms which code for 50 alternating threonines & glycines were found to be conserved in a variety of vertebrates and were polymorphic in man, mole rat and bandicoot (55,56,57,58). A highly polymorphic hypervariable region 3' to the human alpha-globin gene cluster consisting of 17bp repeats was identified by Jarman (7). Similarly, a highly polymorphic AT-rich tandem repeat was

identified in the 3' hypervariable region of apolipoprotein B (59).

1.4 PCR mediated microsatellite amplifications

There are many reports regarding the use of PCR technology for amplification of locus specific microsatellites. For this purpose, initially the genomic library of the specific genotype needs to be screened with micro or minisatellite probes to isolate clones from a homologous system which often increases the polymorphism detected by the probe (46). By sequencing the isolated clones, primers flanking the VNTR loci which are unique can be synthesized and these repeat stretches can be amplified using polymerase chain reaction (PCR) as described by Weber and May (60). The resulting PCR products vary in electrophoretic mobilities differing in the number of core units in the VNTR allele present. PCR amplified microsatellites have several advantages due to their highly informative nature, abundance, codominant inheritance and specificity.

1.5 Micro and minisatellites in plants

It was very exciting when Jeffreys probe 33.6 could detect DNA fingerprints in cultivated rice which were cultivar specific and appeared to be inherited in a Mendelian fashion (61). Later, Zhao & Kochert (62) have demonstrated for the first time in plants the presence, polymorphic nature and stable inheritance of a $(GGC)_n$ microsatellite locus in rice. A RFLP species specific probe for A genome of rice at high stringency which could generate DNA fingerprints at low stringency was isolated by screening a rice genomic library with the human minisatellite sequence 33.6. This probe was shown to have a distinct advantage over the genome specific probes currently available which do not detect RFLPs inherited in a Mendelian manner which makes it unsuitable for placing them on genetic maps (63). When a DNA probe consisting largely of

minisatellites was amplified by PCR using the plasmid DNA containing rice insert as a template, it could generate fingerprint profiles which could differentiate rice cultivars (64). In this study, DraI was found to be the enzyme of choice. Recently, $(TG)_n$ repeats have also been shown to occur every 480 kb and are dispersed in rice genome (65). Extensive microsatellite polymorphisms in indica and japonica rices were shown with 8 mapped SSR markers containing dinucleotide repeats (66). Microsatellites in specific gene sequences were amplified from maize inbred lines and were found to be more informative than RFLPs. Based on these results, microsatellites have been suggested to serve as a valuable tool in maize genome mapping (67). In cotton $(CA)_n$, $(GA)_n$ and $(AT)_n$ microsatellites were shown to be abundant and more polymorphic than RFLP markers (68). Compound microsatellite repeats were shown to be present in the same subclone in the pine genome (69). Compared to abundant $(CA/GT)_n$ repeats in humans (70), $(AC)_n$ & $(AG)_n$ repeats were shown to be present in tropical trees (71) & $(AT)_n$ repeats were abundant in soybean & distributed throughout the genome (72). Further, Yu *et al.* (73) have shown the utility of microsatellite marker assisted screening of germplasm for viral resistance sources in soybean. Recently, Thomas and Scott (74) amplified many microsatellite repeats in grapevine cultivars that detected high level of genetic variation which indicates their utility in mapping and varietal identification. At specific loci, some alleles were found to be more common than others in grapevine cultivars. Microsatellites have been shown to serve as ideal molecular markers in wheat and are being used for construction of a high density linkage map (75). Recently, individual specific DNA fingerprints of *Hevea brasiliensis* clones were generated with the human minisatellite probe 33.6 (76).

TH-1018

1.6 Mechanism for generation of variability

The variability generated by minisatellites and microsatellites have been hypothesized to be a result of one or several processes such as replication slippage and unequal crossing over between sister chromatids (77,78). Mis-alignments in pairing between sister chromatids often result in duplication or deletion of variant repeat units in mini and microsatellites. The recombinator hypothesis proposed for minisatellite generation involves the unwinding of the double stranded DNA by a recombinase enzyme at *chi*-like core sequences creating a nick, DNA repair synthesis, ligation and replication. This results in duplication of the core repeat (1,79). In yeast $(CA)_n$ repeats at telomeres are generated by sequence specific polymerases. Microsatellites are thought to be generated mainly by slippage during DNA replication when the parental and newly synthesized daughter strands unwind and later on mispair due to sequence similarity. The filling of gaps arising due to mispair results in amplification of these simple repeats. The detection of a novel minisatellite specific DNA binding protein (80) and the evidence for exchange between homologous minisatellite alleles (81) suggests that minisatellite VNTRs have a functional role in recombination.

1.7 Transcriptional activity

Although simple repetitive DNA sequences are generally considered to be transcriptionally inactive, there are a few reports of poly $(A)^+$ transcripts that include sequences homologous to the simple tandem repeats. Thus, $(GATA)_n$ and $(GACA)_n$ homologous sequences in mouse (82,83,84) and $(CAC)_n$ homologous sequences in dogs (85) have been reported amongst the transcripts. The significance of expression of simple repeats at the RNA level is not known. Attempts to identify functional protein products encoded by these simple repeat containing

transcripts have not been successful (82). If these sequences are involved in some important functions, then the patterns of expression should change during the course of development in a tissue specific manner (9).

1.8 Methylation studies

Methylation of cytosine residues is common in eukaryotes and has been suggested to be involved in various functions such as regulation of gene expression (86) and tissue specific gene expression (87,88,89). Eventhough, much of our knowledge is based on studies involving animal systems, Gruenbaum *et al.* (90) have shown the target sites of C-methylation in plants. Although, majority of the methylations exist as 5-methyl cytosine in plants (91), adenine methylation has been reported in few cases (92,93). Methylation of hypervariable loci has been shown to be associated with X-chromosome inactivation (94). Polymorphic expression of fragile X syndrome in humans has been linked to cytosine methylation of a single CpG island very near the fragile site (95). DNA fingerprints of humans with a minisatellite probe revealed higher number of identical DNA methylation sites in all tissues compared to limited number of tissue specific methylations (96). Mini and microsatellite probes are ideal to study the development and differentiation related DNA methylation changes as these sequences are dispersed in the genome thereby spanning many loci simultaneously. Cytosine methylation in the vicinity of simple repeats in plants was reported first in Brassica (19). Methylation of cytosine in CG dinucleotide dramatically favours Z-DNA formation. This might be relevant in the regulation of gene expression by methylation (97,98).

1.9 Applications

The ability of micro and minisatellites to detect many hypervariable loci simultaneously and their distribution throughout the genome has made them markers of choice for application to problems in diverse fields some of which are stated below :

Forensic science - By comparing the DNA fingerprints of suspects and the samples collected from the murder site, actual criminal can be identified. It can also be used for paternity testing to exclude nonparental individuals. Similarly, victim and rapist's DNA recovered from samples available at the site of crime can be used to identify the true criminal by using DNA fingerprinting (99).

Sex determination - GATA/GACA repeats were first shown to be sex specifically arranged in *Drosophila*, mouse and snake (83,23). Several hypothesis have been proposed for sex determination involving regulation by simple sequence repeats either by their presence or absence or by copy number variation (100). The hypothesis that GATA/GACA repeats are involved in primary sex determination was supported by the finding that they are linked to a testis determining gene in mouse (84). Simple repeats have been shown to be associated with the differentiation of sex chromosomes in fishes and phorid fly and could generate sex specific hybridization patterns (101,102,103). Jeffreys probe 33.15 could detect sex specific fragments in mice and birds (104,105) and sex linked minisatellites in birds (106). pV47 could generate sex specific hybridization patterns in brown skua, a sea bird (107).

Genetic disease diagnosis - Several neuropsychiatric disorders in humans were recently shown to be caused by expansion of triplet repeats (108) such as $(CCG)_n$ in the case of fragile X syndrome and amplification of $(AGC)_n$ in the case of

myotonic dystrophy and X-linked spinal and bulbar muscular atrophy (Kennedy's disease). In the case of fragile X syndrome, normal individuals have 6 to 60 copies, carriers have 60 to 200 copies and affected individuals have more than 200 copies of the repeat (109,110,111,112). This information is now being used to predict disease occurrence in prenatal diagnosis. The association of trinucleotide repeats with disease causing genes has opened up new opportunities of isolating novel genes associated with genetic disorders (113,114,115).

Individual identification - To reveal genetic variation on an extra ordinarily fine scale, allelic variability in repeat copy number within a minisatellite was analyzed by minisatellite variant repeat mapping (MVR) at the hypervariable locus D1S8 (116,81). MVR mapping involves presence oftwo kinds of repeat units that differ by a single base substitution which creates (a-type) or deletes (t-type) *HaeIII* restriction site. PCR amplification of the allele followed by partial digestion with *HaeIII* provides an unambiguous binary code and reveals high level of variation.

Breeding - DNA fingerprinting can be of great value in monitoring and aiding gene introgression from wild species into elite cultivars by selecting genotypes having maximum similarity to recipient line and minimal similarity to donor line. Genotypes with good combining ability can be identified without evaluation of F1 hybrids in the field to predict heterotic combinations using DNA fingerprinting. Zhang *et al* (117) reported that several microsatellite marker loci detected highly significant effects on yield in elite hybrid rice which may be useful for predicting heterosis.

Patent rights - DNA fingerprinting can be of great value in characterization and evaluation of genetic resources as well as varietal identification. This has great value in protection of breeder's rights/patent rights.

REFERENCES

[1] Jeffreys AJ, Wilson V,Thein SL (1985) Nature 314 : 67

[2] Nybom H, Hall HK (1991) Euphytica 53 : 107

[3] Tautz D, Trick M, Dover GA (1986) Nature 322 : 652

[4] Tautz D, (1989) Nucleic Acids Res 17 : 6463

[5] Wyman AR, White R (1980) Proc Natl Acad Sci USA 77 : 6754

[6] Bell GI, Selby MJ, Rutter WJ (1982) Nature 295 : 31

[7] Jarman AP, Nicholls RD, Weatherall DJ, Clegg JB, Higgs DR (1986) EMBO J 5 : 1857

[8] Capon DJ, Chen EY, Levinson AD, Seebung PH, Goeddel DV (1983) Nature 302 : 33

[9] Epplen JT (1988) J Hered 79 : 409

[10] Ryskov AP, Jincharadze AG, Prosnyak MI, Ivanov PL, Limborska SA (1988) FEBS Lett 233 : 388

[11] Rogstad SH, Patton II JC, Schaal BA (1988) Proc Natl Acad Sci USA 85 : 9176

[12] Zimmerman PA, Lang-Unnasch N, Cullis CA (1989) Genome 32 : 824

[13] Nybom H, Rogstad SH, Schaal BA (1990) Theor Appl Genet 79 : 153

[14] Tzuri G, Hillel J, Lavi U, Haberfeld A, Vainstein A (1991) Plant Science 76 : 91

[15] Beyermann B, Nurnberg P, Weihe A, Meixner M, Epplen JT, Borner T (1992) Theor Appl Genet 83 : 691

[16] Weising K, Kaemmer D, Weigand F, Epplen JT, Kahl G (1992) Genome 35 : 436

[17] Kaemmer D, Afza R, Weising K, Kahl G, Novak FJ (1992) Bio/Technology 10 : 1030

[18] Vosman B, Avens P, Rus-Kortekaas W, Smulder MJM (1982) Theor Appl Genet 85 : 239

[19] Poulsen GB, Kahl G, Weising K (1993) Theor Appl Genet 85 : 994

[20] Schmidt T, Boblenz K, Metzlaff M, Kaemmer D, Weising K, Kahl G (1993) Theor Appl Genet 85 : 653

[21] Thomas MR, Matsumoto S, Cain P, Scott NS (1993) Theor Appl Genet 86 : 173

[22] Singh L, Purdom IF, Jones KW (1980) Chromosoma 79 : 137

[23] Singh L, Phillips C, Jones KW (1984) Cell 36 : 111

[24] Schaffer R, Zischler H, Birsner U, Becker A, Epplen JT (1988) Electrophoresis 9 : 369

[25] Hamada H, Petino MG, Kakunaga T (1982) Proc Natl Acad Sci USA 79 : 6465

[26] Jeang KT, Hayward GS (1983) Mol Cell Biol 3 : 1389

[27] Walmsley RM, Wilkinson BM, Kong TH (1989) Bio/Technology 7 : 1168

[28] Kashi Y, Tickochinsky Y, Genislav E, Iraqi F, Nave A, Beckmann JS,Gruenbaum Y, Soller M (1990) Nucleic Acids Res 18 : 1129

[29] Ellegren H, Andersson L, Johonsson M, Sandberg K (1992) Anim Genet 23 : 1

[30] Lonn M, Tegelstrom H, Prentice HC (1992) Nucleic Acids Res 20 : 1153

[31] Stallings RL, Torney DC, Hildebrand CE, Longmire JL, Deaven LL,Jett JH, Doggett NA, Moyzis RK (1990) Proc Natl Acad Sci USA 87 : 6218

[32] Todd S, Sherman SL, Naylor SL (1993) Genomics 16 : 612

[33] Mills KA, Buetow KH, Xu Y, Weber JL, Altherr MR, Wasmuth JJ, Murray TC(1992) Genomics 14 : 209

[34] Melis R, Bradley P, Elsner T, Robertson M, Lawrence E, Gerken S, Albertsen H, White R (1993) Genomics 16 : 56

[35] Haniford DB, Pulleyblank DE (1983) Nature 302 : 632

[36] Rich A, Nordheim A, Wang AH (1984) Annu Rev Biochem 53 : 791

[37] Hamada H, Seidman M, Howard BH, Gorman CM (1984) Mol Cell Biol 4 : 2622

[38] Wahls WP, Wallace LJ, More PD (1990) Cell 60 : 95

[39] Zischler H, Kammerbauer C, Studer R, Grzeschik KH, Epplen JT (1992)Genomics 13 : 983

[40] Hampe J, Nurnberg P, Epplen C, Jahn S, Grunow R, Epplen JT (1992) Hum Antib Hybridomas 3 : 186

[41] Lagoda PJL, Seitz G, Epplen JT, Issinger OG (1989) Hum Genet 84 : 35

[42] Turner BC, Elder JF, Laughlin TF, Davis WP, Taylor DS (1992) Proc Natl Acad Sci USA 89 : 10643

[43] Vassart G, Georges M, Monsieur R, Brocas H, Lequarre AS, Christophe D(1987) Science 235 : 683

[44] Zeh DW, Zeh JA, Coffroth MA, Bermingham E (1992) Heredity 69 : 201

[45] Blanchetot A (1992) Genome 35 : 681

[46] Haymer DS, Mcinnis DO, Arcangeli L (1992) Genome 35 : 529

[47] Taylor AC, Marshell Gralles JA, Murrows ND, Shermin WB (1991) Biochem Genetics 29 : 355

[48] Longmire JL, Kraemer PM, Brown NC, Harde Kopf LC, Deaven LL (1990)Nucleic Acids Res 18 : 1658

[49] Dolf G, Glowatzki ML, Gaillard C (1992) Animal Genetics 23 : 63

[50] Ehtesham NZ, Das A, Hasnain SE (1992) Gene 111 : 261

[51] Jacobson JW, Guo W, Hughes CR (1992) Insect Biochem Molec Biol 22 : 785

[52] Allshire RC, Gosden JR, Cross SH, Cranston G, Rout D, Sugawara N, Szostak JW, Fantes PA, Hastie ND (1988) Nature 322 : 656

[53] Broun P, Ganal MW, Tanksley SD (1992) Proc Natl Acad Sci USA 89 : 1354

[54] Burr B, Burr FA, Matz EC, Soverson JR (1992) Plant Cell 4 : 953

[55] Shin HS, Bargiello TA, Clark BT, Jackson FR, Young MW (1985) Nature 317 : 445

[56] Georges M,Cochaux P, Lequarre AS, Young MW, Vassart G (1987) Nucleic Acids Res 15 : 7193

[57] Benshlomo R, Shin HS, Nevo E (1993) Heredity 70 : 111

[58] Robinson NA, Murray ND, Sherwin WB (1993) Molec Ecology 2 : 195

[59] Knott TJ, Wallis SC, Pease RJ, Powell LM, Scott J (1986) Nucleic Acids Res 14 : 9215

[60] Weber JL, May PE (1989) Am J Hum Genet 44 : 388

[61] Dallas JF (1988) Proc Natl Acad Sci USA 85 : 6831

[62] Zhao X, Kochert G (1993) Plant Molecular Biology 21 : 607

[63] Dallas JF, Mcintyre CL, Gustafson JP (1992) Genome 36 : 445

[64] Winberg BC, Zhou Z, Dallas JF, Mcintyre CL, Gustafson JP (1993) Genome 36 : 978

[65] Wu KS, Tanksley SD (1993) Mol Gen Genet 241 (1-2) : 225

[66] Yang G, Saghai Maroof MA, Zhang Q (1994) Plant Genome II, Second Intl Conference on Plant Genome, SanDiego, California, January 24-27 : P46, pp 26

[67] Senior ML, Heun M (1993) Genome 36 : 884

[68] Zhao X, Lin Y, Paterson AH (1994) Plant Genome II, Second Intl Conference on Plant Genome, SanDiego, California, January 24- 27 : P238, pp 74

[69] Smith DN, Devey ME, Maran GF (1994) Plant Genome II, Second Intl conference on Plant Genome, SanDiego, California, January 24-27 : F4, pp 79

[70] Weber JL (1990) Genomics 7 : 524

[71] Condit R, Hubbell SP (1991) Genome 34 : 66

[72] Akkaya MS, Bhagwat AA, Cregen PB (1992) Genetics 132 : 1131

[73] Yu YG, Saghai Maroof MA, Buss GR (1994) Plant Genome II, Second Intl Conference on Plant Genome, SanDiego, California, January 24-27 : P 233, pp 73

[74] Thomas MR, Scott NS (1993) Theor Appl Genet 86 : 985

[75] Devos K, Bryan G, Collins A, Smith J, Stephenson P, Gale M (1994) Plant Genome, Second Intl. Conference on Plant Genome, SanDiego, California, January 24-27 : P46 pp 26

[76] Besse P, Lebrun P, Seguin M, Lanaud C (1993) Heredity 70 : 237

[77] Stephan W (1989) Mol Biol Evol 6 : 198

[78] Harding RM, Boyce AJ, Clegg JB (1992) Genetics 132 : 847

[79] Steinmetz M,Uematsu Y, Lindahl KF (1987) Trend In Genetics 3 : 7

[80] Collick A, Jeffreys AJ (1990) Nucleic Acids Res 18 : 625

[81] Jeffreys AJ, MacLeod A, Tamaki K, Neil DL, Monckton DG (1991) Nature 354 : 204

[82] Schaffer R, Boltz E, Becker A, Bartels F, Epplen JT (1986) Chromosoma 93 : 496

[83] Epplen JT, McCarrey JR, Sutlow S, Ohno S (1982) Proc Natl Acad Sci USA 79 : 3798

[84] Singh L, Jones KW (1982) Cell 28 : 205

[85] Hershfeld B, Inouye L, Choder G, Ripps H, Aguirre G (1991) J Heredity 82 : 251

[86] Boyes J, Bird A (1992) EMBO J 11 : 327

[87] Ngernprasirtsiri J, Kobayashi H, Akazawa T (1988) Proc Natl Acad Sci USA 85 : 4750

[88] Razin A, Szyf M, Kafri T, Roll M, Giloh H, Scarpa S, Carotti D, Cantoni GL (1986) Proc Natl Acad Sci USA 83 : 2827

[89] Razin A, Cedar H (1991) Microbiol Rev 55 : 451

[90] Gruenbaum Y, Naveh-Many T, Ceder H, Razin A (1981) Nature 292 : 860

[91] Messeguer R, Ganal MW, Steffens JC, Tanksley SD (1991) Plant Molecular Biology 16 : 753

[92] Pintor-Toro JA (1987) Biochem Biophys Res Comm 147 : 1082

[93] Dhar MS, Pethe VV, Gupta VS, Ranjekar PK (1990) Theor Appl Genet 80 : 402

[94] Boyd Y, Fraser NJ (1990) Genomics 7 : 182

[95] Oberle I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A, Boue J, Bertheas MF, Mandel JL (1991) Science 252 : 1097

[96] Barneman ZN, Raeymaekers P, Vanhoof G, Bockstaele DV, Ramael M, Peetermans ME, Vandenberghe A, Broeckhoven CV (1990) FEBS Lett 255 : 226

[97] Wang AHJ, Quigley GJ, Kolpak FJ, Crawford JL, Van Boom JH, Vander Marel G, Rich A (1979) Nature 282 : 680

[98] Rich A, Nordheim A, Wang AHJ (1984) Ann Rev Biochem 53 : 791

[99] Gill P, Jeffreys AJ, Weatherall DJ (1985) Nature 318 : 577

[100] Chandra HS (1985) Proc Natl Acad Sci USA 82 : 1165

[101] Nanda I, Feichtinger W, Schmid M, Schroder JH, Zischler H, Epplen JT (1990) J Mol Evol 30 : 456

[102] Nanda I, Schartl M, Feichtinger W, Epplen JT,Schmid M (1992) Chromosoma 101 : 301

[103] Willhoeft U, Traut W (1990) Chromosoma 99 : 237

[104] Jeffreys AJ, Wilson V, Kelly R, Taylor BA, Bulfield G (1987) Nucleic Acids Res 15 : 2823

[105] Rabenbold PP, Rabenbold KN, Piper WH, Haydock J, Zack S (1990) Nature 348 : 538

[106] Robenbold PP, Piper WH, Decker MD, Minchella DJ (1991) Genome 34 : 489

[107] Millar CD, Lambert DM, Bellamy AR, Stapleton PM, Young EC (1992) J Hered 83 : 350

[108] Caskey CT, Pizzuti A, Fu YH, Fenwick RG, Nelson DL (1992) Science 256 : 784

[109] Hubert R, Weber JL, Schmitt K, Zhang L, Arnheim N (1992) Am J Hum Genet 51 : 985

[110] Cumming AM, Armstrong JG, Pendry K, Burn AM, Wensley RT (1992) Hum Genet 89 : 194

[111] Nancarrow DJ, Walker GJ, Weber JL, Walters MK, Palmer JM, Hayward NK (1992) Genomics 14 : 939

[112] Richards B, Horn GT, Merrill JJ, Klinger KW (1992) Genomics 9 : 235

[113] Ross CA, Mcinnis MG, Margolis RL, Li SH (1993) Trends In Neurological Sciences 16 : 254

[114] Peltomaki P, Aaltonen LA, Sistonen P, Pylkkanen L, Mecklin JP, Jarvinen H, Green JS, Jass JR, Weber JL, Leach FS, Peterson GM, Hamilton SR, Chapelle A, Vogelstein B (1993) Science 260 : 810

[115] Aaltonean LE, Peltomaki P, Leach FS, Sistonen P, Pylkkanen L, Mecklin JP, Jarvinen H, Powell SM, Jen J, Hamilton SR, Petersen GM, Kinzler KW, Vogelstein B, Chapelle A (1993) Science 260 : 812

[116] Jeffreys AJ, Neumann R, Wilson V (1990) Cell 60 : 473

[117] Zhang Q, Gao Y, Yang S, Ragab R, Saghai Maroof M, Li Z (1994) Plant Genome II, Second Intl Conference on Plant Genome, SanDiego, California, January 24-27 : P166, pp 56

CHAPTER 2

MATERIALS AND METHODS

2.1 Plant material and chemicals

2.2 Media preparation

2.3 Vector and host systems for subcloning

2.4 Isolation of DNA and its digestion with restriction enzymes

2.5 Agarose gel electrophoresis

2.6 Drying, denaturation and neutralization of agarose gels

2.7 Southern blotting

2.8 DNA probes, labelling and purification

2.9 Southern hybridization

2.10 Hybridization of dry gels

2.11 Subcloning of clone $1E_6$

2.12 Colony hybridization

2.13 DNA sequencing

2.14 PCR amplification

2.15 Isolation of total RNA and Poly A^+ RNA

2.16 Dot blot and Northern blot hybridizations

2.17 Analysis of DNA fingerprints

2.1 Plant material and chemicals

Seeds of different cultivars of *O.sativa* & different accessions of wild species of rice as shown in Table 1 were obtained from Directorate of Rice Research (DRR), Hyderabad; Central Rice Research Institute (CRRI), Cuttack; Agricultural Research Station, Vadgaon and Maharashtra Association for cultivation of Science (MACS), Pune. All the rice genotypes were grown in the glasshouse at National Chemical Laboratory (NCL), Pune. Leaf material was harvested, frozen in liquid nitrogen and stored at -70°C.

All chemicals used were of analytical reagent (AR) grade and were obtained from various sources such as Qualigens, British Drug House (BDH), Sarabhai Chemicals, SISCO Research Laboratories and E. Merck, India. Fine chemicals including SDS (Sodium dodecyl sulphate), ampicillin, RNase A (DNase free), X-gal (5-Bromo-4-chloro-3-indolyl-β-D-galactose), IPTG (Isopropyl thiogalactoside) and formamide were obtained from Sigma Chemicals Co., USA. Molecular weight markers such as lambda *HindIII* and φX-173-*HaeIII* digests were from Biogenie Ltd., India and 1kb marker from Bethesda Research Laboratories, USA. All the restriction and modifying enzymes used were from Boehringer Mannheim (Germany), New England Biolabs (NEB) USA, Bethesda Research Laboratories (BRL), USA and Biogenie Ltd., India. Transfer membranes used for blotting of DNA were Hybond N, Hybond N+ (Amersham, U.K.) And Nytran (S & S). Radiolabelled α-P-32-dATP, α-P-32-dCTP and γ-P-32-ATP and multiprime labeling and end labeling kits were obtained from Bhabha Atomic Research Centre (BARC), Bombay,India.

TABLE 1
Rice genotypes used in the present study

Rice genotypes	Type	Genome	Sources
O.sativa Indrayani	Indica	AA	ARS, Vadgaon
Ambemohr	Indica	AA	ARS, Vadgaon
Basmati-370	Indica	AA	CRRI, Cuttack
Tulasi	Indica	AA	DRR, Hyderabad
Pranava	Indica	AA	DRR, Hyderabad
Tambada rambhog	Indica	AA	MACS, Pune
Khadkya	Indica	AA	MACS, Pune
Vargalsal	Indica	AA	MACS, Pune
Adt-27	Indica-Japonica derivative	AA	CRRI, Cuttack
Hakkoda	Japonica	AA	DRR, Hyderabad
Norin-49	Japonica	AA	DRR, Hyderabad
Fujisaka	Japonica	AA	DRR, Hyderabad
Taichung-65	Japonica	AA	CRRI, Cuttack
O.rufipogon	Wild	AA	DRR, Hyderabad
O.nivara	Wild	AA	DRR, Hyderabad
O.punctata	Wild	BBCC	DRR, Hyderabad
O.minuta	Wild	BBCC	DRR, Hyderabad
O.malampuzhansis	Wild	BBCC	DRR, Hyderabad
O.officinalis	Wild	CC	DRR, Hyderabad
O.alta	Wild	CCDD	DRR, Hyderabad
O.granulata	Wild	-	CRRI, Cuttack
O.australiensis	Wild	EE	CRRI, Cuttack
O.branchyantha	Wild	FF	CRRI, Cuttack
O.latifolia	Wild	CCDD	CRRI, Cuttack
O.eichengeri	Wild	CC	CRRI, Cuttack

2.2 Media preparation

Media were prepared in deionized water and were autoclaved at 15psi (pounds per square inch) for 20 minutes. Heat labile components like antibiotics were filter sterilized through autoclaved millipore membranes and were added freshly to the cooled media. For the preparation of solid media, agar was included before autoclaving. When the media was removed from the autoclave, it was swirled gently to distribute the melted agar evenly throughout the solution. The medium was then cooled to 50°C, before adding thermolabile substances like antibiotics. The concentration of the agar used was 2% throughout.

Composition of the medium and concentration of antibiotics used are given below:

LB medium (Luria-Bertani medium) [Per Liter]

Tryptone	:	10gms
Sodium chloride	:	5gms
Yeast extract	:	5gms
pH	:	7.2-7.5, adjusted with 1N NaOH

Plating agar for plates [Per liter]

Tryptone	:	10gms
Sodium chloride	:	5gms
Yeast extract	:	5gms
Agar (bacto-agar)	:	20gms
pH	:	7.2-7.5, adjusted with 1N NaOH

Antibiotics

Ampicillin trihydrate was dissolved in a stock concentration of 100 mg/ml in water. Effective concentration of 100 μg/ml was used for both liquid and solid media. Ampicillin was included in the medium when *E.coli* cells transformed with the plasmids were grown, since the plasmids used in our studies had ampicillin resistance gene.

2.3 Vector and host system for subcloning

It is essential to subclone a specific DNA fragment of interest (in this case clone $1E_6$) when only a part of the DNA fragment is to be characterized. It also helps in sequencing.

The partial genomic library of rice *O.sativa* cv.Malkolam was made in the plasmid vector pUC18 and *E.coli* host DH5α and the details regarding the genotypes, selection markers, size and restriction enzyme sites are given below.

pUC18

The pUC plasmids (Eg. pUC18, pUC19, pUC118, pUC119) are 2.7kbp long and are extremely versatile vectors consisting of an array of unique cloning sites for enzymes such as *EcoRI, BamHI, HindIII, PstI* and *SmaI*, a modified ampicillin resistance gene and an origin of replication derived from pBR322 ligated to a portion of the lac Z gene of *E.coli* (1,2). The pUC plasmids are small and can be easily introduced into *E.coli* cells by transformation even when they contain large DNA inserts. Restriction mapping of recombinant molecules is facilitated by the small size of the vector. Being high copy number plasmids, yields of DNA preparations are high. Moreover, the screening of the

recombinants is easy, wherein the non-recombinants produce blue colonies and recombinants produce white colonies on ampicillin plates containing X-gal (5-Bromo-4-chloro-3- indolyl-β-D-galactoside).

***E.coli* DH5α**

E.coli DH5α is a recombination deficient suppressing strain used for plating and growth of plasmids and cosmids. Alpha complementation with the amino-terminus of β-gal encoded in pUC vectors permits the production of lac Z gene product.

Growth, culturing and storing of the host *E.coli* DH5α

E.coli DH5α cells were inoculated from a solid medium (single colony on a plate) into 10ml liquid LB medium and were allowed to grow at 37°C overnight with shaking (150-200 rpm). This freshly prepared culture was used each time for further inoculation. Cells were stored in stab vials at 4°C or as frozen glycerol suspension (15%) at -70°C.

2.4 Isolation of DNA and its digestion with restriction enzymes

2.4.1. Isolation of high molecular weight DNA from rice

DNA extraction from higher plants has been described by Shure *et al.* (3) and Dellaporta *et al* (4). The major problem in DNA isolated from higher plants is the presence of enzyme inhibiting polysacharides which can be removed by expensive and time consuming cesium chloride density gradient technique. However, Cetyltriethyl ammonium bromide (CTAB) separates polysacharides from nucleic acids based on their differential solubilities. Absence of contaminating material in DNA and digestibility with restriction enzymes have shown that DNA isolation using CTAB was the method of choice. DNA was extracted from leaves of various cultivated and wild rice genotypes as well as root and seedling

of cultivar Indrayani as described by Rogers & Bendich (5). This procedure essentially employs CTAB to disrupt the cell membrane and release DNA into extraction buffer. This DNA is protected from endogenous nucleases by EDTA which chelates magnesium ions which act as cofactors for most nucleases. Chloroform denatures and separates proteins from DNA. High concentrations of sodium chloride in high salt TE is mandatory for DNA to remain in sodium form and DNA is selectively precipitated leaving behind CTAB. 70% ethanol removes residual CTAB and salt. It is essential to remove RNA which is also precipitated along with DNA and may interfere with spectrophotometric estimation of DNA and restriction endonuclease digestions.

Frozen leaf tissue was ground to a fine powder in liquid nitrogen. 30ml of 2X CTAB buffer (2% CTAB, 100mM Tris-HCl, pH 8.0, 20mM EDTA, pH 8.0, 1.4M NaCl, 1% Polyvinyl pyrrolidone) was added to 10g frozen tissue and the suspension was incubated at 60°C for 15 minutes. Equal volume of chloroform:isoamyl alcohol (24:1) mixture was added and mixed thoroughly to form an emulsion and centrifuged for 10 minutes at 10,000 rpm in a SS34 rotor. The supernatant was transferred to a fresh tube containing one tenth volume of 10% CTAB solution (10% CTAB, 0.7M NaCl) and extracted with chloroform. To the supernatant, equal volume of CTAB precipitation buffer (1% CTAB, 50mM Tris-HCl, pH 8.0, 10mM EDTA, pH 8.0) was added, mixed gently and centrifuged at 10,000 rpm. The DNA pellet was dissolved in high salt TE buffer (1M NaCl, 10mM Tris-HCl, pH 8.0, 1mM EDTA, pH 8.0) and was precipitated with twice the volume of chilled ethanol. The DNA precipitate was washed with 70% ethanol and centrifuged and the pellet was redissovled in TE buffer. For removal of RNA, the DNA was incubated at 37°C for 1 hour with RNase, extracted with chloroform, precipitated with ethanol and washed with 70% ethanol. The DNA was pelleted

and traces of ethanol were removed and the pellet was dissolved in TE (Tris-HCl 10mM, EDTA 1mM, pH 8.0). The molecular weights of all the DNAs were more than 23 kbp as determined by comparison with lambda-*HindIII* digest. The DNA concentration was estimated both from gel and absorbance at 260nm.

All DNA preparations were checked for protein and polysaccharide contamination spectrophotometrically. Only those preparations which exhibited an optical density ratio A280/A260 = 0.45-0.55 and A230/A260 = 0.45 were used for further analysis.

2.4.2. Isolation of plasmid DNA

All plasmid DNA extractions were done by the alkaline lysis procedure described by Birnboim & Doly (6). Bacterial cells are lysed with SDS and NaOH where SDS denatures the bacterial proteins while NaOH denatures the plasmid DNA and chromosomal DNA. Plasmid DNA reanneals rapidly in the presence of potassium acetate. Chromosomal DNA and bacterial proteins precipitate along with SDS which forms a complex with potassium and the total precipitate is removed by centrifugation.

A single recombinant colony from the master plate was incubated in 1ml LB medium containing ampicillin (100μg/ml). This was grown to saturation at 37°C, overnight with shaking at 175 rpm. The culture was spun in 1.5ml eppendorf for 10 minutes to pellet cells and supernatant was discarded. The cell pellet was suspended in 100μl solution I [GTE buffer : 50mM Glucose, 25mM Tris-HCl, (pH 8.0), and 10mM EDTA (pH 8.0)], vortexed and rested at room temperature for 10 minutes. Then 200μl of freshly prepared solution II [1% SDS, 0.2N NaOH] was added to the above cell suspension, mixed well by tapping the tube and rested on ice for 10 minutes. The suspension was neutralized by adding 150μl potassium acetate buffer [To 60ml of 5M potassium acetate, add

11.5ml glacial acetic acid and 28.5ml water] and it was again kept on ice for 10 minutes. The suspension was spun for 10 minutes to pellet out the cell debris and chromosomal DNA. Supernantant was transferred to a fresh tube and 1/10th volume of 3M sodium acetate was added to enable DNA to remain in sodium form and DNA was precipitated with two volumes of chilled ethanol. This procedure can be used to isolate plasmid DNA simultaneously from small cultures (minipreps) as well as from large scale cultures (maxipreps).

2.4.3 Restriction endonuclease digestions of DNA

The main tools used to manipulate DNA in genetic engineering are the restriction endonucleases which recognize short DNA sequences and cleave double stranded DNA at specific sites. Genomic libraries of complex eukaryotic genomes can be created by digesting the DNA with a restriction enzyme and ligating with a suitable vector cut with the same restriction enzyme. Restriction enzymes are extremely useful to get the inserts from the clones of the library, to construct restriction maps and to study the organization of specific gene sequences with the help of Southern hybridizations.

In a typical digestion experiment, 1-2μg of DNA was incubated with a suitable restriction enzyme (5-10U/μg DNA) mostly at 37°C or at a specific temperature required for the enzyme used. The salt concentrations of buffers were according to supplier's recommendations. The enzyme to DNA and enzyme volume to reaction volume ratios were adjusted to avoid non-specific (star) activity of restriction enzymes. Following incubation, the enzyme digestion was stopped by addition of 10X reaction terminating buffer (50% glycerol, 100mM EDTA, 0.25% bromophenol blue) to a final concentration of 1X. Control experiments were performed using commercial lambda DNA digested with different restriction enzymes to check the reaction conditions.

For Southern analysis, 5-10μg of genomic DNA was incubated with 35-70U of suitable restriction enzymes using assay conditions as recommended by the supplier. Double digestions were carried out for mapping the restriction endonuclease sites on the target DNA. For this purpose, plasmid DNA was first digested with an enzyme requiring low salt assay buffer, followed by incubation of plasmid DNA with high salt buffer enzyme. Digestions were allowed to proceed at 37°C and the reaction was terminated in the same way using the 10X termination buffer as described above.

2.5 Agarose gel electrophoresis

Agarose is a linear polymer whose basic structure is composed of D-galactose and 3,6-anhydro L-galactose. Electrophoresis through agarose gels is the standard method used to separate, identify and purify DNA fragments. The DNA in the gel can be determined directly by staining the gel with a fluorescent intercalating dye-ethidium bromide, which is visible on UV light. Agarose gels have a lower resolving power than polyacrylamide gels but a greater range of separation. DNAs from 200bp to 20kb in length can be separated on gels of various concentrations. In general agarose gels in TAE or TBE buffers are used to analyze double stranded DNA. To analyze single stranded DNA, agarose gels in alkaline buffers (NaOH/EDTA) are used. The DNA digests were analysed on 1% neutral agarose horizontal slab gels in TAE (40mM Tris acetate, pH 8.0, and 1mM EDTA, pH 8.0) at a constant current of 20-40mA for 12-36 hrs in case of genomic DNAs and 3-5 hrs for plasmid DNAs. Alternatively, TPE buffer was also used (0.09M Tris phosphate and 2mM EDTA). After gel electrophoresis, the gels were stained in dark in ethidium bromide (1μg/ml) and were visualized on

a long wavelength (302nm) UV transilluminator (UV Products, San Gabriel, California, USA) and photographed with a 35mm SLR camera using a red filter on 135/36 B/W ORWO 125 ASA film.

2.6 Drying, denaturation and neutralization of agarose gels

Gel was placed on two sheets of thick filter paper and covered with saran wrap. Using a vacuum gel dryer, the gel was dried without heat for 1 hour or till the gel became completely flat. This gel was heated at 70°C for 1 hour or till it was completely dry and was stored at room temperature till use. Dried gel was soaked in distilled water until saran wrap and filter paper were completely separated. It was denatured in 0.5M NaOH/0.15M NaCl for 30 minutes at room temperature & neutralized with 0.5M Tris-HCl, pH 8.0, 0.15M NaCl for 30 minutes at room temperature and the gel was ready for hybridization.

2.7 Southern blotting

The transfer of electrophoretically separated DNA to a membrane support was as described by Southern (7). DNA was digested with a restriction enzyme and the resulting fragments were separated according to size by agarose gel electrophoresis. The DNA was denatured in situ and transferred to a solid support such as nitrocellulose or nylon membrane. Although nitrocellulose membrane was initially used for transfer of DNA, it is not preferred due to its brittle nature and hydrophobic interaction with DNA. Nucleic acids bind irreversibly and covalently to nylon membrane which can withstand repeated hybridizations. Specific nucleic acid sequences that are immobilized on the membrane can be detected by hybridization of filter bound DNA or RNA with a radioactively labeled probe. In the present work, the transfer of DNA from agarose gels to Hybond (Amersham) membranes or Nytran filters was carried out using vacuum blotting apparatus (Pharmacia LKB Vacu Gene XL). The gel was placed

in the vacuum blotting apparatus on top of the Hybond membrane and a pressure of 55lb. was maintained in the vacuum pump. The gel was treated with 0.25N HCl for 10 minutes for depurination of DNA and the DNA was then denatured in 1.5M NaCl and 0.5N NaOH for 15 minutes. Neutralization of the gel was done in 1M Tris, pH 7.4 and 1.5M NaCl for 15 minutes. The transfer of DNA onto the membrane was carried out for one hour in 20X SSC, after which the membrane was rinsed in 2X SSC, air dried and baked for 2 hours at 80°C.

2.8 DNA probes, labeling and purification

2.8.1 Synthesis and purification of oligonucleotides

Oligonucleotides were synthesized on a gene assembler plus (Pharmacia) and desalted on a NAP-5 (Sephadex) column. An aliquot of the oligonucleotide was 5' end labelled with γ-P-32-ATP and checked on a 20% denaturing polyacrylamide gel. If bands lower than the desired length of the oligomer were observed, the major band was purified by polyacrylamide gel electrophoresis as follows :

50ml of 20% polyacrylamide gel solution was made with 25ml of 40% acrylamide solution (38:2 acrylamide/bis acrylamide), 5ml of 10XTBE, 21g urea, 0.3ml ammonium persulphate (freshly prepared) and 30ul TEMED. The gel solution was poured in a gel casting unit and molten 1% agarose gel was used to seal the precasted glass plates and the polyacrylamide gel solution was poured in it. 1mm comb was inserted and the gel was polymerized for 45 minutes. Wells were washed with 1X TBE after the removal of comb. The gel was run for 15-30 minutes at 290V in 1X TBE. Equal volume of formamide and xylene cyanol/bromophenol blue mixture was added to the oligonucleotide and heated to 55°C for 5 minutes to disrupt secondary structure before loading the samples. The gel was run for 3 hours or until the bromophenol blue reached one end and

was stained with ethidium bromide (1μg/ml). Using the UV transilluminator, the DNA band of interest was cut with a scalpel and the gel slice was cut into many fine pieces in a 1.5ml eppendorf tube. 500ul of TE was added and incubated overnight at 37°C. The tube was spun at 10,000 rpm for 10 minutes and the supernatent containing the purified oligonucleotide was taken out in a fresh tube for further analysis.

2.8.2 Multiprime labeling

The labeled probe when hybridized to unlabeled DNA strand on the filter gives a signal on the autoradiogram depending on sequence homology between the two fragments. The efficiency of such hybridization experiments depend on the specific activity of the radio labeled probe. In the present experiment, DNA cloned in plasmid vector was labeled with α-P-32-dCTP (BARC) by random priming method of Fienberg and Vogelstein (8,9). This method is based on the annealing of mixtures of all the possible hexanucleotides to the DNA to be labeled at different positions on the template followed by extension by Klenow fragment of *E.coli* DNA polymeraseI in the presence of α-P-32-dCTP and other dNTPs to generate radioactive probe uniformly labeled on both the strands. Klenow fragment essentially catalyzes the addition of nucleotides in a 5'-3' direction to the primer annealed to the template resulting in a phosphodiester bond formation between the nucleotides with release of pyrophosphate. Since input DNA serves as a template and remains intact during the reaction, it is possible to label minimal amounts of DNA (10ng) to a high specific activity (10^8cpm/μg DNA).

25-50ng of the probe in a suitable volume of TE (Tris 10mM, EDTA 1mM, pH 8.0) was denatured by boiling for 10 minutes and was immediately chilled on ice. To the denatured DNA fragment, following reagents from BARC multiprime labeling kit were added : 5μl primer, 5μl labeling buffer, 4μl each of

dATP, dGTP and dTTP, 5μl (50μCi) of α-P-32-dCTP (specific activity 3000 Ci/m mole) and 2μl klenow fragment (4U). Final volume was made to 50μl with sterile water. Reaction was carried out at 37°C for 2 hours and terminated by adding 10X stop dye (10% SDS, 125mM EDTA, 0.25% Bromophenol blue) to a final concentration of 1X.

To prevent background or non-specific hybridization, the probe was separated from unincorporated labeled nucleotides by purification using spun column chromatography. A 1.5ml microfuge tube was packed with Sephadex G-50 to the level of 1ml by centrifugation in a Remi table-top microfuge at 1000 rpm. The entire mixture containing labeled probe and free nucleotides was loaded on this exclusion column and spun at 1000 rpm for 10 seconds. 100ul of TE was added on top of the column and again spun at 1000 rpm for 10 seconds. This was repeated thrice since the exclusion volume of the column is around 300-400μl. After spinning, the probe eluted in a column of about 400μl was collected and used for hybridization.

2.8.3 End labeling

Oligonucleotides were 5' end labeled with γ-P-32-ATP using T4 poly-nucleotide kinase which catalyzes transfer of terminal gamma phosphate of ATP to the 5' hydroxyl termini of DNA. 20 picomoles of oligonucleotide, 5μl of 10X kinase buffer (670mM Tris-HCl pH 8.0, 100mM MgCl2, 100mM DTT), 5ul of γ-P-32-ATP, 20 U of T4 polynucleotide kinase (Biogenie or NEB) and sterile water were added to make up volume to 50μl and incubated for 40 minutes at 37°C. End labeling reaction was terminated by adding 5μl of 0.2M EDTA. The labeled oligonucleotide probe was separated from γ-P-32-ATP by ion exchange chromatography on a DE-52 column. Labeled probe was loaded on 200μl of DE-52

column packed in a 1.5ml eppendorf tube. The column was washed with 4ml of TE and 4ml of 0.2N NaCl in TE. 5' end labeled oligonucleotide was finally eluted by 2 washes of 500μl each of 1N NaCl in TE and stored at -20°C till use.

2.9 Southern hybridization

To prevent the non-specific binding of the probe onto membranes, filters were prehybridized for 2 hours. Prehybridization was carried out in heat sealed plastic bags containing the following mixture : 5 X SSPE, 0.1% SDS, 5X Denhardt's solution (0.1% Ficoll, 0.1% Polyvinyl pyrrolidone) and 0.2X BLOTTO (1X BLOTTO is 5% defatted milk powder in water). When hybridization was carried out at 42°C, formamide was added to reduce the Tm (melting temperature) of nucleic acid hybrids. By including formamide in the hybridization solution, the hybridization temperature (Tm-25°C) can be reduced to 30-42°C which has several advantages - the probe is more stable at lower temperatures & there is better retention of non-covalently bound nucleic acids on the filter. After prehybridization, the solution was removed and replaced by hybridization solution. The latter had the same constituents of the prehybridization mixture except that it contained the probe. The purified probe was denatured by boiling it for 10 minutes and was immediately chilled on ice. Hybridization of the probe to the DNAs on the filter was continued overnight with gentle shaking. After overnight hybridization, probe was discarded and filters were washed to remove unbound/unhybridized probe first with 2X SSC, 0.1% SDS for 15 minutes twice at room temperature with gentle shaking. Hot wash was given at hybridization temperature with 2X SSC, 0.1% SDS for 10 minutes to 1 hour depending on the probe. Washing was continued with stringent washes with 1X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS depending on autoradiographic signals.

Moist filters were saran wrapped and exposed to XAR-5 X-ray film (Fuji) for 15 minutes to 7 days depending on the signals at -70°C using Cronex Lightening-Plus intensifier screens.

2.10 Hybridization of dry gels

Neutralized gel was equilibrated in 5XSSPE for 5 minutes & no pre-hybridization was required. Hybridization was carried out in heat sealed plastic bags containing 5X SSPE, 0.1% SDS, 5X Denhardt's solution, 0.2X BLOTTO and the labelled probe at Tm-5°C. Tm was calculated according to the formula Tm = 4°C (number of C or G) + 2°C (number of A or T). After overnight hybridization, solution was stored at 4°C and reused. Gels were initially washed thrice for 30 minutes at room temperature with 5XSSC/0.1% SDS and for 2 minutes at hybridization temperature. The gel was transferred to a filter paper, dried to remove excess liquid, covered with saran wrap and exposed at -70°C. Based on the signal, further stringent washes were given.

2.11 Subcloning of $1E_6$

2.11.1 RE digestion and band elution

The 1.7 kbp clone 1E6 was digested with *HindIII* and was size fractionated on horizontal gels which were run in 1X TAE buffer (0.09M Tris acetate and 2mM EDTA). A 3.5kbp band containing pUC18 and 0.8kbp of insert was transferred from agarose gel onto Whatman DE81 paper (2.3cm, Whatman Ltd, Maidstone, England) (10). DNA was eluted from the paper by incubating it with high salt elution buffer (1.5M NaCl in TE, pH 8.0) at 42°C for 15 minutes and this step was repeated twice. DNA was precipitated using two volumes of ethanol, dissolved in TE (Tris.HCl 10mM, and EDTA 1mM, pH 8.0) and quantified

by band-intensity on ethidium bromide stained gels using a defined amount of standard DNA (lambda/*HindIII* fragments). This DNA fragment was used for ligation.

2.11.2 Ligation

DNA ligases catalyze the formation of phosphodiester bond between 5' phosphate and 3' hydroxy group in duplex DNA. The source of energy for T4 DNA ligase is ATP. The 3.5kbp DNA fragment of $1E_6$ was used for ligation in a reaction volume of 10μl containing 1μl 10X bacteriophage T4 DNA ligase buffer (500mM Tris-HCl, pH 7.4, 100mM DTT, 10mM spermidine, 1mg/ml BSA), 1μl 0.1M ATP, 1ul 0.1M MgCl2 and 2.0 U T4 DNA ligase (BRL). Ligation was carried out at 16°C for 16 hours and the ligation product was directly used to transform *E.coli* DH5α competent cells.

2.11.3 Preparation of competent cells and transformation of *E.coli* DH5α with chimeric plasmid

Preparation of highly 'competent' host cells is a key step towards a good efficiency of transformation. Competent *E.coli* cells were prepared as described by Mandel & Higa (11), Dagert & Ehrlich (12) and they were transformed with 50-100ng of ligated DNA. Along with the ligated samples, 25ng of uncut plasmid DNA (pUC18) and host DH5α cells were used as positive and negative controls, respectively. A single colony of *E.coli* DH5 α was inoculated in 10ml LB and was grown overnight at 37°C with shaking at 175 rpm. One ml of this saturated overnight grown culture was inoculated at 37°C (approximately 2 hours) till the culture obtained an O.D. of 0.45-0.55 (cell density of 5 x 10^7 cells/ml) at 590nm. The culture was immediately chilled on ice for 20 minutes to stop further cell growth. Cells were then pelleted by centrifugation at 10,000 rpm for 10 minutes at 4°C in a Sorvall RC 5B centrifuge. The supernatant was poured off and the

cell pellet was gently suspended in 25ml ice-cold 100mM CaCl2 sterile solution and rested on ice. Exposure to calcium ions enables the cells to take up DNA or to become competent. Cell suspension was again centrifuged for 10 minutes at 10,000 rpm at 4°C. The supernatant was discarded and cell pellet then suspended in 2ml ice-cold 100mM $CaCl_2$ solution. 200μl of competent cells were then dispensed into pre-chilled sterile eppendorf tubes. The ligated and control plasmid DNAs were added to the cell aliquots and rested on ice for 30 minutes. The transformation procedure was essentially as described by Hanahan (13). Cells were then heat shocked at 42°C for two minutes. The heat shock treatment enlarges the cell pore size enabling the DNA in the suspension to enter the cells. Immediately after this step, 800μl of sterile LB medium was added to each tube, and tubes were incubated for 1 hour at 37°C. This helps the host cells to recover and express antibiotic resistance. 100ul aliquots of this transformation culture were spread on LB agar plates (2% agar) containing ampicillin (100μg/ml) and X-gal (40μg/ml). When the plates were dry, they were incubated at 37°C for 12-16 hours. Colonies of *E.coli* cells harboring plasmid with rice insert were selected on the basis of X-gal ampicillin screening procedure. White colonies generally represent the recombinants.

2.12 Colony hybridization

2.12.1 Transfer and replication of colonies onto filters

Hybond-N filters were smoothly placed on the LB agar plates (containing 100μg/ml ampicillin) avoiding any air bubbles between the membrane and the medium. Cultures were inoculated on the membrane from the 96 well microtitre plates of the *PstI* partial rice genomic library using microtitre plate combs. These replica plates with the membrane were incubated at 37°C for 16

hours for growth of individual colonies which appeared on the filters on the next day. Filters were further treated and used for screening the library by colony hybridization.

2.12.2 Screening the library by colony hybridization

In 1975, Grunstein and Hogness (14) described a method for *in situ* lysis of bacterial colonies on nitrocellulose filters and non-covalent attachment of released DNA to the filters. The DNA could then be hybridized to labeled nucleic acid probes. This method remains the most commonly used technique to date to identify individual bacterial colonies carrying plasmids or cosmids that contain sequences of interest and to screen the library for different types of sequences. The first step involved the lysis of bacterial colonies, wherein one piece of whatman paper was saturated with 10% SDS and placed at the bottom of a petriplate and excess liquid was poured off. Hybond-N filter having overnight grown recombinants was peeled off from the plate using blunt-ended forceps and was placed on the SDS impregnated Whatman paper with colony side up for 3 minutes. This treatment limits the diffusion of the plasmid DNA during denaturation and neutralization resulting in a sharper hybridization signal. Filter was then transferred to the second Whatman sheet, saturated with the denaturing solution (0.5N NaOH, 1.5M NaCl), for 5 minutes. This brings about *in situ* denaturation of released plasmid DNA. It was later transferred to the third sheet of Whatman paper, saturated with the neutralizing solution (1.5M NaCl, 0.5M Tris-HCl, pH 8.0) by incubation for 5 minutes. Filter was rested on a dry sheet of Whatman paper with colony side up to dry at room temperature. Thus liberated and denatured DNA on the filter was fixed by baking at 80°C for two hours. These filters were used for hybridization with P-32-labeled probes for screening the library.

2.13 DNA Sequencing

DNA sequencing as described by Sanger *et al* (15) was carried out using sequenase version 2.0 kit (USB) according to manufacturers instructions. This involves the use of a specific primer for extension by a DNA polymerase, chain termination by dideoxynucleotide phosphates and the use of polyacrylamide gels to separate single stranded DNA chains capable of resolving a single nucleotide difference. 2', 3' ddNTPs differ from dNTPs in the absence of a hydroxyl residue at 3' position of deoxyribose which prevents the formation of a phosphodiester bond with the succeeding dNTPs, resulting in termination of the reaction. A small amount of ddNTP is included with the four conventional dNTPs in a reaction mixture alongwith DNA polymerase which results in competition between extension of the chain and infrequent but specific termination. By using four different ddNTPs in four separate reactions, oligonucleotides are generated that terminate at positions occupied by every A, C, G or T in the template strand. The DNA polymerases generally used for sequencing include Klenow fragment of *E.coli* DNA polymeraseI and Sequenase version 2.0 enzyme (a genetically engineered and modified T7 DNA polymerase to eliminate 3' -> 5' exonuclease activity).

Denaturation of template DNA : 2μg of plasmid DNA was dried in an eppendorf tube and dissolved in 40μl denaturation buffer (0.2M NaOH, 0.2mM EDTA, pH. 8.0) and kept at 37°C for 30 minutes. 4μl of 3M sodium acetate, pH. 5.2 was added, followed by 100μl of chilled ethanol and precipitated for 30 minutes at -70°C. The sample was spun at 10,000 rpm for 10 minutes at 4°C. Supernatant was discarded and pellet was washed with ethanol, dried and dissolved in 7μl of sterile water.

Annealing of sequencing primer : 1μl of pUC sequencing or reverse sequencing

primer was added to the template DNA. 2μl of 5X reaction buffer was added and incubated at 65°C for 2 minutes.

Labeling reaction : To the annealed template-primer, 1μl DTT (0.1M), 2μl labeling nucleotide mix, 5μCi of α-S-35 or α-P-32-dATP and 3.25 units of sequenase version 2.0 enzyme were added. The sample was incubated for 2-5 minutes at room temperature.

Chain termination : Four eppendorf tubes were labeled G,A,T,C. 2μl of respective dideoxynucleotide mixture was taken in the four labeled tubes and warmed to 37°C. 3.5μl of the template primer mixture was added to each of the tubes labeled G,A,T,C and mixed and incubated at 37°C for 5 minutes. 4μl of formamide buffer was added to stop the reaction.

Casting and electrophoresis of sequencing gel - The gel was cast as in sect 2.12 except that 6% polyacrylamide gel was made instead of 20% as follows : 40g urea, 12ml of 40% polyacrylamide, 8ml of 10X TBE. Volume was made upto 80ml. The spacers used were of 0.4mm thickness and the gel casting unit was 20cm x 60cm. The solution was degassed and 450μl of 10% freshly prepared ammonium persulphate and 60μl TEMED (N,N,N',N'-tetramethyl ethylene diamine) were added and poured in the gel casting unit held by tape and clamps. The comb was immediately inserted. After 30 minutes, tapes were removed and the gel was clamped to the electrophoresis unit which was maintained at 37°C using circulating water bath. The samples were heated to 80°C for 5 minutes and were loaded and electrophoresed at 2000V till the bromophenol blue ran to one end of the gel. Samples were loaded for the second time and electrophoresis was continued till bromophenol blue touched the lower edge of the gel. Two loadings of the sample will ensure that more sequence can be read. The gel was covered with saran wrap and dried on a vaccum gel dryer at 80°C for 2 hours and exposed to X-ray film at -70°C overnight.

2.14 PCR amplification

The polymerase chain reaction (PCR) is used to amplify a DNA segment that lies between two known sequences (16). Specific primers flanking the target sequence are made to anneal to the template DNA followed by extension and denaturation. After 30 cycles the target sequence is amplified 10^6 fold.

Polymerase chain reactions were performed in a total volume of 25µl with 1.3U of *Taq polymerase*, 1X buffer (Biogenie), 250ng rice genomic DNA, 0.25µM of each primer, 200µM of dGTP, dCTP and dTTP, 150µM of dATP and 0.1µl of α-P-32-dATP (BARC, Bombay). 1 cycle at 95°C for 2 minutes and 30 cycles at 95°C for 1 minute, 55°C for 2 minutes and 72°C for 2 minutes followed by a final extension at 72°C for 2 minutes were performed in a Perkin Elmer Cetus thermocycler. The amplified products were loaded on a 4% (or) 6% sequencing gel depending on the size of the fragment to be analyzed.

2.15 Isolation of total RNA and poly(A)+ RNA

2.15.1 Total RNA Isolation

A good preparation of RNA is an essential step in gene analysis and gene expression studies. The main problems in RNA isolation is the presence of ribonucleases in all the tissues. Therefore, the various steps involved in RNA isolations include inhibition of ribonucleases, deproteinization and separation of RNA from other components of the homogenate. In the protocol used here, aurintricarboxylic acid (ATA) has been used as a nuclease and protein inhibitor. PVP 40 and proteinase K help in removal of carbohydrates and proteins. Deionized water was treated with 0.1% diethylpyrocarbonate (DEPC) and all the glassware was washed with DEPC treated water and autoclaved to inhibit RNase activity. Leaf and dehusked seed tissues were crushed in liquid nitrogen to a fine powder. 10ml RNA isolation buffer (100mM Tris-HCl, pH 8.5, 20mM ATA, 200mM LiCl,

100mM EDTA, 100mM β-mercaptoethanol, 0.5% PVP 40 and 0.5mg proteinase K) was added to a gram of tissue and suspended well. Homogenate was incubated at 37°C for 2 hours and centrifuged at 8,000 rpm for 20 minutes at 4°C. The supernatent was deproteinized with phenol-chloroform : isoamyl alcohol mixture (25:24:1). One third volume of 8M LiCl was added to the aqueous phase and left overnight at 4°C for precipitation facilitating selective separation of RNA from DNA. RNA precipitate was collected by centrifugation and washed with 2M LiCl and resuspended at 55°C for 10 minutes in 2% potassium acetate. This suspension was clarified by centrifugation and RNA was precipitated with 2.5 volumes of absolute ethanol at -20°C overnight. RNA precipitate, obtained after centrifugation, was dissolved in sterile water.

2.15.2 Isolation of poly(A)+ RNA

Messenger RNAs (mRNAs) constitute 1 to 5% of total RNA and are heterogeneous in both size and sequence. Presence of a poly(A) tail in most eukaryotic RNAs enables them to be separated on oligo (dT) - cellulose by affinity chromatography. The mRNAs so prepared virtually code for all the polypeptides synthesized in the cell. Poly $(A)^+$ RNA was isolated as described by Sambrook *et al* (17).

A siliconized 1ml pipette tip was washed with 10ml of 5M NaOH and rinsed with sterile diethylpyrocarbonate (DEPC) treated water. 5ml of dry oligo (dT) cellulose powder was added to 1ml of 0.1M NaOH and the slurry was poured into column and rinsed with 10ml of sterile water. The column was equilibrated with 10 to 20ml loading buffer (0.5M LiCl, 10mM Tis-HCl, pH 7.5, 1mM EDTA, 0.1% SDS) till the pH of the eluant was 7.5. Approximately 2mg total RNA was heated at 55°C for 10 minutes to remove the secondary structure. After the addition of LiCl to a final concentration of 0.5M, the RNA was loaded onto oligo

(dT) cellulose column. The column was washed with 1ml of loading buffer and the eluant was collected and again passed through the column twice to ensure that all the poly(A)$^+$ RNA was bound to the oligo (dT). The column was rinsed with 2ml of middle wash buffer (0.15M LiCl, 10mM Tris-HCl, pH 7.5, 1mM EDTA, 0.1% SDS) followed by elution of poly(A)$^+$ RNA in a fresh tube with 2mM EDTA, 0.1% SDS. One tenth volume of 3M sodium acetate was added to the eluant followed by 2.5 volumes of ethanol and stored at -20°C for few hours. Poly(A)+ RNA was precipitated and collected by centrifugation. The pellet was allowed to air dry and poly(A)+ RNA was suspended in TE buffer.

2.16 Dot blot and Northern blot hybridizations

2.16.1 Dot blot hybridization

In dot blot hybridization, an excess of radiolabeled probe is hybridized to RNA that has been immobilized on a solid support. This gives a quick estimate of the presence or absence of homologous gene transcripts of the probe used. RNA samples were heated to 60°C for 5 minutes in three volumes of the following solution : formamide - 500µl, formaldehyde - 162µl, 10X MOPS buffer - 100µl, and were chilled on ice. Equal volume of cold 20X SSC was added. On a dry filter paper, Hybond-N wetted om 20X SSC was placed and RNA samples were spotted in small aliquots. The membrane was washed in 0.5M NaCl for 5 minutes, dried thoroughly and baked for 2 hours at 80°C.

2.16.2 Northern blotting and hybridization

The size and amount of specific mRNA in preparations of total or poly(A)$^+$ RNA can be determined by Northern blotting and hybridization. RNA is separated according to size by electrophoresis through a denaturing agarose gel and is then transferred to a nylon membrane. The RNA of interest can be detected by hybridization with a radiolabeled probe.

Total RNA (6μl) was incubated at 65°C for five minutes with 12.5μl formamide, 2.5μl 10X MOPS buffer (0.2M 3-[N-morpholino] propane-sulphonic acid, 0.05M sodium acetate pH 7.0, 0.01M EDTA) and 4μl formaldehyde and chilled on ice. 2.5μl of 50% [v/v] glycerol containing 0.1mg/ml bromophenol blue and xylene cyanol was added. 1gm agarose was dissovled in sterile water and cooled to 50°C. 10ml of 10X MOPS buffer and 17ml formaldelyde were added, mixed and poured immediately. The gel was run at 40mA in 1X MOPS running buffer and stained with ethidium bromide, destained and photographed on UV transilluminator. Capillary blotting was done in 20X SSC and filters were baked at 80°C for 2 hours.

The second method to transfer poly(A)$^+$ messages onto a nylon membrane without using oligo (dT) column is use of Hybond-mAP (messenger affinity paper) from Amersham which is poly-U charged and binds to poly(A)$^+$ messages only. To RNA in TE, 0.2 volumes of loading buffer (50% glycerol, 50mM EDTA pH 6.5, 0.1% bromophenol blue and xylene cyanol) was added and the samples were heated to 60°C for five minutes before loading. Electrophoresis was done in 1% agarose in MOPS buffer. The gel was placed in transfer buffer (0.5M NaCl, 10mM Tris-HCl, pH 7.5) for 30 minutes prior to blotting for equilibration. Hyb-mAP was wetted in transfer buffer and capillary blotting was performed as described earlier. Northern hybridizations and post hybridization washings were carried out as described in section 2.9 except that *E.coli* DNA was used instead of 0.2X BLOTTO in the hybridization solution.

2.17 Analysis of DNA fingerprint patterns

Differences in band patterns were scored on the basis of absence or presence of bands. Analysis was carried out on DNA fragments over 2kb in length. A similarity index D expressing the probability that a fragment in one variety

is also found in another for all pairwise comparisons was calculated. ($\overline{X}_D$ represents average similarity index for all pairwise comparisons). Probability that the DNA fingerprints of two varieties of rice will be identical by chance was then estimated as $(XD)^n$ (18).

REFERENCES

[1] Viera J, Messing J (1982) Gene 19 : 259

[2] Yanisch-Perron C, Viera J, Messing J (1985) Gene 33 : 103

[3] Shure M, Wessler S, Federoff N (1983) Cell 35 : 225

[4] Dellaporta SL, Wood J, Hicks JB (1983) Plant Molecular Biology Reporter 1 (4) : 19

[5] Rogers SO, Bendich AJ (1988) Plant Molecular Biology Manual A6 : 1 Ed. Gelvin SB Schilperoort RA, Kluwer Academic Publishers,Dordrecht

[6] Birnboim HC, Doly J (1979) Nucleic Acids Res 7 : 1513

[7] Southern EM (1975) J Mol Biol 98 : 503

[8] Feinberg AP, Vogelstein B (1983) Anal Biochem 132 : 6

[9] Feinberg AP, Vogelstein B (1984) Anal Biochem 137 : 266

[10] Dretzen G, Bellard M, Sassone-Corsi P, Chambon P (1981) Anal Biochem 112 : 295

[11] Mandel M, Higa A (1970) J Mol Biol 53 : 159

[12] Dagert M, Ehrlich SD (1974) Gene 6 : 23

[13] Hanahan D (1983) J Mol Biol 53 : 159

[14] Grunstein M, Hogness D (1975) Proc Natl Acad Sci USA 72 : 3981

[15] Sanger F, Nicklen S, Coulson AR (1977) Proc Natl Acad Sci 74 : 5463

[16] Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA (1986) Science 324 : 163

[17] Sambrook J, Fritsch EF, Maniatis T (1989) In: Molecular Cloning: A Lab Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

[18] Wetton JH, Carter RE, Parkin S, Walters D (1987) Nature 327 : 147

CHAPTER 3

RESULTS

3.1 **DNA fingerprinting in rice with oligonucleotides specific for simple repetitive DNA sequences.**

3.2 **DNA fingerprinting in rice using minisatellites and a cattle probe R18.1**

3.3 **Demonstration of gene transcripts in rice homologous to $(CAC)_5$**

3.4 **Status of DNA methylation in rice using some microsatellites and minisatellites as probes**

3.5 **Identification and characterization of a microsatellite from rice using $(GATA)_4$ probe**

3.1 DNA FINGERPRINTING IN RICE WITH OLIGONUCLEOTIDES SPECIFIC FOR SIMPLE REPETITIVE DNA SEQUENCES

High molecular weight DNAs from several rice cultivars as well as wild rice genotypes were isolated and digested with restriction enzymes such as *AluI, HaeIII, HinfI, TaqI, ClaI, DraI, EcoRI, EcoRV, HindIII* and *XbaI*. Six oligonucleotides - $(GATA)_4$, $(GACA)_4$, $(GGAT)_4$, $(CAC)_5$, $(GAA)_6$ and $(TG)_{10}$ were explored for their potential as DNA fingerprinting probes in rice. These probes were used for both dry gel and blot hybridizations. Dry gels always gave better signals as compared to those obtained with blots (1). All the simple sequence motifs used in the present study were present in rice genome although the level of informativeness and fingerprinting potential varied with different probe and enzyme combinations.

Restriction enzymes with four base pair specifities are widely used in DNA fingerprinting. In the present study, however, only *RsaI* (Fig 1) gave rise to a distinct band pattern. For example, a strong band at 2.5kb was present in lanes 1 and 2 (cultivars Pranava and Fujisaka) while it was absent in wild rice genotypes. Similarly, a strong band was present only in cultivar Pranava at 3kb. All the other tetracutters that were used gave rise to a background smear in low molecular weight range with faint bands which were not scorable (Fig 2). This suggests that these repeat motifs are not present in long uninterrupted stretches with lack of tetracutting restriction enzyme sites. This observation is interesting since much larger and distinct fragments have been obtained with restriction enzymes with a four base pair recognition site in humans and animals suggesting clusterring of tandemly repeated micro/minisatellites (2,3,4).

FIG 1 Hybridization patterns detected by probe $(CAC)_5$ in *RsaI* digests of genomic DNA of rice. Lanes 1 and 2 - *O.sativa* cultivars Pranava & Fujisaka; Lane 3 - *O.rufipogon*; Lane 4 - *O.punctata*; Lane 5 - *O.officinalis* and Lane 6 - *O.alta* accession. Molecular size markers in kb are indicated in the right margin.

FIG 2 Hybridization patterns detected by probe $(GATA)_4$ in *TaqI* digests of genomic DNA of rice. Lanes 1 through 4 are as described in figure 1.

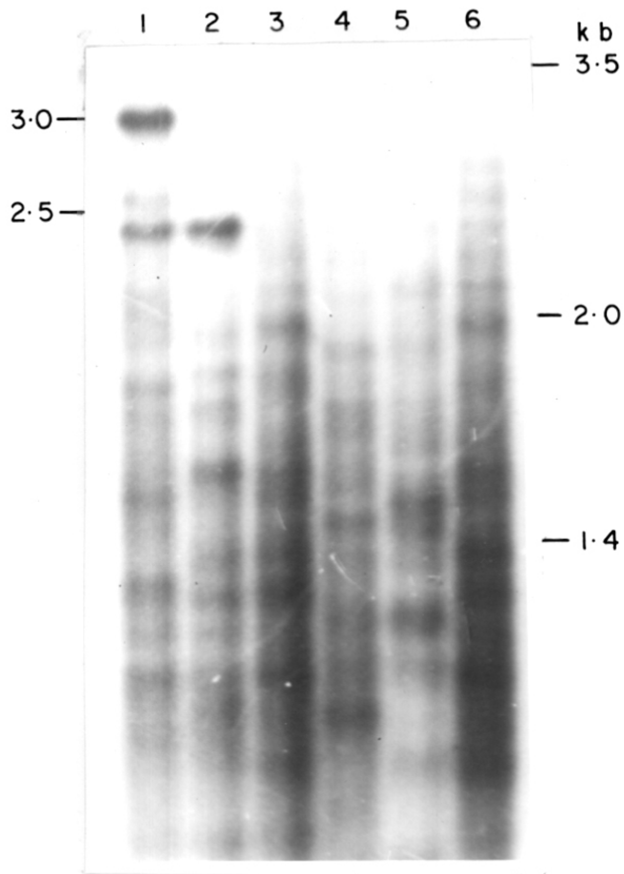


FIG. 1

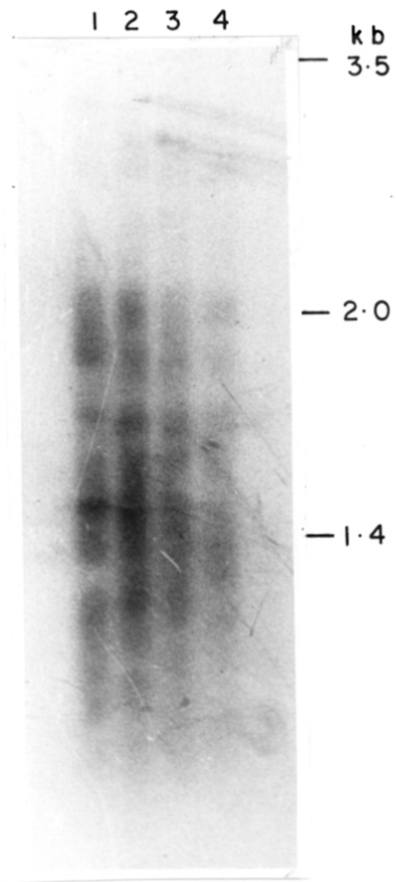


FIG. 2

All the rice genotypes in the present study could be distinguished with the restriction enzymes having six base pair specificities such as *ClaI, DraI* and *HindIII* with most of the probes. Table 2 shows the number of resolvable fragments detected by different oligonucleotide probes in cultivar Basmati-370. In most cases, distinct bands could not be detected in the lower molecular weight region (less than 2kb) and hence gels were run till the 2kb band of lambda-*HindIII* marker ran off the gel. Figures 3A-F represent hybridization patterns detected in *DraI* digests of four cultivars of *O.sativa* and one accession each of *O.rufipogon* (AA) and *O.officinalis* (CC) with $(GATA)_4$, $(CAC)_5$, $(GAA)_6$, $(GGAT)_4$, $(GACA)_4$ and $(TG)_{10}$ respectively. As seen clearly from the above figures, hybridization with $(GATA)_4$, $(CAC)_5$, $(GACA)_4$, $(GAA)_6$ and $(GGAT)_4$ resulted in strong highly polymorphic bands whereas hybridization with $(TG)_{10}$ gave rise to a heavy background smear indicating larger number of loci containing TG repeats. Except for $(GATA)_4$, most of the polymorphic bands were faint indicating the presence of slightly diverged simple repeat sequences in rice.

Different fingerprinting parameters calculated from the hybridization patterns of different oligonucleotide probes with rice genotypes are depicted in Table 3. As indicated in this table, probability of identical match by chance was found to be 2.5×10^{-9} for $(GATA)_4$, 5.5×10^{-9} for $(GAA)_6$, 1.1×10^{-11} for $(GGAT)_4$, 7.5×10^{-7} for $(CAC)_5$ and 2.1×10^{-5} for $(GACA)_4$. The mean probability that two different rice genotypes have identical fingerprints for probes $(GACA)_4$, $(GATA)_4$, $(CAC)_5$, $(GAA)_6$ and $(GGAT)_4$ is $2.1 \times 10^{-5} \times 2.5 \times 10^{-9} \times 7.5 \times 10^{-7} \times 5.5 \times 10^{-9} \times 1.1 \times 10^{-11} = 2.4 \times 10^{-39}$. This estimation is based on the assumption that the DNA fingerprint bands identified by different oligonucleotide probes do not overlap with each other. Based on the above assumption, upto 10^{39} rice genotypes can be distinguished using these five oligonucleotide probes. A proper

TABLE 2

Number of bands hybridizing to different oligonucleotide probes in cultivar Basmati-370 digested with different restriction enzymes

	DraI	*ClaI*	*HindIII*
$(GATA)_4$	30	15	27
$(CAC)_5$	34	17	23
$(GAA)_6$	37	16	19
$(GGAT)_4$	38	13	18
$(GACA)_4$	24	15	15

FIG 3A-F DNA fingerprints of rice genotypes using oligonucleotide probes a $(GATA)_4$, b $(CAC)_5$, c $(GAA)_6$, d $(GGAT)_4$, e $(GACA)_4$ and f $(TG)_{10}$. DNA was digested with *DraI*. Lanes 1 through 6 contain DNA from *O.sativa* cultivars Indrayani (1), Adt-27 (2), Basmati-370 (3), Hakkoda (4), accessions of *O.rufipogon* (5) and *O.officinalis* (6). Molecular size markers in kb are indicated in the right margin.

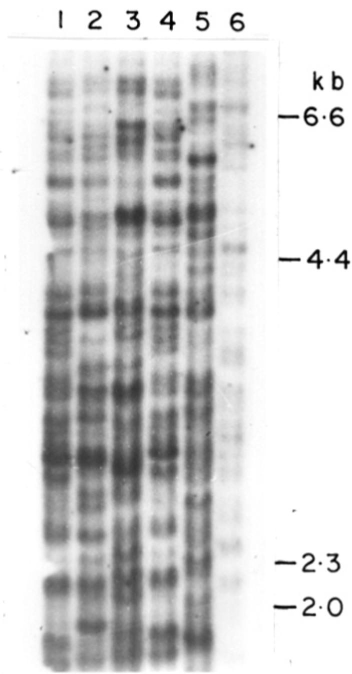


FIG. 3 A

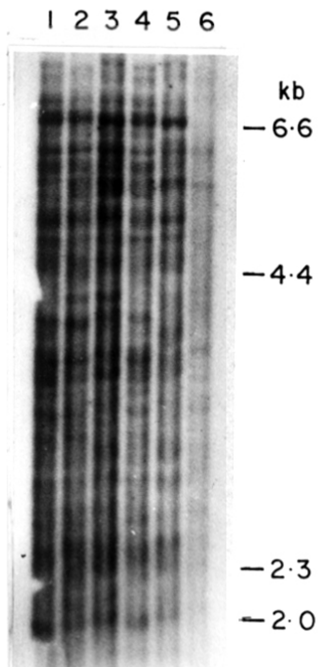


FIG. 3 B

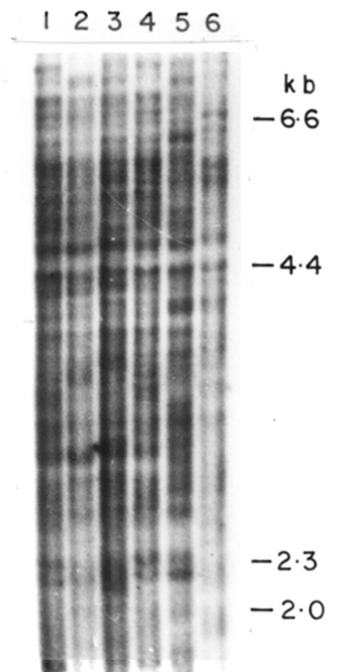


FIG. 3 C

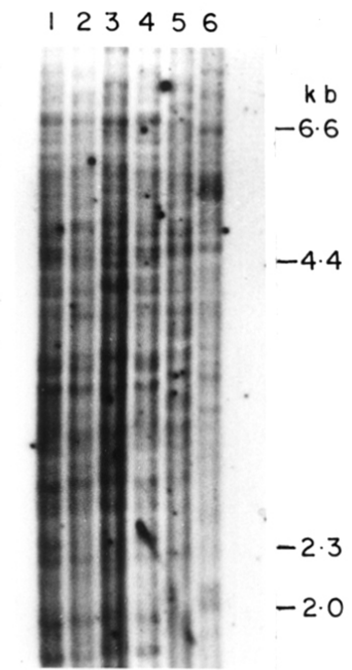


FIG. 3 D

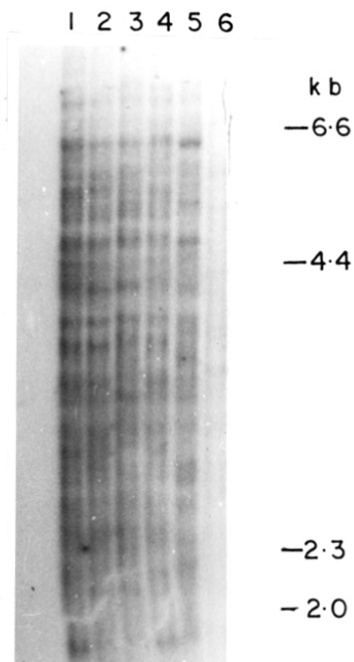


FIG. 3 E

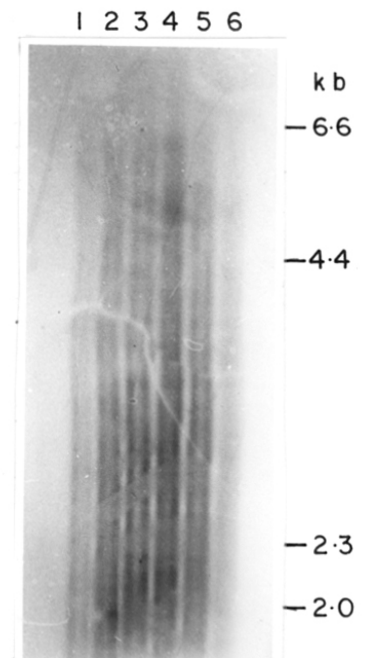


FIG. 3 F

TABLE - 3

Analysis of DNA Fingerprints using different oligonucleotide probes

	$(GATA)_4$	$(GAA)_6$	$(GGAT)_4$	$(CAC)_5$	$(GACA)_4$
Average no. of bands (n) $\pm$ SD	32.2 ± 3.4	33.8 ± 2.9	36.7 ± 2.3	31.2 ± 2.3	24.2 ± 1.7
Average no. of polymorphic bands between pairs $\pm$ SD	29.5 ± 8.3	28.7 ± 10.1	36.4 ± 12.9	22.5 ± 10.0	17.4 ± 9.1
Average similarity index $(\overline{X}_D) \pm$ SD	0.54 ± 0.14	0.57 ± 0.17	0.5 ± 0.17	0.64 ± 0.29	0.64 ± 0.18
Probability of identical match by chance $(\overline{X}_D)^n$	2.5×10^{-9}	5.5×10^{-9}	1.1×10^{-11}	7.5×10^{-7}	2.1×10^{-5}

Similarity index was calculated as $$X_D = \frac{2N_{AB}}{(N_A + N_B)}$$

Where N_{AB} is the number of bands present in both lanes.

N_A is the total number of bands in lane A and N_B is the total number of bands in lane B

combination of probe and enzyme is probably the key to efficient application of DNA fingerprinting in both varietal identification and breeding. Based on the maximum number of specific bands detected, $(GGAT)_4$ would be the best choice. However, probes $(GATA)_4$, $(GAA)_6$ and $(CAC)_5$ are preferred for better band resolution.

In order to investigate differences between individual plants of the same cultivar, DNAs were isolated from individual plants of the cultivar Indrayani, digested with *HindIII* and probed with the oligonucleotides used in this study. All the individual plants exhibited identical hybridization profiles and no differences could be detected (data not shown). Next, to examine the somatic stability of the DNA fingerprints, different tissues viz. seed, seedling and root derived from the cultivar Indrayani were examined. In this study also no differences were observed with the various oligonucleotide probes tested suggesting that the DNA fingerprint patterns were somatically stable (data not shown).

In an attempt to find out if the oligonucleotide probes can distinguish more genotypes of rice, we extended our studies to more number of wild rice genotypes as well as cultivars representing both indica and japonica subspecies. Hybridization with probe $(GATA)_4$ was carried out as it gave DNA fingerprints with the highest resolution. Figure 4 represents hybridization patterns detected by $(GATA)_4$ in *HindIII* digests of *O.sativa* cultivars from subspecies indica (lanes 1-4), japonica (lanes 5-7) and six wild rice genotypes (lanes 8-13) representing different genomes of rice as described in Materials and Methods. Although the bands in lanes 8, 10 and 11 are very faint, it was possible to score them on original autoradiograms. With $(GATA)_4$ as probe, all 13 rice genotypes could be distinguished from each other based on the divergent fingerprint profiles. Table 4 shows the similarity matrix generated from the DNA fingerprint profiles in Fig

FIG 4 Hybridization patterns of detected by probe $(GATA)_4$ in *HindIII* digests of DNA from several rice genotypes. Lanes 1 through 13 are *O.sativa* cultivars Indrayani (1), Basmati-370 (2), Pranava (3), Tambdarambhog (4), Fujisaka (5), Hakkoda (6), Norin-49 (7), accessions of *O.nivara* (8), *O.rufipogon* (9), *O.minuta* (10), *O.punctata* (11), O.officinalis (12) and *O.alta* (13). Molecular size markers in kb are indicated in the right margin.

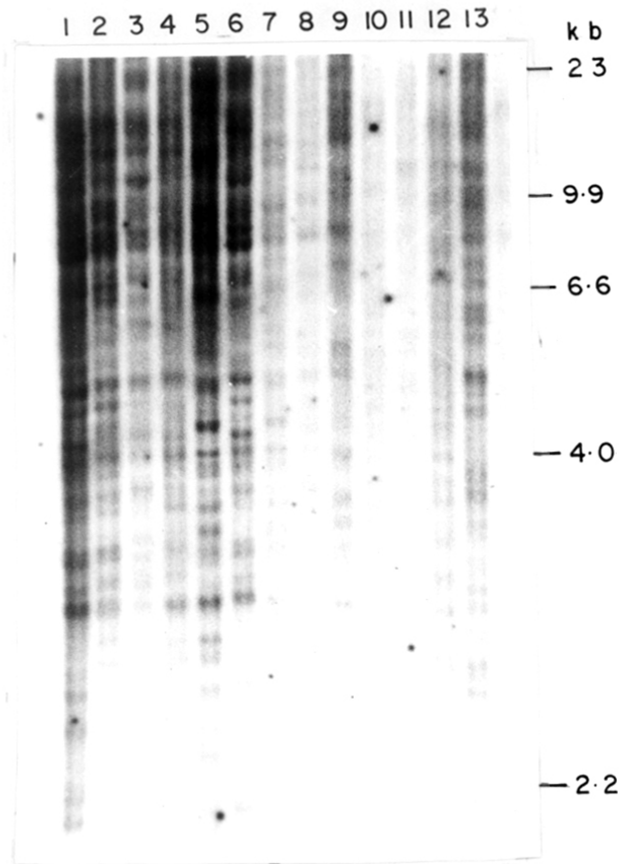


FIG. 4

TABLE 4

Similarity index values for different rice genotypes with probe $(GATA)_4$ and enzyme *HindIII*

	Indr	Basm	Pran	T-65	Fuji	Hakk	N-49	O.n	O.r	O.m	O.p	O.o
Basmati	0.59											
Pranava	0.74	0.67										
Taichung-65	0.54	0.59	0.51									
Fujisaka	0.43	0.38	0.38	0.3								
Hakkoda	0.46	0.33	0.39	0.4	0.47							
Norin-49	0.4	0.35	0.35	0.38	0.57	0.52						
O.nivara	0.3	0.28	0.3	0.28	0.37	0.37	0.42					
O.rufipogon	0.35	0.35	0.3	0.43	0.39	0.34	0.36	0.47				
O.malampuzhansis	0.19	0.22	0.21	0.16	0.17	0.22	0.25	0.39	0.24			
O.punctata	0.19	0.21	0.14	0.21	0.26	0.25	0.24	0.34	0.15	0.22		
O.officinalis	0.21	0.25	0.29	0.22	0.27	0.30	0.36	0.24	0.31	0.28	0.27	
O.alta	0.21	0.20	0.25	0.17	0.19	0.26	0.29	0.09	0.29	0.22	0.22	0.27

Indr - Indrayani, Basm - Basmati, Pran - Pranava, T-65 - Taichung 65, Fuj - Fujisaka, Hakk - Hakkoda, N-49 - Norin 49, *O.n - O.nivara, O.r - O.rufipogon, O.m - O.malampuzhansis, O.p - O.punctata, O.o - O.officinalis*

4. Thus, from the fingerprint profiles of Figs 3A - E, Fig 4 and Table 4, based on the similar patterns among cultivars of indica and cultivars of japonica subspecies and divergent DNA fingerprints of wild rice species, it can be concluded that indica and japonica subspecies appear to form two distinct groups with reference to GATA containing loci. Secondly, there is a clear distinction between the wild and cultivated rice genotypes. In addition, the level of variation observed between the cultivated rice and wild species belonging to AA genome specially *O.rufipogon* is less as compared to other genomes of rice.

If the polymorphic bands hybridizing to oligonucleotide probes have to be useful in rice breeding programs, they should be inherited in a Mendelian manner. Therefore, in a F_2 population derived from a cross between Basmati-370 (indica) and Taichung-65 (japonica), the segregation of polymorphic bands hybridizing to these oligonucleotide probes was studied. Figure 5 shows representative autoradiogram with probe $(GATA)_4$ hybridizing to the DNA of F_2 population digested with *BglI*. Polymorphic segregating bands were inherited in a Mendelian fashion in a 3:1 ratio indicative of monogenic inheritance.

3.2 DNA FINGERPRINTING IN RICE USING MINISATELLITES AND A CATTLE PROBE R18.1

To investigate the presence and polymorphic nature of minisatellites and related sequences alongwith a cattle probe containing GT repeats, rice DNA was initially digested with several restriction enzymes such as *AluI, HaeIII, HinfI,TaqI,BamHI, ClaI, DraI, EcoRI, EcoRV, HindIII, RsaI* and *XbaI* and was hybridized to eight different probes. Cultivars Indrayani, Basmati-370, Pranava, Tambda rambhog belonging to indica subspecies and Fujisaka, Hakkoda,

FIG 5 DNA fingerprints of a F_2 population of a cross between Basmati-370 (1) and Taichung-65 (2) with $(GATA)_4$ probe. Lane 3 to 11 contain BglI digested F_2 DNAs.

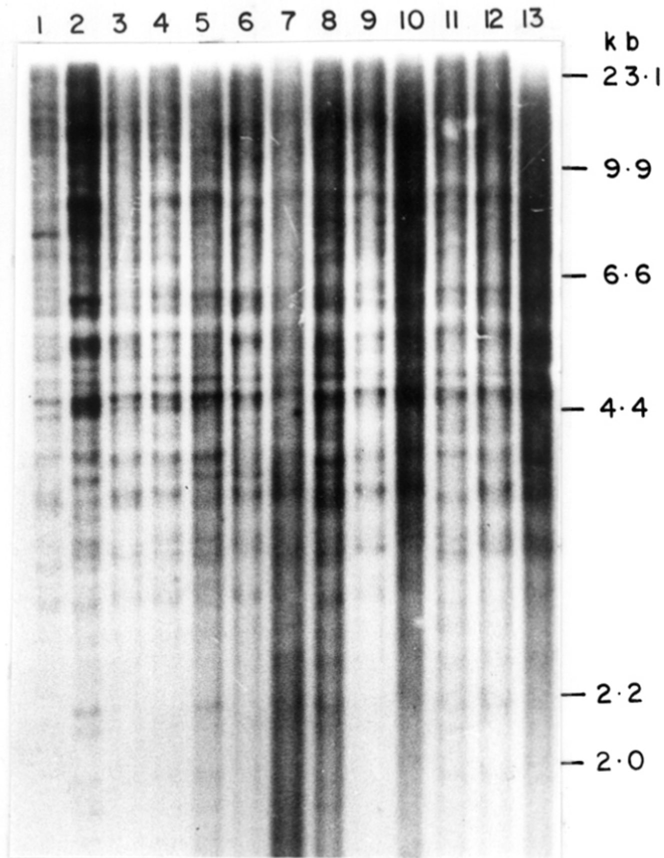


FIG. 5

Norin-49 belonging to japonica subspecies of *O.sativa* and accessions of wild species of rice belonging to *O.nivara, O.rufipogon, O.minuta, O.punctata, O.officinalis* and *O.alta* were used in this study.

Most of the restriction enzymes with four base pair specificities gave rise to a background smear in low molecular weight range whereas restriction enzymes with six base pair specificities such as *DraI* gave the highest level of polymorphism with good resolution similar to that observed with oligonucleotide probes. A probe with two tandem repeats of GGTGGTGG corresponding to the consensus core repeat of *chi* sequence of *E.coli* detected limited variability in rice cultivars, while interspecific variability was more pronounced. The sequence ACCGGGACAGGAACTGGA corresponding to the repeat in "*Per*" gene of *Drosophila* (5) showed less polymorphism in all the genotypes of rice used in our study. Similarly, repeats in the 3' hypervariable region of apolipoprotein B TTTTATAATTAATATTTT could detect relatively weak signals. A probe consisting of three tandem repeats of the sequence TTAGGG which constitute telomeric repeats, previously shown to be conserved in eukaryotes, was however, observed to be remarkably polymorphic. Strong hybridization signals appeared above 30kb in most cases which could not be resolved on agarose gel suggesting a lack of restriction enzyme cutting site in telomeres. In cultivar Adt-27, an interesting observation was made in *DraI* digest of DNA where the average length of telomeric sequences appeared to be shorter as compared to that in other rice genotypes (Figure 6A). Similar observations were also made in case of cultivar Fujisaka and wild rice species - *O.alta* (Figure 6B). This is not the result of DNA degradation because the patterns of restriction enzyme digested DNAs on ethidium bromide stained gel were similar in all rice genotypes. Similar results and conclusions

FIG 6A Hybridization patterns detected by a telomeric repeat probe in *DraI* digested rice DNAs. Lanes 1 through 5 are as described in figure 3.

FIG 6B Hybridization patterns detected by a telomeric repeat probe in *HindIII* digested rice DNAs. Lanes 1 through 13 are as described in figure 4.

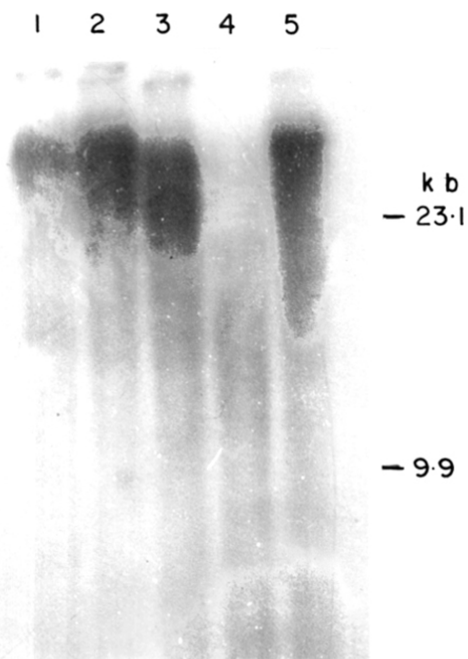


FIG. 6A

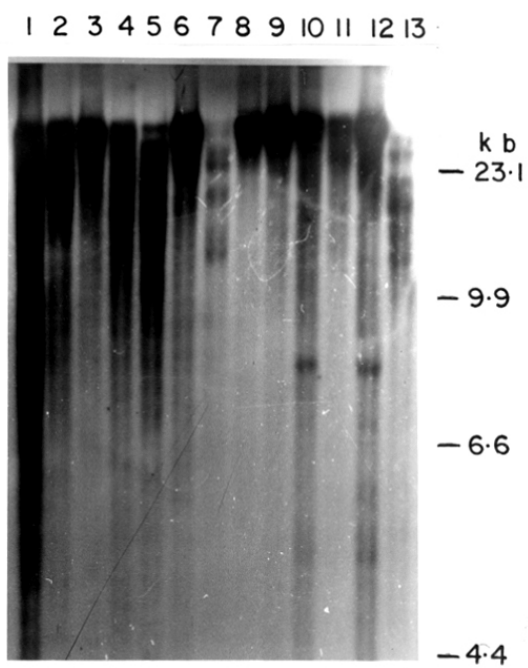


FIG. 6B

were made in maize (6). The use of mouse rDNA probe corresponding to the small subunit of ribosomes although showed polymorphism could not differentiate all the 14 rice genotypes (Figure 7).

M13 repeat probe of 280bp length, pV47 - a human minisatellite probe and an oligonucleotide GAGGGTGGCGGTTCT consisting of the consensus core repeat of M13 revealed complex hybridization patterns with a good polymorphism among rice varieties. Diffuse background regions were reported when DNA fingerprinting in plants was compared to human DNA (7). Background cross hybridizing sequences in plants indicate that either they are divergent from mini or microsatellite repeats or fewer subrepeats are dispersed throughout the genome. R18.1 containing six poly (GT) stretches hybridized to several restriction fragments and could generate highly polymorphic DNA fingerprints among rice genotypes indicating the presence of limited number of loci with poly (GT) stretches. Previously, we have shown that a simple $(TG)_{10}$ probe resulted in a heavy background smear indicating the presence of large number of loci containing $(TG)_{10}$ repeat (8). Recently, Wu and Tanksley (9) have shown the presence of $(TG)_n$ repeats once in every 480 kb of the rice genome. These microsatellites were shown to be dispersed in the genome and were useful to detect polymorphisms in rice genotypes. The results showed here that all the DNA sequences used in our studies although present in rice genome, occur at different abundances. In view of these results we will now describe the data obtained mainly with R18.1, pV47 and M13 probes.

Figures 8A & B represent hybridization patterns detected by R18.1 in *DraI* & *HindIII* digests of rice genotypes respectively while Fig 9 & 10 show DNA fingerprints generated by pV47 & M13 repeat probes respectively in *DraI* digests of rice genotypes as described in figure legends. Some hybridization signals

FIG 7 Hybridization patterns detected by mouse rDNA probe in *DraI* digested rice DNAs. Lanes 1 through 13 are as described in figure 4.

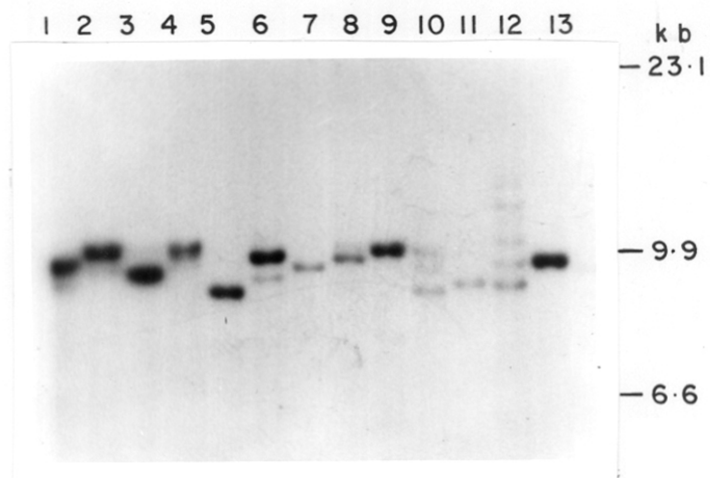


FIG. 7

FIG 8A DNA fingerprints of *DraI* digested rice DNAs probed with R18.1. Lanes 1 through 13 are as described in figure 4.

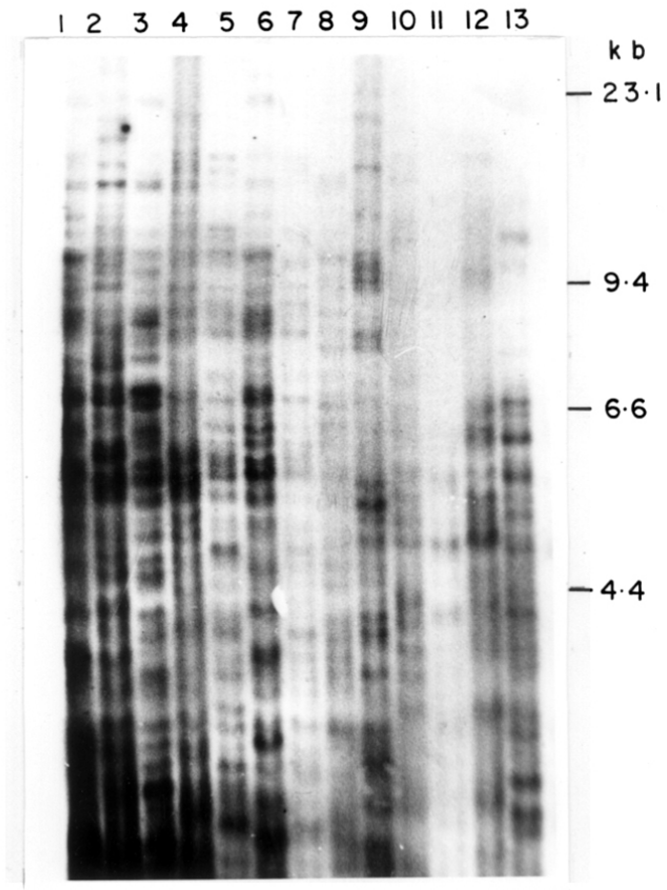


FIG. 8 A

FIG 8B DNA fingerprints of *HindIII* digested rice DNAs probed with R18.1. Lanes 1 through 13 are as described in figure 4. Lane 14 contains DNA from *O.malampuzhansis*.

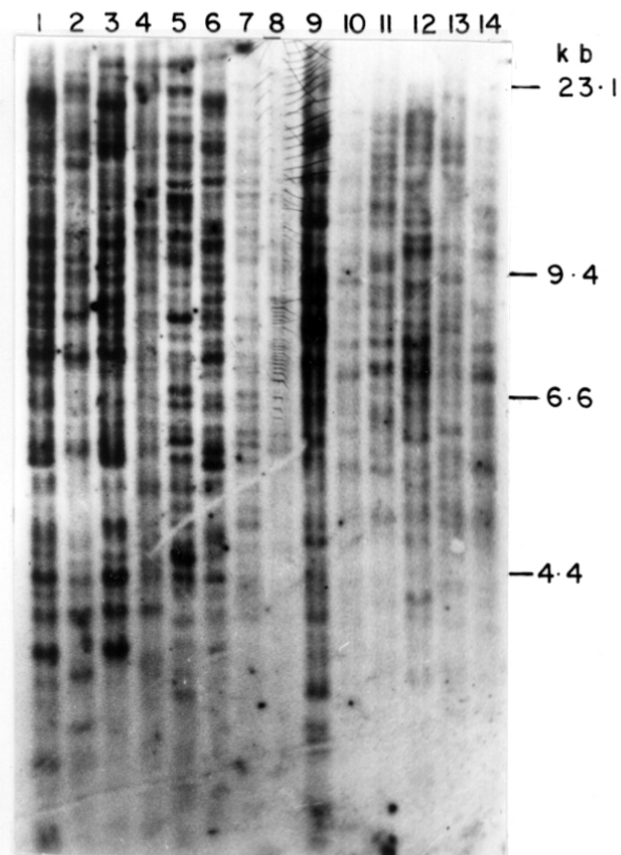


FIG. 8 B

though weak in the figures were visible on the original x-ray film. As seen from the Figs 9 & 10, the two probes, namely pV47 and M13 could detect polymorphisms in comparatively high molecular weight region (more than 4kb) in all the rice genotypes while the banding pattern was mostly monomorphic in the lower molecular weight region. Intraspecific variability among the indicas and japonicas as well as extensive interspecific variability were detected between the cultivated and wild rices with R18.1, pV47 and M13 repeat probes. However, probe R18.1 produced the best resolution of scorable bands as compared to pV47 and M13.

The DNA fingerprint profiles in Fig. 8A & B, 9 & 10 and the similarity index values calculated from these figures indicate Basmati-370 to be closely related to the indicas than to japonicas and wilds as compared to the isozyme analysis by Glaszmann (10), where Basmati-370 is the only variety with indica morphology placed in Group V and closer to japonicas. Secondly, there is a clear distinction between the wild and cultivated rice genotypes. Different fingerprinting parameters calculated from the hybridization patterns in Fig. 8A, 9 and 10 are depicted in Table 5. Among the three probes the number of bands detected with R18.1 was higher (38.8) and the level of band sharing between rice genotypes was lower (0.31) indicating higher level of polymorphism with R18.1.

A proper combination of probe and enzyme is thus a key to efficient DNA fingerprinting which in turn finds application in both varietal identification and breeding. A probe suitable for DNA fingerprinting should distinguish genotypes to such an extent that the probability is very low that two genotypes will show identical fragment profiles. We have estimated probability of identical fingerprints by chance to be 2.5×10^{-20}, 1.5×10^{-11} and 2.0×10^{-10} for genus *oryza* and 1.0×10^{-17}, 1.1×10^{-8} and 4.4×10^{-7} for rice cultivars with R18.1, pV47

FIG 9 DNA fingerprints of *DraI* digested DNA from rice genotypes probed with pV47. Lanes 1 through 14 are as described in figure 8B.

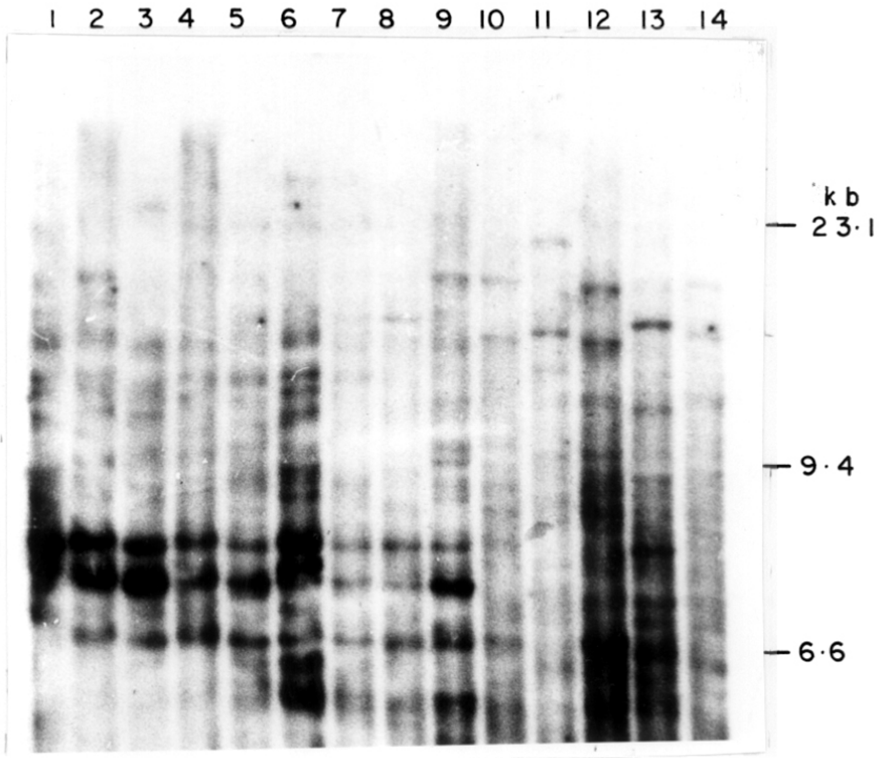


FIG. 9

FIG 10 DNA fingerprints of *DraI* digested DNA from rice genotypes probed with M13 repeat probe. Lanes 1 through 14 are as described in figure 9.

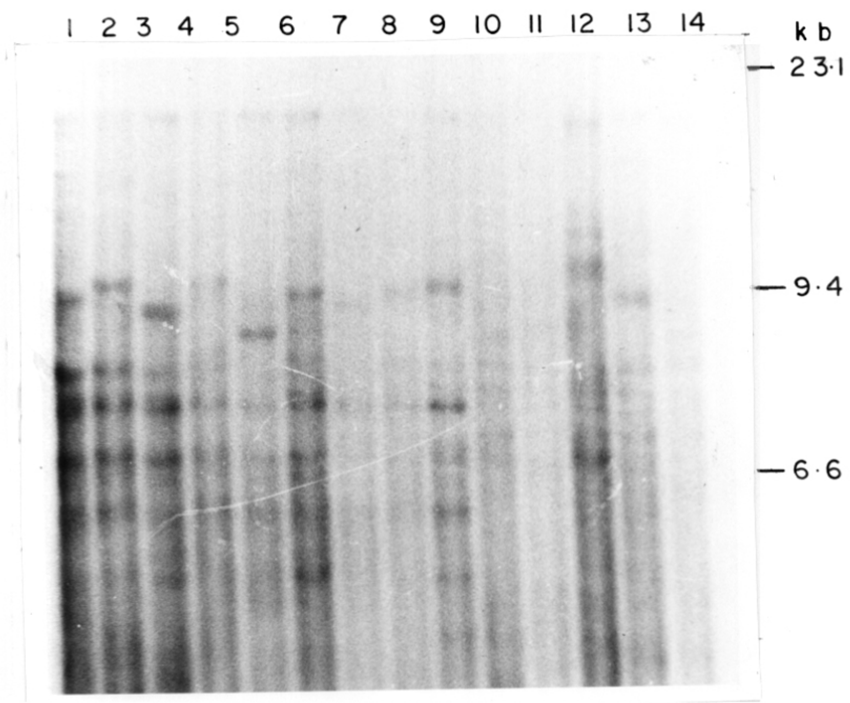


FIG. 10

TABLE 5

Analysis of DNA fingerprints of rice using M13, pV47 and R18.1.

	M13	pV47	R18.1
Average number of bands (n) ± SD	29.7 ± 2.0	27.2 ± 3.6	38.8 ± 4.5
Number of comparisons	91	91	78
Average number of polymorphic bands between pairs ± SD	31.4 ± 7.5	32.6 ± 7.3	53.5 ± 9.1
Average similarity index ($\overline{X}_D$) ± SD	0.47 ± 0.14	0.4 ± 0.13	0.31 ± 0.08
Probability of identical match by chance ($\overline{X}_D$)^n	2×10^{-10}	1.5×10^{-11}	2.5×10^{-20}

Similarity index was calculated as $$X_D = \frac{2N_{AB}}{(N_A + N_B)}$$

Where N_{AB} is the number of bands present in both lanes.

N_A is the total number of bands in lane A and N_B is the total number of bands in lane B

and M13 repeat probes respectively. The mean probability that two different rice genotypes have identical fingerprints for probes R18.1, pV47 and M13 is 2.5 x 10^{-20} x 1.5 x 10^{-11} x 2.0 x 10^{-10} = 7.5 x 10^{-41}. This estimation is based on the assumption that the DNA fingerprint bands identified by different probes do not overlap with each other. Based on the above assumption, upto 10^{41} rice genotypes can be distinguished using R18.1, pV47 and M13 probes.

In order to investigate differences between individual plants of the same cultivar, DNAs were isolated from individual plants of the cultivar Indrayani, digested with *HindIII* and probed with R18.1, pV47 and M13. No differences were detected among individual plants mentioned above (data not shown). We also examined the fingerprint patterns of different tissues viz. root, seedling and leaf of cultivar. Here again, no differences were observed with R18.1, pV47 and M13 indicating somatic stability of fingerprints (data not shown).

To gain insight into the mode of inheritance of DNA fingerprints detected by R18.1, the most polymorphic probe in the present study, 20 plants of a F_2 population of a cross between Basmati-370 (indica) and Taichung-65 (japonica) were analyzed. The parents differ in many bands with different restriction enzymes as shown in Figure 11A. All the bands in the offspring could be traced back to the parents. Representative DNA fingerprints are shown in Figure 11B. Here, due to non availability of nytran membrane, the quality of autoradiogram is not as good as in figure 11A. The segregation ratios did not deviate significantly from 3:1 indicating a Mendelian pattern of inheritance.

FIG 11A DNA fingerprints of Basmati-370 and Taichung-65. Basmati-370 DNA (lanes 1,3,5,7,9) and Taichung-65 DNA (lanes 2,4,6,8,10) were digested with *BamHI* (lanes 1,2), *BglI* (lanes 3,4), *EcoRI* (lanes 5,6), EcoRV (lanes 7,8) & *PstI* (lanes 9,10).

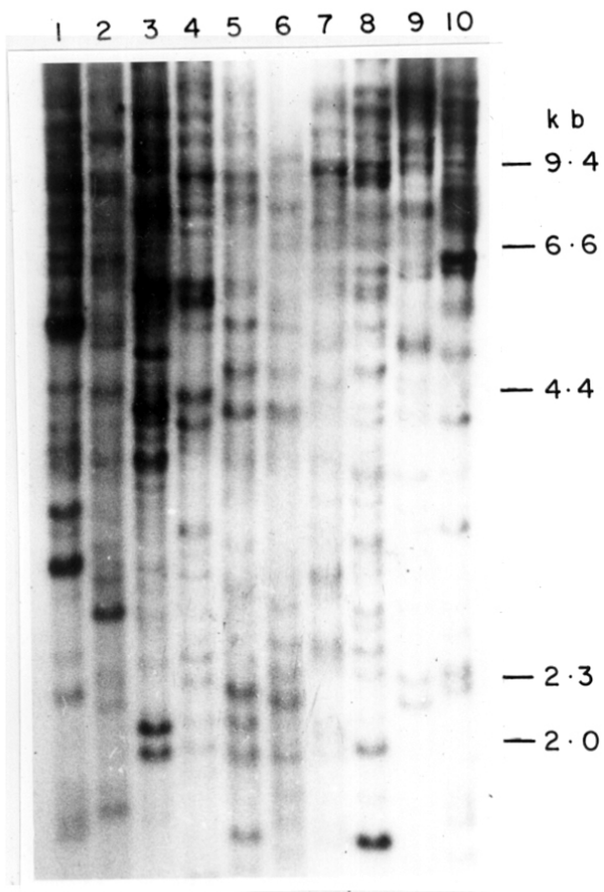


FIG. 11 A

FIG 11B DNA fingerprints of a F_2 population of a cross between Basmati-370 (1) and Taichung-65 (2) with R18.1 as probe. Lanes 3 to 11 contain EcoRV digested F_2 DNAs. Bands specific to B-370 and T-65 are indicated by arrows and dots respectively.

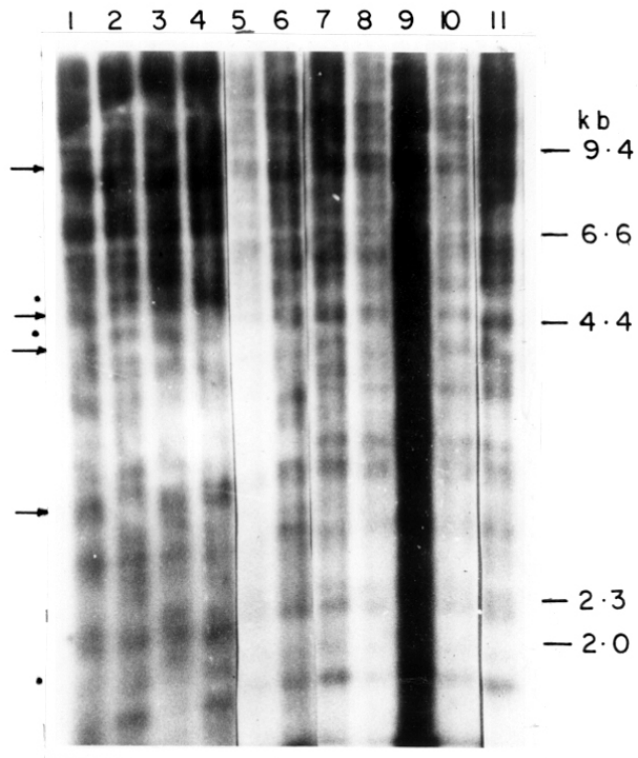


FIG. 11 B

3.3 DEMONSTRATION OF GENE TRANSCRIPTS IN RICE HOMOLOGOUS TO $(CAC)_5$

Although simple repeats are present abundantly in the genome and detect high levels of polymorphisms, their functional significance remains unknown. Detection of transcriptional activity of these repeats forms the basis for a possible role in cellular function including gene expression. To test whether simple repetitive DNA sequences are transcribed in rice, oligo- nucleotide probes specific for simple repeats viz $(CAC)_5$, $(GATA)_4$, $(GACA)_4$, and $(TG)_{10}$ were hybridized to dot blot of increasing concentration of RNAs ranging from 0.5 to 8μg of DNAs isolated from two different tissues viz. seed and leaf of rice. *E.coli* RNA and DNA were taken as negative control since simple repeats are presumably absent in *E.coli*. All the four probes hybridized to DNA, but only $(CAC)_5$ hybridized to the RNAs at the concentration as low as 0.5μg RNA as shown in Figs 12A & B thus indicating the possible presence of transcriptionally active sequences homologous to $(CAC)_5$ amongst the microsatellites tested in our study.

In order to determine the size class of RNAs homologous to $(CAC)_5$, total RNAs isolated from two different tissues of rice such as seed and leaf were electrophoresed, blotted and hybridized with end labelled $(CAC)_5$ (Fig. 13A). One band at 1.2kb was observed to be common in both seed and leaf RNA. However, in case of seed tissue (lane a), a band at 1.9kb was additionally present, while in leaf tissue (lane b), a 2.3kb band was very intense indicating the presence of tissue specific transcripts in seed and leaf. The hybridization of $(CAC)_5$ oligonucleotide probe with those of total RNA species in a tissue specific manner raised an obvious question whether sequences complementary to $(CAC)_5$ were a part of poly$(A)^+$ species in rice. On hybridization of poly$(A)^+$ RNAs with $(CAC)_5$, strong signals corresponding to 1.2kb RNA were obtained in case of seed and leaf. However,

FIG 12A Dot blot hybridization of $(GATA)_4$ to total RNA and DNA from different tissues of rice. Different concentrations (0.5μg, 2μg & 8μg) of RNAs extracted from seed and leaf as well as 8ug of DNA from each of the above tissues of rice as well as RNA and DNA from *E.coli* were dot blotted on to Hybond N membrane and hybridized with 5' end labeled oligo-nucleotide $(GATA)_4$ at 35°C as described in Materials and Methods. Lane a - *E.coli* RNA and DNA; Lane b - Rice seed RNA and DNA; Lane c - Rice Leaf RNA and DNA.

FIG 12B The same blot was reprobed with 5' end labelled $(CAC)_5$ at 45°C as described in Materials and Methods.

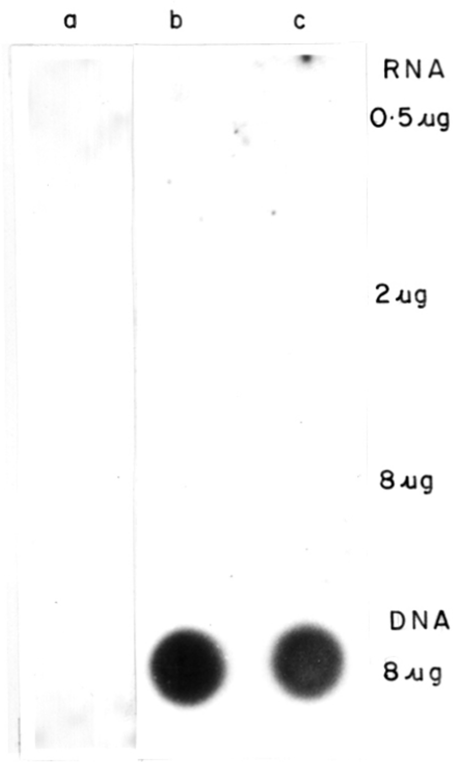


FIG. 12 A

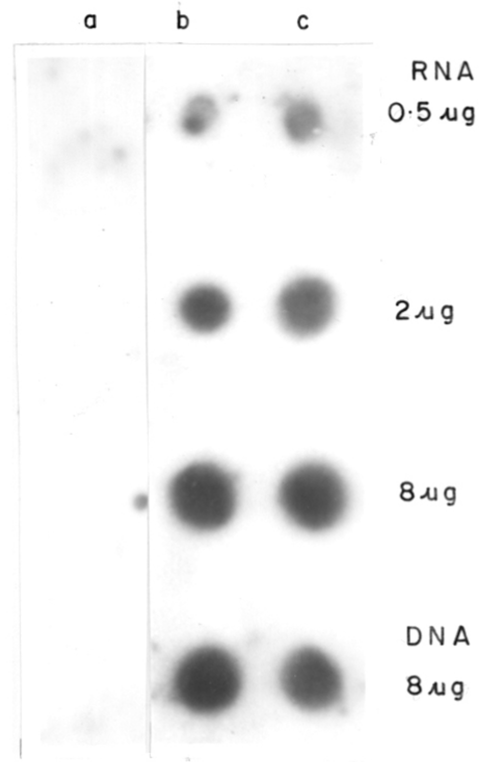


FIG. 12 B

FIG 13A Northern blot hybridization of rice RNA from different tissues probed with end labelled $(CAC)_5$. 40μg of total RNA was electrophoresed, blotted and hybridized as described in Materials and Methods. Lane a - Rice seed RNA; Lane b - Rice leaf RNA.

FIG 13B Northern blot hybridization of rice poly$(A)^+$ RNA with end labelled $(CAC)_5$. Blotting, hybridization and washing of the above filter was carried out as described in Materials and Methods. Lane a - Rice seed poly$(A)^+$ RNA; lane b - Rice leaf poly$(A)^+$ RNA.

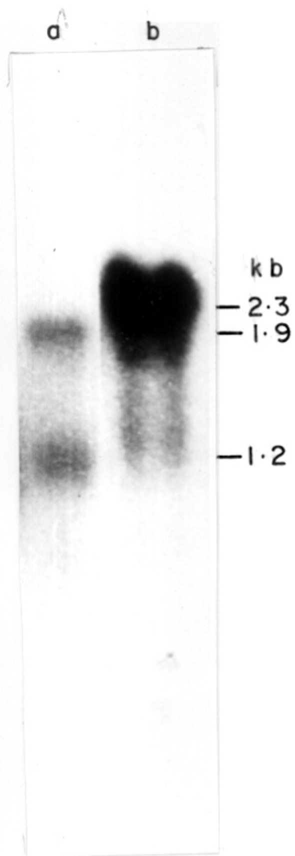


FIG. 13A

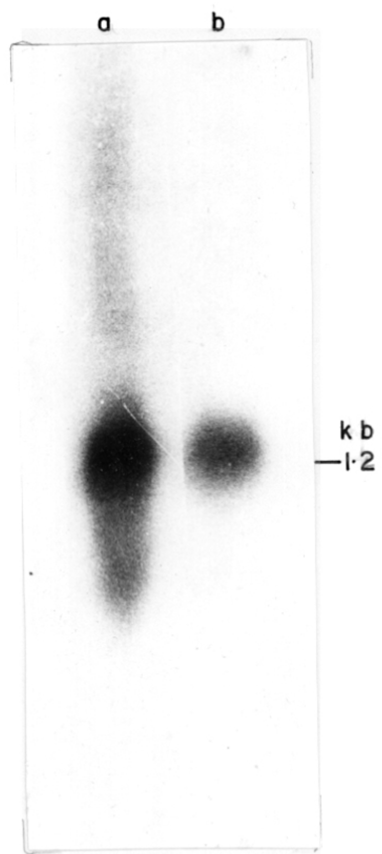


FIG. 13B

hybridization to heterogeneously sized RNA was observed in seed and this was totally absent in leaf. This 1.2kb band which gave an intense hybridization signal in poly(A)$^+$ RNA (Fig. 13B) was observed to be very faint in total RNA (Fig. 13A). This might be due to the low proportion of poly(A)$^+$ RNA (400ng) in total RNA (40μg) in Fig 13A as compared to the amount of poly(A)$^+$ RNA (5μg) loaded on gel in Fig 13B. This blot was reprobed with soybean rDNA in order to eliminate the possibility of hybridization to carry over contamination of non poly(A)$^+$ RNAs and no signals were observed (data not shown).

The hybridization patterns in total RNA (Fig. 13A), especially the signals at 1.9 and 2.3kb and poly(A)$^+$ RNA (Fig. 13B), indicate that sequences homologous to $(CAC)_5$ are present in poly(A)$^-$ and poly(A)$^+$ RNA species. When a search was made for the presence of $(CAC)_5$ in the ribosomal DNA and histone gene sequences of rice, using the computer software program DNASIS (Pharmacia), no significant homology was found to $(CAC)_5$. Since all the poly(A)$^-$ RNA will not account for transcripts such as rRNAs and histone RNAs, it is possible that repetitive sequences may have resulted in a population of RNAs of sizes 2.3 and 1.9kb giving hybridization to $(CAC)_5$.

3.4 STATUS OF DNA METHYLATION IN RICE USING SOME MICROSATELLITES AND MINISATELLITES AS PROBES

DNA methylation of cytosine and adenine residues is crucial for various functions including gene regulation. Methylation of hypervariable loci was shown to be associated with activation of diseases in humans. Although the presence and functional significance of methylations at hypervariable loci in plants is very poorly studied, the possiblity of specific function for these methylations makes this study interesting. 5-methylcytosine accounts for 20-30% of total cytosine in plant genomes and most of the methylation of DNA is at the CpG

dinucleotide. However, there have been a few reports on adenine methylation in plants (11,12). DNA methylation can be conveniently studied by digestion of genomic DNA with restriction endonuclease isoschizomer pairs differing in their sensitivities to cytosine methylation such as *MspI* and *HpaII* at the sequence CCGG and to both cytosine and adenine methylation by *MboI, Sau3AI* and *DpnI* at the sequence GATC followed by electrophoretic separation, blotting and hybridization with specific probes. Digestion with *HpaII* is inhibited when the internal and\or external cytosine residue in its CCGG recognition sequence is methylated whereas digestion with *MspI* is inhibited only when external C is methylated in its CCGG recognition sequence. Hence the differential pattern by *MspI* and *HpaII* gives information regarding CpG dinucleotide methylation in the sequence CCGG. Digestion with *Sau3AI* is inhibited when the cytosine residue in GATC is methylated whereas *MboI* is insensitive to C-methylation but inhibited by adenine methylation. In contrast to this, *DpnI* restriction enzyme cuts the DNA at the same recognition site GATC only when adenine is methylated. Thus the digestion patterns by *MboI* and *Sau3AI* can be studied for extent of cytosine methylation whereas *MboI* and *DpnI* digestion patterns can be studied for adenine methylation levels in the sequence 5'-GATC-3'.

In the present study, a total of four DNA probes including $(GATA)_4$, R18.1, pV47 and mouse rDNA were used to investigate the methylation status at the loci detected by these probes in rice. Since ribosomal DNA probe is a conserved repeat sequence, it was used for comparison in these experiments as control. In all the experiments, the rice cultivars used for the analysis were Adt-27, Basmati-370, Indrayani and Taichung-65. The DNA samples of all the four cultivars were digested with restriction enzyme isoschizomers *MspI, HpaII, MboI, Sau3AI* and *DpnI*. The DNA samples from leaf tissue of each of the cultivar,

digested with the above five restriction enzymes were loaded in the same order for electrophoresis, were blotted and were hybridized with various probes. Figure 14A & B show the methylation status of sequences homologous to R18.1 in different rice cultivars. From Fig 14A, it is clear that in case of hybridization of R18.1 probe with *Sau3AI* digested Adt-27 DNA, the signals were observed in comparatively high molecular weight region than in case of *MboI* digested DNA indicating presence of cytosine methylation in the sequence GATC. In case of Indrayani (Figure 14B), the hybridization patterns in *MboI* as well as *Sau3AI* digested DNAs were identical indicating neither A nor C methylation was present in sequence GATC. However, Basmati-370 and Taichung-65 showed some interesting results. The *MboI* digested DNAs gave a few additional bands in the range of 1.5-2.5kbp when hybridized with R18.1 which were absent in *Sau3AI* lane suggesting presence of A-methylation in some of the GATC sequences. But to our surprise, this did not reflect in the *DpnI* digested lanes where the hybridization signals were present only in a very high molecular weight region. The only possible explanation for this observation is the less frequency of GATC sequences where A is methylated. No two GATC sites are very closely spaced in the genome where A is methylated thus giving high molecular weight signal in *DpnI* digestion. However, such A methylated GATC sites are spaced within 1.5-2.5kbp distance with nonmethylated GATC sites on either sides giving signals in the 1.5-2.5kbp region in case of *MboI* and are absent in case of *Sau3AI*. Most hybridizing fragments in *HpaII* digested DNA were in the higher molecular weight region which were absent in *MspI* digested DNA suggesting a high frequency of CpG methylation in the sequence CCGG. Several bands were common in *MspI* and *HpaII* digested DNAs indicating that these cutting sites were not methylated in all the four cultivars.

FIG 14A DNA methylation patterns detected in cultivars-Adt-27 & Basmati-370 digested with *MspI, HpaII, MboI, Sau3AI, DpnI* & probed with R18.1. Lanes 1 through 5 contain cultivar Adt-27 DNA digested with *MspI* (1), HpaII (2), *MboI* (3), *Sau3AI* (4) and *DpnI* (5). Lanes 6 to 10 contain cultivar Basmati-370 DNA digested with *MspI* (6), *HpaII* (7), *MboI* (8), Sau3AI (9) and *DpnI* (10). Molecular size markers in kb are indicated in the right margin.

FIG 14B DNA methylation patterns detected in cultivars Indrayani and Taichung-65 digested with *MspI, HpaII, MboI, Sau3AI, DpnI* and probed with R18.1. Lanes 1 through 10 are as described in figure 14A.

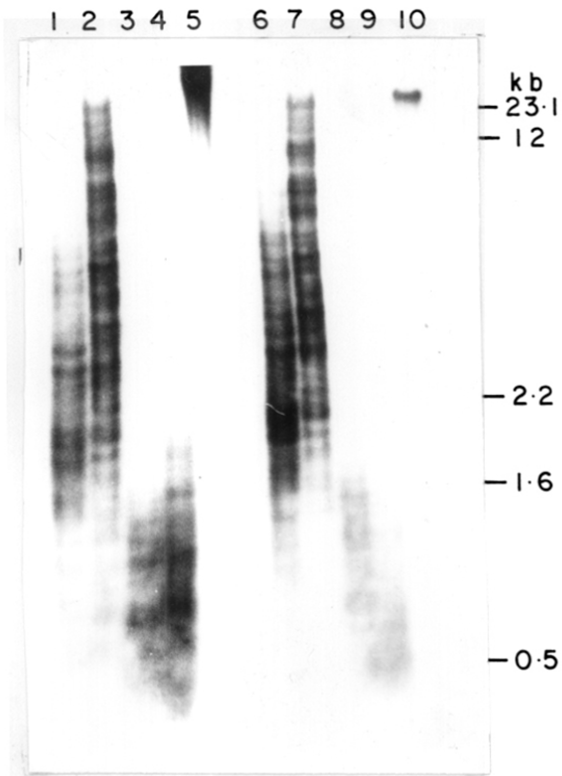


FIG. 14 A

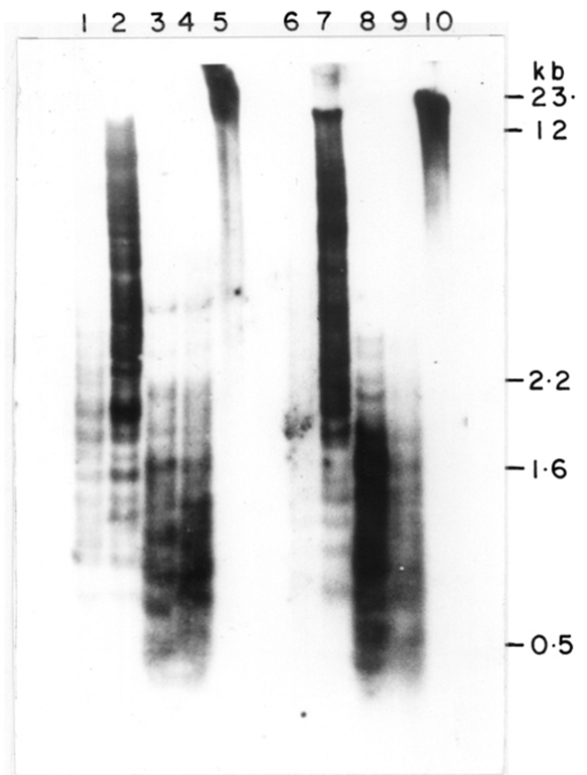


FIG. 14 B

Similar analysis was performed for other three probes namely pV47, $(GATA)_4$ and mouse rDNA. All of these probes showed similar pattern of CpG methylation in the sequence CCGG in all the four cultivars. The results obtained by comparative analysis of digestion pattern with methylation specific enzymes and the hybridization with above probes in all the four cultivars are summarized in Table 6. From the Table it is evident that

[1] All the four cultivars with all the probes show high levels CpG methylation in the sequence CCGG without any exception.

[2] All the four probes indicate higher levels of C methylation in the sequence GATC in the cultivar Adt-27. In cultivars Basmati-370 and Indrayani C-methylation is observed with rDNA probe.

[3] Adenine methylation in the sequence GATC is absent in Adt-27 and Indrayani with all the four probes. Only exception is Indrayani with $(GATA)_4$ probe.

[4] Contrary to above observation, A methylation in the sequence GATC is observed in Basmati-370 and Taichung-65 with all the four probes without exception.

[5] Thus it can be inferred from Table 6 that the indica japonica derivative Adt-27 shows different levels of C methylation. Basmati-370 which is an indica rice shows methylation pattern quite similar to Taichung-65 which is a japonica rice, whereas Indrayani an indica rice shows minimum levels of A or C methylation in the sequence, GATC.

Attempts were made to study tissue specific C methylation patterns in the sequence CCGG using R18.1 and $(GATA)_4$ probes in the case of one cultivar

TABLE 6

DNA methylation patterns in different rice cultivars

	R18.1		pV47		rDNA		$(GATA)_4$	
	CCGG	GATC	CCGG	GATC	CCGG	GATC	CCGG	GATC
		•A •C		•A •C		•A •C		•A •C
Adt 27	•CpG	X •C	•CpG	X •C	•CpG	X •C	•CpG	X •C
Basmati 370	•CpG	•A X	•CpG	•A X	•CpG	•A •C	•CpG	•A X
Indrayani	•CpG	X X	•CpG	X X	•CpG	X •C	•CpG	•A X
Taichung 65	•CpG	•A X	•CpG	•A X	•CpG	•A X	•CpG	•A X

•A - adenine methylation

•C - cytosine methylation

X - no methylation

namely Indrayani. For this purpose, DNAs from root, seedling and leaf tissues were used. However, the probes under our study did not show much of C methylation in different tissues under consideration.

3.5 IDENTIFICATION AND CHARACTERIZATION OF A MICROSATELLITE FROM RICE USING $(GATA)_4$ PROBE

In the previous section, the use of mini and microsatellite probes in DNA fingerprinting in rice was established. The probes used in our studies were mostly standard probes, reported earlier from various systems. Such probes are versatile and can be used almost for all the eukaryotic systems including plants to distinguish between various cultivars. It is generally necessary to use many enzyme and probe combinations to get the DNA fingerprint which can provide as "passport" information about the particular cultivar. However, the probes identified from homologous system always increase the probability of detecting higher levels of polymorphism (13) and to develop locus specific probes. It is with this objective, an attempt was made in the present study to identify a homologous probe from rice system. In this section, identification and characterization of such a rice probe having microsatellite sequences GATA, TG and AG are described. A *PstI* genomic library of cultivar Malkolam (a land race from Maharashtra State, India) in pUC18 was screened with probes $(GATA)_4$, $(TG)_{10}$, R18.1, pV47 and M13, the most informative probes in rice. A total of 40 positive clones were identified. Figure 15A shows two GATA positive clones - $1E_6$ and $1C_9$ identified by colony hybridization. Based on strong hybridization signal, $1E_6$ was selected for further analysis. Since *PvuII* sites are located on either side of the multiple cloning site (MCS) region, plasmid DNAs from 40 positive clones including $1E_6$ were isolated and digested with *PvuII* to release the inserts as shown in Figure 15B. This method was chosen as *PstI* did not give complete digestion probably

FIG 15A Colony hybridization showing two GATA positive clones. Colony filters of the *PstI* partial genomic library were hybridized with $(GATA)_4$.

FIG 15B Plasmid DNAs from $(GATA)_4$, $(TG)_{10}$, M13, pV47 and R18.1 positive clones. Plasmid DNAs were cut with *PvuII* to release the insert.

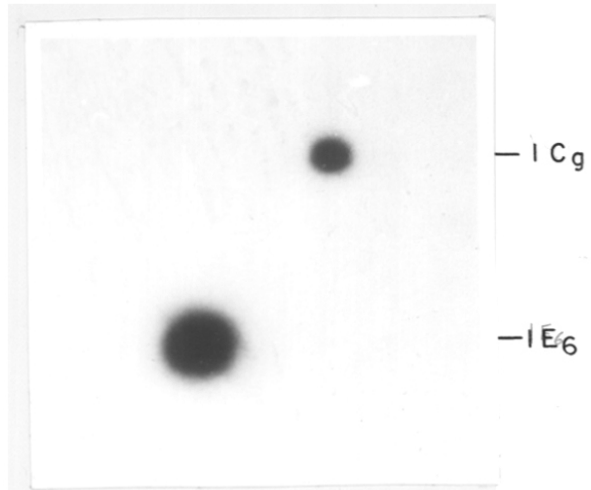


FIG. 15 A

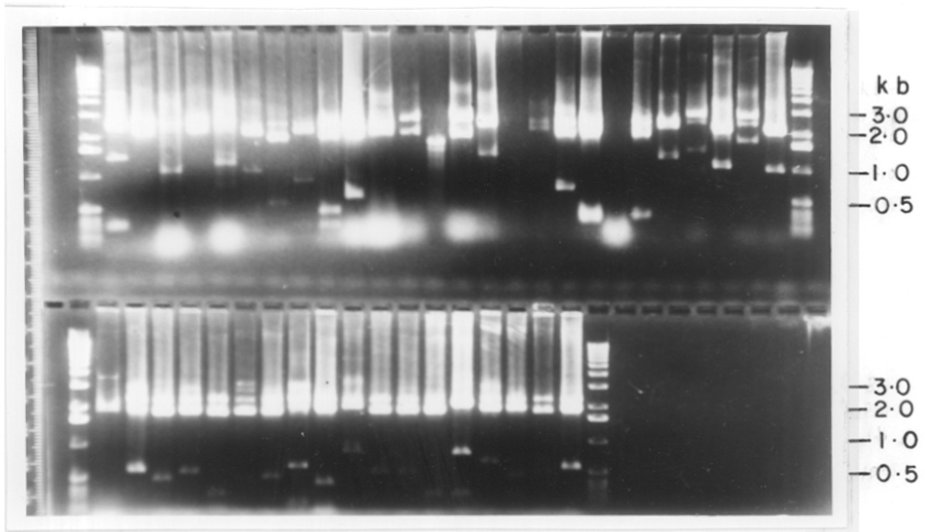


FIG. 15 B

due to methylation of the enzyme cutting sites in some of the clones. Insert size of 1E_6 clone was thus determined to be 1.7kbp. In order to construct a restriction map for this clone, IE_6 was digested with several restriction enzymes in the multiple cloning site of pUC18 and also with *PvuII* (Figure 16A). The clone showed sites for *PvuII* and *HindIII* within the insert which could be mapped with the help of double digestions such as *PvuII-PstI* and *PvuII-HindIII* (Figure 16B). Based on hybridization of this gel with $(GATA)_4$, GATA repeats in the clone were inferred to be present in the 0.9kbp *HindIII-PstI* fragment (Figure 16C). The restriction map of the clone is shown in Figure 17. The *HindIII-PstI* fragment of 0.9kbp was subcloned into pUC18 for further analysis. The subclone was digested with *PvuII* and *HindIII* (Figure 18A) and was hybridized with $(GATA)_4$ (Figure 18B). The hybridization signal at 1.0kbp *PvuII* fragment (0.8kbp insert + 0.2kbp flanking plasmid) and 3.5kbp *HindIII* fragment (*HindIII* linearized the clone giving one band of 3.5kbp having 2.7kbp plasmid and 0.8kbp insert together) confirmed the presence of GATA repeats in the 0.8kbp *HindIII-PstI* fragment. This fragment also hybridized to $(TG)_{10}$ suggesting a clustering of both GATA and TG repeats in the clone. DNA sequencing of the subclone was carried out in order to find out the actual number and the nature of arrangement of these repeats in this clone. Based on the available sequence obtained using universal and reverse sequencing primers as well as a primer synthesized from the sequence, thirteen imperfect GATA repeats and nine AC/TG repeats were identified in the clone. Apart from these expected repeats, twelve AG/TC repeats were also identified in the sequence shown in Fig 19. Thus the 1E_6 clone identified from the rice library having three microsatellite sequences clustered in it was interesting and it was necessary to exploit this information further to analyze fine variations in the rice genome. Usually the microsatellite sequences are highly polymorphic due to the presence

FIG 16A Restriction enzyme digestions of clone $1E_6$. Lanes 1 through 6 shows plasmid DNA digested with *HindII* (1), *PvuII* (2), *PstI* (3), *HindIII* (4), *BglI* (5) and *EcoRI* (6). 1kb ladder was loaded as molecular weight marker (7).

FIG 16B Double digestions of clone $1E_6$. Lanes 1 through 7 contain plasmid DNA digested with *PvuII-PstI* (2), *PvuII* (3), *PvuII-HindIII* (4), HindIII (5) and *RsaI* (6). Molecular weight markers include 1kb ladder (1) and ϕX173-*HaeIII* digest (7).

FIG 16C The gel shown in 16B was hybridized with $(GATA)_4$. Lanes 1 through 5 correspond to lanes 2 to 6 described in figure 16B.

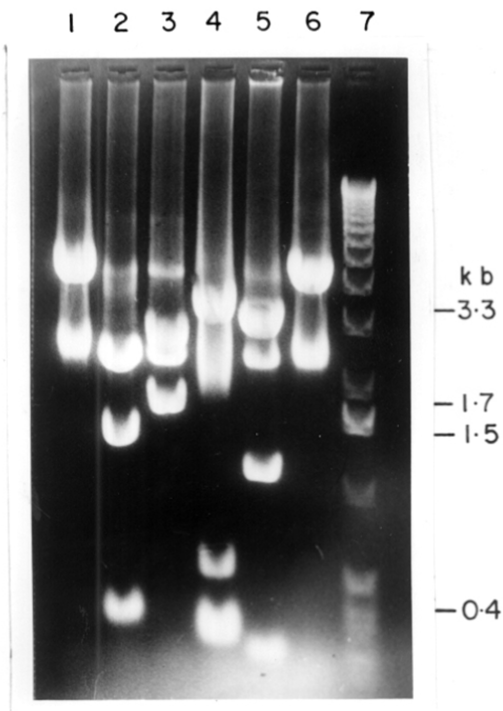


FIG. 16 A

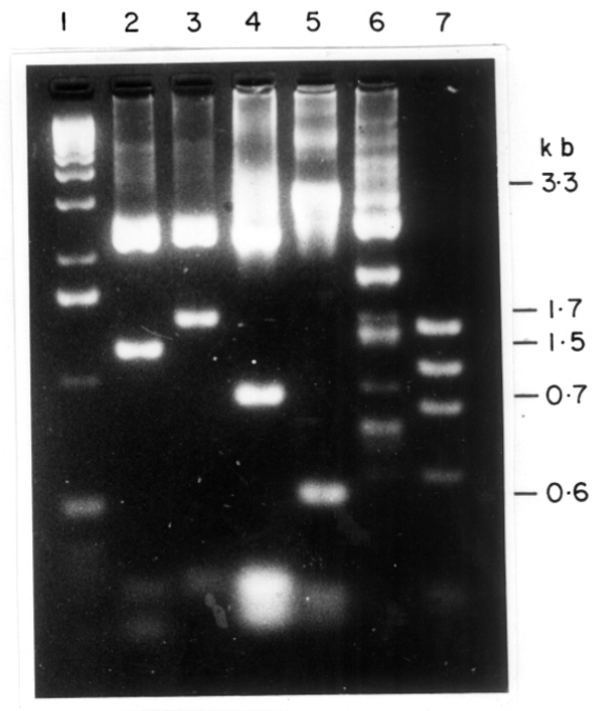


FIG. 16 B

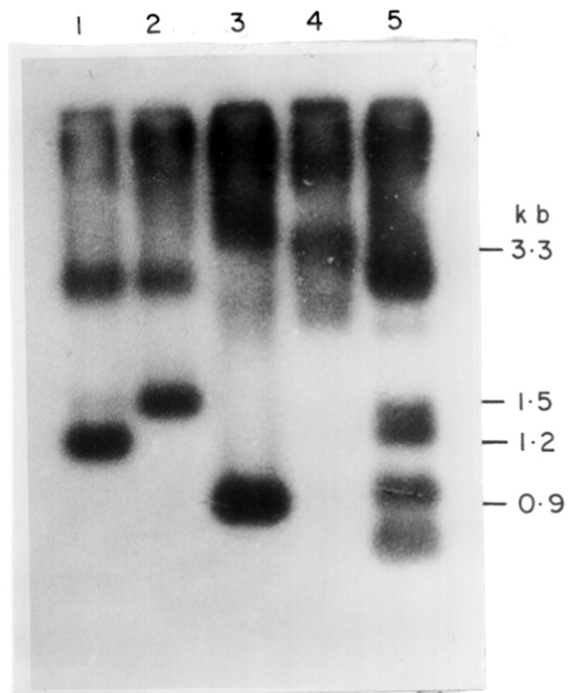


FIG. 16 C

FIG 17 Restriction map of the clone $1E_6$. The position of various repeats and the primers used for PCR amplification are shown in the figure.

0.2kb 0.7kb 0.8kb

PstI PvuII HindIII <= ++ ** PstI

—> <—
primer 1 primer 2

** Twelve (AG/TC) repeats

++ Ten (GATA/CTAT) repeats

=> Nine (AC/TG) repeats

Fig 17

FIG 18A Subclones of $1E_6$ digested with *PvuII* and *HindIII*. Lanes 1 through 6 contain plasmid DNA (1,2) digested with *PvuII* (3,4) and *HindIII* (5,6). Molecular weight markers are lambda-*HindIII* and ϕX173-*HaeIII* digests(7)

FIG 18B Plasmid DNA of $1E_6$ subclones digested with *PvuII* and *HindIII* and hybridized to $(GATA)_4$. Lanes 1 through 6 are as described in figure 18A.

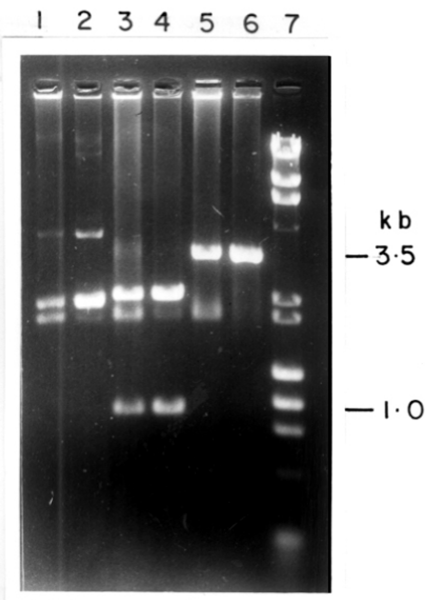


FIG.18 A

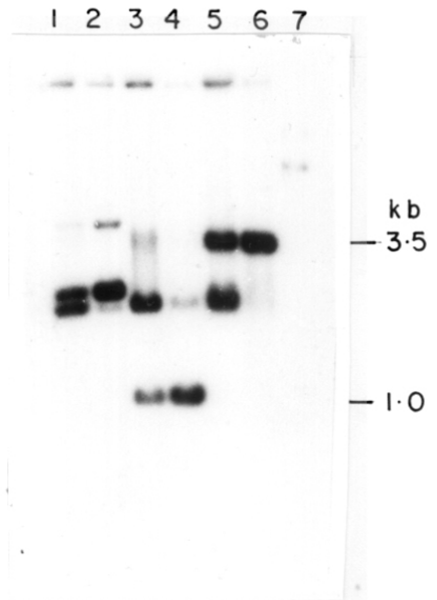


FIG.18 B

FIG 19 Partial DNA sequence of clone $1E_6$ showing the presence of simple repeats. Bold letters indicate the position of these repeats

5‘-CCTACACAAGTTTCTGTTGGGCAAAGCGAATAGTACAATTACACATCAAACACA
CCCTCACAC**AGAGAGAGAGAGAGAGAGAGAGAGAG**AATCCCTAGAACACCTGCTT
AATCAACAGGTTACAGCTTAGGGTTGTGGCAGAGAACATATATATATAGCAACAT
TTATGCCTCTCCAGCATAGCACTGCGTATATCTAGCAGCAGTAGAGGGAGATGAT
CACACTGCCAGTTAATCACGACAAAGCACACCG**GATAGATAGATT**CAGG**GATAG**
ATAAATAGATAGATAGATAGATTTGATTGATTTGAGGAAGAAGGACGTAGGT
GGTGGATATATAGGTAGATAGATCGATGTAGACATGTGGTAGTAGTAGTGGCAC
GA**ACACACACACACTCACAC**TAGCATGGCTGCATGACAGAAGAG -3‘

FIG 19

of variable number of tandem repeats present in them. These variations/ polymorphisms can be detected by using primers flanking the microsatellite sequences for PCR amplification and by separating the amplified products on DNA sequencing gel in order to detect few base pair variations. Different rice genotypes as described in figure legends including indicas, japonicas, wilds and somaclonal variants with respect to yield were used for PCR amplifications using unique primers 1 and 2 (Fig 17) flanking GATA and TG repeats from the DNA sequence of $1E_6$ clone (Fig 19). A 215bp fragment was amplified from all rice genotypes except *O.minuta* and low yielding somaclonal variant SC2 (Fig 20).

To detect genetic variation on an extraordinarily fine scale, the PCR amplified product was digested with *Sau3AI* and loaded on a 6% sequencing gel (Fig 21). The following observations were made

[1] Two strong bands at 155bp and 139bp appear to be common in all the genotypes, while a prominent 157bp band was present only in SCE1 and Indrayani.

[2] A 95bp fragment was present only in SCE1 and Indrayani and absent in others while a 73bp band was common in all.

[3] A single band at 65bp was present only in Indrayani and SCE1 whereas an additional band at 67bp was present only in SCE1.

The presence of many bands in *Sau3AI* digests of PCR amplified products can be explained by copy number variation amongst different rice genotypes, with *Sau3AI* sites being created in some copies of the amplified PCR product.

FIG 20 PCR amplified products from rice genotypes using primers flanking GATA and TG simple sequence repeats at $1E_6$ locus. Lanes 1 through 24 contain PCR amplified product from Hakkoda (1), Taichung-65 (2), Norin-49 (3), Adt-27 (4), Indrayani (5), Ambemohr (6), vargalsal (7), Tambdarambhog (8), a high yielding somaclone SCE1 (9), low yielding somaclones SC2 (10), SC5 (11), cultivar pranava (12), wild rice accession of *O.granulata* (13), *O.brachyantha* (14), *O.austaliensis* (15), *O.eichengeri* (16), *O.officinalis* (17), *O.minuta* (18), *O.punctata* (19), *O.latifolia* (20), *O.malampuzhansis* (21), *O.longistaminata* (22), *O.nivara* (23) and *O.rufipogon* (24). Molecular weight in bp is given in the right margin.

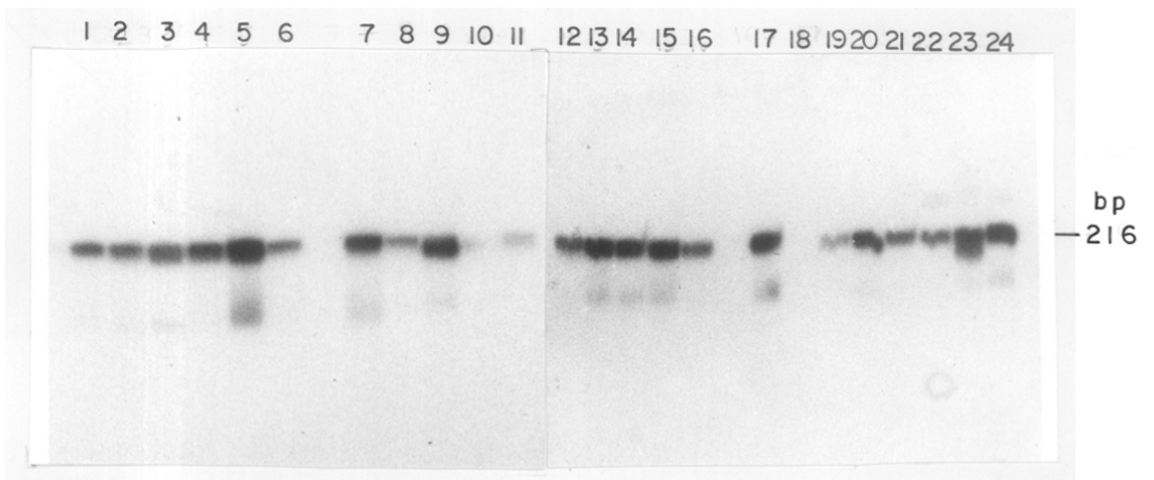


FIG. 20

FIG 21 Rice genotypes were PCR amplified as described in figure 20 and digested with *Sau3AI*. Lanes 1 through 13 contain *O.rufipogon* (1), *O.longistaminata* (2), *O.malampuzhansis* (3), *O.officinalis* (4), *O.australiensis* (5), *O.brachyantha* (6), *O.granulata* (7), SC5 (8), SCE1 (9), vargalsal (10), Indrayani (11), Norin-49 (12), Taichung-65 (13). Molecular weight was calculated from M13 sequence (A & G) (lanes 14 & 15).

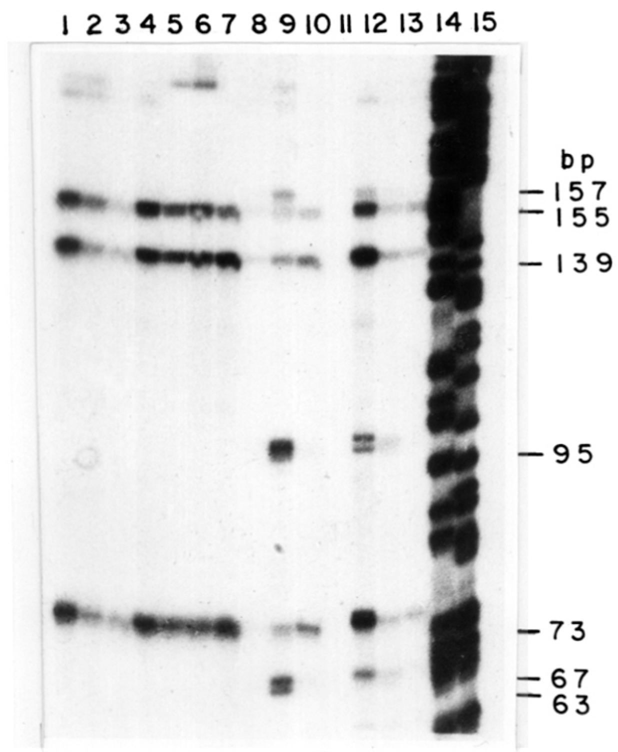


FIG. 21

REFERENCES

[1] Tsao SGS, Brunk CF, Pearlman RE (1983) Anal Biochem 131 : 365

[2] Kashi Y, Tikochinsky Y, Genislav E, Iraqi F, Nave A, Beckman JS, Gruenbaum Y, Soller M (1990) Nucleic Acids Res 18 : 1129

[3] Georges M, Lequarre AS, Catelli M, Henset R, Vassart G (1988) Cytogenet Cell Genet 47 : 369

[4] Zischler R, Nanda J, Schafer R, Schmid M, Epplen JT (1989) Hum Genet 82 : 227

[5] Georges M, Cochaux P, Lequarre AS, Young MW, Vassart G (1987) Nucleic Acids Res 15 : 7195

[6] Burr B, Burr FA, Matz EC, Saverson JR (1990) Plant Cell 4 : 953

[7] Rogstad SH, Patton II JC, Schaal BA (1988) Proc Natl Acad Sci USA 85 : 9176

[8] Ramakrishna W, Lagu MD, Gupta VS, Ranjekar PK (1994) Theor Appl Genet (In Press)

[9] Wu KS, Tanksley SD (1993) Mol Gen Genet 241 (1-4) : 225

[10] Glaszmann JC (1987) Theor Appl Genet 74 : 21

[11] Pintor Toro JA (1987) Biochem Biophys Res Comm 147 : 1082

[12] Dhar MS, Pethe VV, Gupta VS, Ranjekar PK (1990) Theor Appl Genet 80 : 402

[13] Haymer DS, Mcinnis DO, Arcangeli L (1992) Genome 35 : 529

CHAPTER 4

DISCUSSION

4.1 Detection of polymorphism using mini and microsatellite probes

4.2 Detection of fine variations at a microsatellite locus

4.3 Transcription and methylation of mini and microsatellites

4.4 Applications of mini and microsatellite probes in plant breeding

4.1 Detection of polymorphism using mini and microsatellite probes

Using morphological and isozyme markers, genetic differentiation of cultivars and construction of linkage maps have been reported in rice (1,2,3,4). The use of restriction fragment length polymorphisms (RFLPs) has proved to be of great value in detecting and manipulating genetic variation and construction of linkage maps (5,6). RFLP markers provide a distinct advantage over morphological markers in that they are not affected by environment and stage of development. The use of tightly linked RFLP markers can serve as a valuable tool for marker assisted breeding programs (7).

The use of ten base oligonucleotide primers to amplify specific target regions in the genome using PCR to generate polymorphisms constitute the Random Amplified Polymorphic DNAs (RAPDs). Fikuoko *et al.* (8) have reported the use of RAPDs for identification of rice accessions. RAPDS involve simpler technology, lack of radioactive work and high level of polymorphism. RAPD markers are inherited in a Mendelian fashion and although are dominant, have been used for efficient mapping in Arabidopsis (9). Compared to conventional RFLPs which are generally diallelic, micro and minisatellites have assumed great importance as they detect multiple loci/alleles which increase the polymorphic information content. The immense potential of microsatellite and minisatellite generated DNA fingerprints has been realized to some extent in humans as described in Chapter I. However, the potential remains unexplored mostly in the case of plants.

We have shown for the first time that minisatellites, microsatellites and several other multilocus probes are present in the rice genome and further support the concept of ubiquitous appearance of these DNA sequences in eukaryotes. Although the occurrence of sequences homologous to M13, R18.1 and simple

repeats has been shown in plants (10,11,12), there is no report about demonstrating the use of human minisatellite probe pV47 and a *rec* BCD dependent recombination signal in *E.coli* i.e. *chi* sequence in detection of polymorphism in plants with special reference to rice. The ability of the *chi* sequence repeat to decipher polymoprhism in rice is comparable to that in humans (13).

The probability of identical fragment profiles by chance using probes reported in this study compares quite favourably when chance occurrence of identical fragment patterns in rice cultivars with Jeffreys probe 33.6 is considered which is about 10^{-11} (14). Therefore, we show in this work a new range of molecular genetic markers which can be efficiently used in cultivar identification and rice breeding programs in addition to the RFLP markers presently available. It is clear from our work that micro and minisatellite probes hybridize to many restriction fragments (multiple loci) and reveal DNA fingerprints to detect genetic variability in rice.

The differences in the band patterns, as observed from our data, represent variations in repetitive DNA sequences in rice. It is yet unclear as to why some repeats are organized differently and are more polymorphic than others. R18.1 containing poly-TG repeats provides distinct polymorphic banding patterns as compared to other probes. Individual stretches of poly-TG are often characterized by highly polymorphic length variation due to variable number of tandem repeats of the TG unit. Poly-TG has a number of unique physico-chemical properties (15,16) and may serve regulatory functions (17). Mini and microsatellite variability is thought to result from unequal crossing over and replication slippage (18). The biological function and the conserved nature of the minisatellite core sequences, both in eukaryotes and even in prokaryotes where repeats constitute a small portion of the genome, remain largely unknown. However, their homology

to the *chi* sequence of *E.coli* and bacteriophage lambda has led to speculation that minisatellites may be involved in homologous recombination as shown in humans (19,20). Most of the probes used in our study have homology with the *chi* sequence of *E.coli* and the breakpoints of oncogene translocations (21). The average similarity index value ($\bar{X}_D$) (Table 3 & 5) shows that a relatively high level of polymorphism (in terms of fingerprint identity) has been maintained in cultivated rice. The potential of these probes, even individually to distinguish rice genotypes, is comparable with the probability of obtaining identical fingerprints by chance using M13 probe, Jeffreys probes and oligonucleotide probes in other eukaryotes reported so far. However, the fact that micro and minisatellite derived DNA fingerprints do not show variations between individuals of the same cultivar or between different tissues indicates rather slow turnover rate of these repeats which makes them suitable for application in breeding programs.

4.2 Detection of fine variations at a microsatellite locus

Locus specific amplifications which were initiated first in humans (22,23) were later extended to plants (24). This involves screening a genomic library, sequencing of the clones, designing of primers and PCR amplification. The utility of this approach in construction of genetic linkage maps and detection of mutations at disease causing loci in humans has been well documented. Recently, this method has been successfully applied to several plants including rice as described in Chapter I. However, the polymorphism detected by this approach was very limited in our studies. The detection of variability within minisatellite alleles based on the presence or absence of *HaeIII* site in the PCR amplified allele has tremendously increased the sensitivity of this technique (20). We have used a similar technique to detect high levels of genetic variation at the $1E_6$ specific locus in rice. The digestion of PCR amplified fragment with *Sau3AI*

(or) *HinfI* detected distinct polymorphism amongst various rice genotypes. These polymorphisms from many such loci in rice can be utilized for various applications in plant breeding programs.

4.3 Transcription and methylation of mini and microsatellites

Simple repeats seem to be hybridizing to heterogeneously sized RNA as reported in mouse and canine (25,26). Hershfield *et al.* (26) have shown the presence of tissue specific transcripts in canines homologous to the probe $(CAC)_n$. A prominent polyadenylated message at 6kb, complementary to $(CAC)_n$, was present in the retina, spleen and kidney but absent in the heart or liver of canine. Oligomers represent simple tandem repeats and one such repeat $(CAC)_5$ reveals tissue specific differences in rice. Analogous to the role proposed for other repetitive sequences interspersed with single copy DNA which have been hypothesized to be involved in gene regulation and evolution by formation of RNA-RNA duplexes and control the production of messenger RNA (27), a similar role can be envisaged for $(CAC)_5$ in rice.

Mini and microsatellites which serve as highly informative molecular markers have been suggested to be hot spots of recombination in humans (19,28,29,18). Similarly, methylated cytosines are also known to serve as recombinational hot spots (30,31). Thus, a high level of methylation at mini-satellite and microsatellite loci in rice further supports the concept of involvement of minisatellites in recombination. However, the differential methylations observed in different rice genotypes with micro and minisatellite probes points to a specific role in cellular function. Methylation of these loci may have a role in altering the chromatin structure thereby regulating gene expression (32).

4.4 Applications of mini and microsatellite probes in plant breeding

The results presented here that oligonucleotide fingerprinting in rice can serve as a powerful tool for the identification of different rice genotypes as most of the genome can be investigated using a collection of different oligonucleotide probes. Such analysis will assist in characterization and evaluation of rice genetic resources, testing the homogeneity of inbred lines and identification of somatic hybrids and chromosome addition lines. DNA fingerprinting can be of great value in monitoring and aiding gene introgression from wild rices into modern cultivars. As DNA fingerprint loci (DFP) are extremely polymorphic, genotypes having maximum similarity to recipient line and minimal similarity to donor line can be selected in crosses involving wild and cultivated genotypes (33). This ensures large number of heterozygous DNA fingerprint loci thereby reducing the required number of backcross generations in gene introgression breeding programs. Such DNA fingerprint loci can also aid in removal of both marker linked donor DNA flanking the gene of interest (to minimize the linkage drag) and unlinked DNA (7). Correlation between variability at the DNA level measured by RFLP and heterosis has proved to be of great value in allocating maize genotypes to different heterotic groups (34,35). Similarly, by using mini and microsatellite generated DNA fingerprints, rice varieties with good combining ability can be identified and the success of the crosses can be predicted without evaluation of F_1 hybrids in the field to predict heterotic combinations. Linkage between RFLP markers and quantitative trait loci (QTL) for improvement of an elite hybrid has been reported in maize (36). Similar linkage has been established between DNA fingerprint bands and QTL in chickens (37) and hence DNA fingerprinting can serve as an efficient method for such analysis. In this aspect, our study is of significant importance as most of the probes show extensive interspecific variability. Alleles of minisatellite loci are best suited for tagging the entire genome as they are scattered throughout the genome (38,33). Recently,

the use of microsatellites in plant breeding programs has been proceeding at a rapid pace, but it is a longway to go before the unexplored potential is totally exploited, especially in the area of marker assisted breeding.

REFERENCES

[1] Glaszmann JC (1988) Genome 30 : 782

[2] Oka HI (1953) Jpn J Breed 3 : 33

[3] Wu KS, Glaszmann JC, Khush GS (1988) Biochem Genet 26 : 303

[4] Kinoshita T (1986) In : Rice Genetics, Proc Int Rice Genet Symp,May, IRRI, Phillipines, 215

[5] McCouch S, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD (1988) Theor Appl Genet 76 : 815

[6] Saito A, Yano M, Kishimoto N, Sugiura M (1991) Jpn J Breed 41 : 665

[7] Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) Bio\technology 7 : 257

[8] Fikuoka S, Hosaka K, Kamijima O (1992) Jpn J Genet 67 : 243

[9] Reiter RS, Williams JGK, Feldmann KA, Rafalski JA, Tingey SV, Scolnik PA (1992) Proc Natl Acad Sci 89 : 1477

[10] Rogstad SH, Patton II JC, Schaal BA (1988) Proc Natl Acad Sci USA 85 : 9176

[11] Tzuri G, Hillel J, Lavi U, Haberfeld A, Vainstein A (1991) Plant Science 76 : 91

[12] Bayermann B, Nurnberg P, Weihe A, Meixner M, Epplen JT, Borner T (1992) Theor Appl Genet 83 : 691

[13] Ehtesham NZ, Das A, Hasnain SE (1992) Gene 111 : 261

[14] Dallas JF (1988) Proc Natl Acad Sci USA 85 : 6831

[15] Haniford DB, Pulley blank DE (1983) Nature 302 : 632

[16] Htun H, Dahlberg JE (1989) Science 243 : 1571

[17] Braatan DC, Thomas JR, Little RD, Dickson KR, Goldberg I, Schlesinger D, Ciccodicola A, D'Urso M (1988) Nucleic Acids Res 16 : 865

[18] Jarman AJ, Wells RA (1989) Trends in Genet 5 : 367

[19] Wahls WP, Wallace LJ, More PD (1990) Cell 60 : 95

[20] Jeffreys AJ, Neumann R, Wilson V (1990) Cell 60 : 473

[21] Krowczynska AM, Rudders RA, Krontiris G (1990) Nucleic Acids Res 18 : 1121

[22] Weber JL, May PE (1989) Am J Hum Genet 44 : 388

[23] Tautz D (1989) Nucleic Acids Res 17 : 6463

[24] Akkaya MS, Bhagwat AA, Cregan PB (1992) Genetics 132 : 1131

[25] Schafer R, Boltz E, Becker A, Bartels F, Epplen JT (1986) Chromosoma 93 : 496

[26] Hershfield B, Inouye L, Chader G, Ripps H, Aquirre G (1991) J Heredity 82 : 251

[27] Davidson EH, Posakony JW (1982) Nature 297 : 633

[28] Jeffreys AJ, MacLeod A, Tamaki K, Neil DL, Monckton DG (1992) Nature 354 : 204

[29] Chandley AC, Mitchell AR (1988) Cytogenet Cell Genet 48 : 152

[30] Coulondre C, Miller JH, Farabough PJ, Gilbert W (1978) Nature 274 : 775

[31] Doerfler W (1983) Ann Rev Biochem 52 : 93

[32] Selker EU (1990) Trends in Biochem Sci 103

[33] Lee M, Godshalk EB, Lamkey KR, Woodman WW (1989) Crop Sci 29 : 1067

[34] Hillel J, Schaap T, Haberfeld A, Jeffreys AJ, Plotzky Y, Cahaner A, Lavi U (1990) Genetics 124 : 783

[35] Godshalk EB, Lee M, Lamkey KR (1990) Theor Appl Genet 80 : 273

[36] Zehr BE, Dudley JW, Chojecki J, Saghai Maroof MA, Mowers RP (1992) Theor Appl Genet 83 : 903

[37] Plotsky Y, Cahaner A, Haberfeld A, Lavi U, Lamont SJ, Hillel J (1993) 24 : 105

[38] Beckmann JS, Soller M (1988) Theor Appl Genet 76 : 228