Rice Genomics: Use of DNA markers for phylogenetic and hybrid performance analysis

THESIS SUBMITTED TO THE UNIVERSITY OF PUNE

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BIOCHEMISTRY

SWATI PRABHAKAR JOSHI

Plant Molecular Biology Unit,

Division of Biochemical Sciences,

National Chemical Laboratory,

Pune- 411008

JANUARY 2001

DECLARATION

Certified that the work incorporated in this thesis entitled “Rice Genomics: Use of DNA markers for phylogenetic and hybrid performance analysis,” submitted by **Ms. Swati Prabhakar Joshi** was carried out under my supervision. The material obtained from other sources has been duly acknowledged in this thesis.

(Vidya Gupta)

Research Guide

INTER- INSTITUTIONAL COLLABORATIVE RESEARCH EFFORT

Research work embodied in this thesis was carried out at:

National Chemical Laboratory
Pune, (Maharashtra), India

In collaboration with:

International Rice Research Institute (IRRI)
Philippines

Konkan Krishi Vidyapeeth (KKV)

Dapoli (Maharashtra) India

Scientist Involved From Each Institute

NCL: Dr. V.S. Gupta P. K. Ranjekar

IRRI: Dr. D. S. Brar

KKV: Dr. B. L. Dhonukshe

INDEX

ACKNOWLEDGEMENTS

It gives me a great pleasure as I take this opportunity to thank all those who have been directly or indirectly instrumental in making this endeavor a satisfying experience for me. It is not a big secret that any piece of work cannot be successfully concluded single-handed. Each and every person around me at home and in lab has in some way contributed in overcoming all the difficulties faced while fulfilling this task.

Working under the able guidance of Dr. Vidya Gupta was a great privilege. She has been more than a guide and her friendly and motherly attitude has always made interactions with her extremely enjoyable at all times. Her meticulous approach and analytical mind has always inspired me in dealing with problems instantly and with logical reasoning. I express my gratitude to her for her commitment and help which has greatly helped me in pursuing this work.

I received valuable guidance from Dr. Ranjekar. I thank him for his constant encouragement and interest in my work. He has been extremely particular in going through all the manuscripts and thesis and helped to mold them appropriately making them more pertinent.

During my work I have interacted with several scientists in India and abroad. Though it would be dificult to mention each and every one, I would like to reiterate in the warmest words my thanks to Dr. D. S. Brar for providing wild rice germplasm from IRRI. Collaborative efforts with him gave me a perspective in the field of “Phylogeny and Evolution of Rice”. I express my grateful thanks to Dr J P Gustafson, Dr.Aggarwal and Dr. McNally for going through my manuscripts and giving valuable suggestions for their improvement. Discussions with Prof. E A Siddiq and Dr. Bapat were extremely useful in planning my experiments especially on hybrid work. I am thankful to Dr. Dhonukshe and Dr. Bhave KKV, for performing crosses for hybrid studies and providing the field data for replicated trial of F_1 hybrids as well as the leaf tissue of parental lines. Due credit lines appear at appropriate places in the text for all those whose names may not have appeared here.

I received valuable help from some of the best of friend I met at NCL. I do not have appropriate words to thank Aditi, Anjali, Aparna, Armaity, Bhushan, Suresh, Venkat and Vrinda.

I have also enjoyed the caring and cheerful companionship of Abhay, Ajay, Ajit, Aparna D, Archana Deshpande, Arundhati, Ashok, Ashok A, Bimba, Dipak, Gauri Apte, Jakir, Juri, Kanchan, Maneesha B, Manisha R, Madhumita, Meena M, Milind, Monali, Mukund, Rahul, Rajesh, Raju, Rakhee, Rangarajan, Renu, Renuka, Ritika, R K, Sami, Sanjay, Sastry, Shashi, Shubhada and Vijay.

My heartfelt thanks to Dr. Meena, Dr. Mohini, Dr. Lalitha, Dr. Nirmala, Mr. B.G. Patil and Ms. Usha, for their help and special thanks to Sadhana, Suvarna and Indira for their cooperation and help in handling all the official paper work. I appriciate the help extended by Karunakaran, Jagtap, Kamble, Karanjkar, Kamthe and Trehan for technical assistance for instrumentation.

I am greatly indebted to Dr. Paul Ratnaswami, Director NCL, for providing the lab facility during the tenure of my work. I acknowledge CSIR for providing financial assistance. Thanks to Rockefeller foundation for financial support.

I acknowledge all those friends and well wishers whom I have not mentioned here but have always been helpful at appropriate time.

Lastly this acknowledgement would be incomplete without expressing my indebtedness to my family members who have been a constant source of inspiration and support all through my life. My mother and the memories of my late father and grandfather have always guided me to achieve my ambitions. I thank Rahul, Kanchan, Vrinda, Devdatta, and Aji, who always stood by me. I thank my in-laws for their backing during the last stages of my work. Last but not the least I thank my husband Shrirang, who has been understanding supportive and helpful in every way to see my work come through.

Swati Joshi

List of Abbreviations

AFLP	Amplified Fragment Length Polymorphism
AP-PCR	Arbitrary Primed Polymerase Chain Reaction
ASAP	Allele Specific Associated Primers
bp	Base pairs
CAPS	Cleaved Amplified Polymorphic sequences
cDNA	Complementary Deoxyribo Nucleic Acid
cM	Centi Morgan
CMS	Cytoplasmic Male Sterility
CMS 25A	Cytoplasmic Male Sterile Line IR 58025 A
CMS 29A	Cytoplasmic Male Sterile Line IR 62829 A
DAF	DNA Amplification Fingerprinting
DAMD	Directed Amplification of Minisatellite- region DNA
dATP	Deoxyadenosine 5' triphosphate
dCTP	Deoxycytidine 5' triphosphate
dGTP	Deoxyguanosine 5' triphosphate
dNTP	Deoxynucleotide 5' triphosphate
dTTP	Deoxythymidine 5' triphosphate
DNA	Deoxyribose Nucleic acid
EDTA	Ethylene Diamine Tetra Acetic acid
EST	Expressed Sequence Tags
GD	Genetic Distance
HVR	Hyper Variable Repeats
IRRI	International Rice Research Institute
ISSR	Inter Simple Sequence Repeats
kb	kilobases
MAS	Marker Assisted Selection
µl	Microliter
ml	Milliliter
mM	Millimolar
NIL	Near Isogenic Lines
ng	nanograms

PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
RAMPO	Randomly Amplified Microsatellite Polymorphism
RAPD	Randomly Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
SCAR	Sequence Characterized Amplified Region
SNP	Single Nucleotide Polymorphism
SSCP	Single Strand Conformation Polymorphism
STMS	Sequence Tagged Microsatellite site
SSR	Simple Sequence Repeats
STS	Sequence Tagged Sites
STR	Short Tandem Repeats
TAE	Tris Acetate EDTA Buffer
TBE	Tris Borate EDTA Buffer
Tm	Melting Temperature
Tris	Tris Hydroxymethyl Amino Methane
QTL	Quantitative Trait Loci
VNTR	Variable Number of Tandem Repeats
WA	Wild Abortive

List of Tables

Chapter 2: Section I

Chapter 2: Section II

Chapter 3

List of Figures

Chapter 2: Section I

Chapter 2: Section II

Chapter 3:

Figure 3: PCR amplification profile obtained using anchored microsatellite primer ISSR 856 [$(AC)_8$ YA].

Figure 4: Dendrogram generated by UPGMA cluster analysis based on specific heterozygosity markers.

Figure 5: Dendrogram generated by UPGMA cluster analysis based on ISSR genotype profiles.

Figure 6 to 13:

Graphs showing the relationship between genetic distance and hybrid performance, heterosis and heterobeltiosis in CMS IR 58025 A and CMS IR 62829 A, crosses with the 14 restorer lines for all the traits being studied.

Thesis Abstract

Genetic variation in the germplasm is extremely important for the improvement of crop plants, be it through traditional breeding practices or through molecular approaches. Since phenotypes are not always effective means for identifying general variability or specific heterozygosity, molecular markers are now looked upon as convenient tools not only for genotyping and variability analysis, but also for mapping various traits, cloning important genes and identifying appropriate sources for gene introgression.

In our laboratory, rice genome analysis was an important theme of research for over a decade. In continuation with the ongoing efforts, I undertook my thesis work with following objectives:

1. To study the utility of molecular markers (Inter simple sequence repeats) for assessing genetic variations in the genus *Oryza* and phylogenetic organization of the germplasm being analyzed.
2. To identify variable alleles for members of cyclophilin gene family and other CAC microsatellite containing ESTs in the genus *Oryza.*
3. To exploit molecular tools for prediction of hybrid performance, thereby determining their utility in selection of parents for a three-line hybrid system.

Important Research Findings of My Work Are:

1. Demonstration of Utility of ISSR Polymorphism in Fingerprinting of Cultivated and Wild Species Germplasm and in Understanding Evolutionary Relationships in *Oryza*

Inter simple sequence repeat polymorphism (ISSR) was used to determine genetic diversity and phylogenetic relationships in *Oryza*. Forty-two genotypes including seventeen wild species, representing *AA, BB, CC, EE, FF, GG, BBCC, CCDD*, and *HHJJ* genomes, two cultivated species, *O.sativa* (AA) and *O.glaberrima* (AA), and three related genera, *Porteresia coarctata, Leersia* and *Rhynchoryza subulata* were used in ISSR analysis.

A total of 30 ISSR primers were screened representing di, tri, tetra and penta nucleotide repeats of which 11 polymorphic and informative patterns were

selected to determine the genetic diversity. The consensus tree constructed using binary data from banding patterns generated by ISSR-PCR could cluster 42 genotypes according to their respective genomes. ISSR analysis suggested that the genus *Oryza* might have evolved following a polyphyletic path, O.*brachyantha* (FF genome) was the most divergent species in *Oryza* and *O.australiensis* (EE genome) did not fall under the *Officinalis* complex. DNA profiles based on ISSR markers revealed potential diagnostic fingerprints for various species, genomes and also for individual accessions / cultivars. ISSR also revealed 87 putative genome / species specific molecular markers for 8 of the 9 genomes of *Oryza*.

2. Detection of Genetic Variability Among $(CAC)_n$ Containing ESTs in *Oryza*

Based on previous work in our laboratory, microsatellite motif $(CAC)_n$ was shown to be present in poly A^+ transcripts of rice. This motif showed a differential hybridization with heterogeneously sized poly A^+ RNA, specifically in rice endosperm tissue. In my work, I generated ESTs containing $(CAC)_n$ microsatellite motifs, by isolating expressing sequences from rice developing seed (endosperm) cDNA library. Genetic diversity and phylogeny of these ESTs were studied in forty-six genotypes of *Oryza* including wild and cultivated species representing *AA, BB, CC, EE, FF, GG, BBCC, CCDD* and *HHJJ* genomes, three related genera and four out-group members, wheat, maize, barley and oat. A total of six ESTs were analyzed, of which, three were found to be the members of the cyclophilin gene family and one was a gene containing Leucine zipper homeo domain. Upon sequence homology search, the remaining two were found to be totally unknown gene sequences.

The genetic variability among ESTs can be useful for finding new gene alleles, especially for genes like cyclophilin. This would be extremely useful because these genes play important role in the life cycle of rice plants like development, stress response and other house keeping functions and provide target gene sources for future introgression breeding programs for crop improvement.

3. Correlation of DNA Markers Based Genetic Distance and Prediction of Heterotic Combination in a Three Line Hybrid System

Two Cytoplasmic Male Sterile lines were crossed with fourteen restorer lines of rice widely grown in the western regions of Maharashtra, India, to produce 28 F_1 hybrids and were evaluated for eight agronomically important traits, contributing to yield potential, in replicated field trials. The hybrid performance was recorded along with heterosis and heterobeltiosis. All the rice lines under investigation were subjected to marker-based variability analysis. An attempt was made to correlate genetic distance based on random markers and specific markers for each trait individually, as well as average genetic distance based on all specific markers, with hybrid performance and heterosis, by regression analysis.

The results indicated that ISSR primers yielded fingerprints and putative variety-specific bands. While, specific markers could cluster the parental lines in different groups and showed significant correlation with hybrid performance, the data also supported the proposition that epistasis is the basis of heterosis however, it revealed lack of significant predictive values for field application.

Chapter I

Review of Literature

Molecular Markers and their Utility in Crop Improvement Programs

Improvement in Rice Production: A Challenge to Break Yield Barriers

Rice is one of the most important members of the *Gramminae* family and forms a staple diet of more than 3rd of the world population, imparting it agronomically and nutritionally important status among the cereal crops. With Green Revolution, a series of high yielding rice varieties became available and, most of the rice growing countries in Asia where almost 92% of the world's rice is grown, became self sufficient in rice production (Lampe 1991). However, a fear of food shortage looms again as the release of high yielding crops has reached a plateau and now the major challenge faced by today's breeders is breaking the yield barrier. Biotechnology is looked upon as one of the potential tools to significantly strengthen and complement the rice- breeding program. During the last decade, rice biotechnology has gained importance and a useful knowledge database on rice molecular genetics is becoming available which can be used as a tool for rice improvement programs. Breeders and biotechnologists have begun to concentrate their efforts on modifying the useful traits of rice like resistance to insects and diseases, tolerance to biotic and abiotic stresses, high yield and better nutritional quality, in diverse environments. One of the most recent examples is "Golden rice" which produces Pro-vitamin A (β carotene) imparting a high nutritional value (Ye *et al* 2000) to rice.

Traditionally plant breeding involves two phases, viz. evolutionary phase and evaluation phase (Toenniessen and Khush 1991). In the evolutionary phase, the time- tested method of hybridization is used to create variability (on some occasions mutation induction may also be used), while, in the evaluation phase, variable germplasm is selected from the above population by various

methods such as bulk population breeding, pedigree selection, recurrent selection and single seed descent. Biotechnology tools will not be able to replace these time tested methods, but can definitely enhance their efficiency. For example, wide hybridization through embryo rescue / somatic hybridization / somaclonal variation / genetic engineering can be useful in evolutionary phase, and molecular markers linked to genes of economic importance can aid and increase the efficiency of evaluation phase. The marker-based selection holds particularly high promise for monitoring the traits that are governed by quantitative trait loci (QTLs) and those that are difficult to evaluate with presently available techniques. Since the main focus of my work has been on evaluation of DNA markers for certain aspects of rice genome analysis, I have highlighted in this chapter, their applications with special reference to phylogeny and evolution, diversity analysis of exotic germplasm and prediction of heterotic combinations in hybrid crop breeding programs.

DNA Markers: A Boon to Rice Researchers

Introduction to DNA Markers

Genetic polymorphism is classically defined as the simultaneous occurrence of two or more discontinuous variants or genotypes of a trait in the same population. Although DNA sequencing is a straightforward approach for identifying variations at a locus, it is very expensive and laborious. A wide variety of techniques have, therefore, been developed in the past few years for visualization of DNA sequence polymorphism.

The term DNA-fingerprinting was introduced for the first time by Sir Alec Jeffery in 1985, to describe bar-code-like DNA fragment patterns generated by hybridization with multilocus probes after electrophoretic separation of genomic DNA fragments. The emerging patterns make up a unique feature of the analyzed individual and are currently considered to be the ultimate tool for biological individualization. Recently DNA fingerprinting / profiling term is used to describe the combined use of several single locus detection / DNA marker systems which are being used as versatile tools for investigating various aspects of plant genomes. These include characterization of genetic variability, genome fingerprinting, genome mapping, gene localization, analysis of genome

evolution, population genetics, taxonomy, plant breeding, and diagnostics.

Some of the desirable properties for ideal DNA markers are:

- Highly polymorphic nature
- Co-dominant inheritance (determination of homozygous and heterozygous states of diploid organisms)
- Frequent occurrence in genome
- Selective neutral behavior (The DNA sequences of any organism are neutral to environmental conditions or management practices)
- Easy access (availability)
- Easy and fast assay
- High reproducibility
- Easy exchange of data between laboratories

It is extremely difficult to find a molecular marker, which would meet all the above criteria. Hence depending on the type of study to be undertaken, a marker system can be identified, that would fulfill at least a few of the above characteristics (Weising *et al* 1995).

Various types of molecular markers are generally classified as hybridization based markers and polymerase chain reaction (PCR) based markers. In the hybridization-based markers, the DNA profiles are visualized by hybridizing the restriction enzyme digested DNA, to a labeled probe, which is a DNA fragment of known origin or sequence. PCR based markers involve *in vitro* amplification of particular DNA sequences or loci, with the help of specifically or arbitrarily chosen oligonucleotide sequences (primers) and a thermostable DNA polymerase enzyme. The amplified fragments are separated electrophoretically and banding patterns are detected by different methods such as staining and autoradiography. PCR is a versatile technique that was invented during mid 1980s (Saiki *et al* 1985), while the thermostable DNA polymerase was introduced in 1988 (Saiki *et al* 1988). Owing to its sensitivity and high speed, PCR has opened up a multitude of new possibilities in molecular biology research and clinical laboratories.

Occurrence of Wide Spectrum of DNA Markers

Single or Low Copy Probes

Restriction Fragment Length Polymorphism (RFLP)

RFLPs are simply inherited Mendelian characters. They have their origin in the DNA rearrangements that occur due to the evolutionary processes, point mutations within the restriction enzyme recognition site sequences, insertions or deletions within the fragments and unequal crossing over (Schlotterer and Tautz 1992).

In RFLP analysis, restriction enzyme digested genomic DNA is resolved by gel electrophoresis and then blotted (Southern 1975) on to a nitrocellulose membrane. Specific banding patterns are then visualized by hybridization with a labeled probe. The latter are mostly species specific single locus probes of about 0.5-3.0 kb in size and are obtained from a cDNA or a genomic library. The appropriate source for RFLP probe varies with the requirement of an application. Though it seems that genomic library probes may exhibit more variability than the gene probes from cDNA libraries, a few studies reveal the converse (Landry and Michelmore 1987, Miller and Tanksley 1990). The cDNA probes not only detect variation in the coding regions of the corresponding genes, but also the regions flanking the genes and the introns of the gene.

Botstein *et al* (1980) used RFLPs for the first time to construct a genetic map. RFLPs being co-dominant markers can detect coupling phase of DNA molecules, as DNA fragments from all homologous chromosomes are detected. They are very reliable and informative markers in linkage analysis and breeding, as they can easily determine if a linked trait is present in a homozygous or heterozygous state in an individual, an information highly desirable for recessive traits (Winter and Kahl 1995). The utility of RFLPs has been hampered due to the large amount of DNA required for restriction digestion and southern blotting. Also, the requirement of radioactive isotope or non-radioactive kit makes the analysis relatively hazardous and / or expensive, respectively. The assay is time consuming and labor-intensive and only one out of several markers may be polymorphic, which renders them inconvenient

especially for crosses between closely related species. Their inability to detect single base changes, restrict their use in detecting point mutations occurring within the regions at which they are detecting polymorphism.

RFLP Markers Converted into PCR Based Markers

1] Sequence Tagged Sites (STS)

RFLP probes specifically linked to a desired trait can be converted into PCR based STS markers based on the nucleotide sequence of the probe giving a polymorphic band pattern, to obtain specific amplicon. Using this strategy, the tedious hybridization procedure involved in RFLP analysis can be overcome. This approach is extremely useful for studying the relationship between several species (Bustos *et al* 1999). When these type of markers are linked to some specific traits, they can be easily integrated into plant breeding programs for marker assisted selection of the trait of interest.

2] Allele Specific Associated Primers (ASAP)

To obtain an allele specific marker, specific allele (either in homozygous or heterozygous state) is sequenced and specific primers are designed for amplification of DNA template to generate a single fragment at stringent annealing temperature. These markers tag specific alleles in the genome and are more or less similar to SCARs (Gu *et al* 1995).

3] Expressed Sequence Tag Markers (EST)

Adams et al (1991) first introduced the term “Expressed Sequence Tags”, which are obtained by partial sequencing of random cDNA clones. Once generated, they are useful in cloning specific genes of interest and synteny mapping of functional genes in various related organisms. ESTs are being used in full genome sequencing and mapping programs, which are underway for a number of organisms, for identifying active genes and isolating new genes.

4] Single Strand Conformation Polymorphism (SSCP)

This is a powerful and rapid technique for gene analysis particularly for detection of point mutations and for typing of DNA polymorphism (Orita *et al*

1989). SSCP can identify heterozygosity of DNA fragments of the same molecular weight and can even detect changes of a few nucleotide bases as the mobility of the single stranded DNA changes with the change in its GC content, due to its conformational change. Furthermore, to over come, the problems of re-annealing and complex banding patterns, an improved technique called asymmetric-PCR SSCP was developed (Ainsworth *et al* 1991). In this approach, the denaturation step was eliminated and a large sized sample could be loaded for gel electrophoresis making it a potential tool for high throughput DNA polymorphism, which is useful in detection of heritable human diseases. In plants, however, SSCP is not well exploited although its application in discriminating progenies can be extremely useful once suitable primers are designed for agronomically important traits (Fukuoka *et al* 1994).

Repetitive DNA

Microsatellites and Minisatellites

About 30-90% of the genomes of virtually all species is constituted of repetitive DNA, which are highly polymorphic in nature. These regions contain genetic loci comprising several alleles, differing from each other with respect to length, sequence or both. The repetitive DNA regions play an important role in absorbing mutations in a genome and inherited mutations are vital in evolution or polymorphism. Thus repetitive DNA and mutational forces together form the basis for a number of marker systems that are useful for several applications in plant genome analysis.

The major forms of repetitive DNA are microsatellites and minisatellites. The term, microsatellites was coined by Litt and Lutty (1989), while the term minisatellites, was introduced by Jeffrey (1985). They are multilocus probes creating complex banding patterns, usually non-species specific, and occur ubiquitously. The patterns generated by these probes are also known as oligonucleotide fingerprints. The methodology has been derived from RFLP and specific fragments are visualized by hybridization with a labeled micro or minisatellite probe.

Minisatellites are tandem repeats with a monomer repeat length of about 11-60

bp, while microsatellites or short tandem repeats / simple sequence repeats (STRs / SSRs) consist of 1 to 6 bp long monomer sequence that is repeated several times. These loci contain tandem repeats that vary in the number of repeat units between genotypes and are referred as variable number of tandem repeats (VNTRs) i.e.: a single locus that contains variable number of tandem repeats between individuals (Nakamura *et al* 1987). They are some times called hyper variable regions (HVRs) i.e.: numerous loci containing tandem repeats within a genome generating high levels of polymorphism between individuals (Jeffrey *et al* 1985). Microsatellites and minisatellites form an ideal marker system creating complex banding patterns by detecting multiple DNA loci simultaneously. They are dominant fingerprinting markers and co-dominant STMS (sequence tagged microsatellites) markers. They exist as many alleles in a population, their level of heterozygosity is high and they follow Mendelian inheritance.

Minisatellite and Microsatelite Sequences Converted into PCR Based Markers

a] Sequence Tagged Microsatellite Sites (STMS)

This method includes DNA polymorphism using specific primers designed from sequence data of a specific locus. Primers complementary to the flanking regions of the simple sequence repeat loci (Weber and May 1989) yield highly polymorphic amplification products. Polymorphisms appear because of variation in the number of tandem repeats (VNTR loci) in a given repeat motif. Tri and tetra nucleotide microsatellites are more popular for STMS analysis, because they present a clear banding pattern after PCR and gel electrophoresis (Hearne *et al* 1992). However, dinucleotides are generally abundant in the genomes and have been used for diversity analysis (Rafalski *et al* 1993).

2] Directed Amplification of Minisatellite-Region DNA (DAMD)

Heath *et al* (1993) introduced this technique for the first time where minisatellites were used as primers for DNA amplification. It has been explored as a means of generating DNA - Probes that are useful for detecting

polymorphism. The DAMD PCR product clones have yielded individual specific DNA finger printing pattern and thus have demonstrated the potential as markers for species differentiation and cultivar identification (Somers *et al* 1996).

3] Inter Simple Sequence Repeat Markers (ISSR)

Primers based on microsatellites are used in this type of analysis for amplifying inter-SSR DNA sequences. Zietkiewicz *et al* (1994) reported this technique for the first time. Where, microsatellites anchored at the 3' end are used for amplifying genomic DNA increasing their specificity. They are mostly dominant markers, though occasionally a few of them exhibit co-dominance. An unlimited number of primers can be synthesized for various combinations of di, tri, tetra and penta nucleotides [$(3)^3 = 27$, $(4)^4 = 256$] etc with an anchor made up of a few bases. I have extensively employed these markers in my work and exploited their power in producing putative diagnostic fingerprints (Chapter I: Section I).

Arbitrary Sequence Markers

Randomly Amplified Polymorphic DNA Markers (RAPD)

In 1990, Welsh and McClelland developed a PCR based genetic assay namely Randomly Amplified Polymorphic DNA (RAPD) in which nucleotide sequence polymorphisms is detected by using a single primer of arbitrary nucleotide sequence. In this reaction, a single species of primer anneals to the genomic DNA at two different sites on complimentary strands of the DNA template. If these priming sites are within the amplification range of each other, a discrete DNA product is formed through thermo-cyclic amplification. On an average, each primer directs amplification of several discrete loci in the genome, making the assay useful for efficient screening of nucleotide sequence polymorphism between individuals (Tingey *et al* 1993).

Some Variations in the RAPD Technique Include

A] DNA Amplification Fingerprinting (DAF)

Caetano- Anolles *et al* (1991) employed single arbitrary primers as short as 5

bases to amplify the DNA using polymerase chain reaction. In a spectrum of products obtained, the patterns that are simple are useful as genetic markers for mapping, while more complex patterns are useful for DNA fingerprinting. The band patterns are reproducible and can be analyzed, using polyacrylamide gel electrophoresis and silver staining. DAF requires careful optimization of parameters, however, it is extremely amenable to automation and fluorescent tagging of primers for early and easy determination of amplified products thus it has been useful in genetic typing and mapping.

B] Arbitrary Primed Polymerase Chain Reaction (AP -PCR)

This is a special case of RAPD, wherein discrete amplification patterns are generated by employing single primers of 10-50 bases in length in PCR of genomic DNA (Welsh and McClelland 1991). In the first two cycles, annealing is carried out under non-stringent conditions. The final products are structurally similar to RAPD products.

C] Sequence Characterized Amplified Regions for Amplification of Specific Band (SCAR)

Michelmore and co-workers (1991) and Martin *et al* (1991) introduced a procedure wherein the RAPD marker termini are sequenced and longer primers are designed (22-24 nucleotide bases long) for specific amplification of a particular locus. These are similar to STS markers (Olson *et al* 1989) in construction and application and are more reproducible compared to RAPD. The presence or absence of the band indicates a variation in the sequence. SCARs are usually dominant markers, however, some of them can be converted into co-dominant markers by digesting them with tetra cutting restriction enzymes and polymorphism can be deduced by either denaturing gel electrophoresis or single strand conformation polymorphism (Rafalski and Tingey 1993). Being co-dominant, SCARs are useful in mapping and map based cloning as they can be used to screen pooled genomic libraries by PCR. SCARs also allow comparative mapping or homology studies among related species, thus making this strategy an extremely adaptable concept in near future.

D] Cleaved Amplified Polymorphic Sequences (CAPS): Co-Dominant Markers

Here, the electrophoretic patterns are generated by restriction enzyme digestion of PCR products and are compared for their differential migration during electrophoresis (Koniecyzn and Ausubel 1993, Jarvis *et al* 1994). PCR primers for this process can be synthesized based on the sequence information available in databank of genomic or cDNA sequences or cloned RAPD.

E] Randomly Amplified Microsatellite Polymorphisms (RAMPO)

In this PCR based strategy, genomic DNA is first amplified using arbitrary (RAPD) primers. The amplified products are then electrophoretically separated and the dried gel is hybridized with microsatellite oligonucleotide probes. Several advantages of oligonucleotide fingerprinting (Epplen 1992), RAPD (Williams *et al* 1990) and Microsatellite primed -PCR (Weising *et al* 1995, Gupta *et al* 1994a) are thus combined in RAMPO. Some of the advantages include speed of the assay, high sensitivity, high level of variability detected and no requirement of prior DNA sequence information (Richerdson *et al* 1995).

F] Amplified Fragment Length Polymorphism (AFLP)

In 1993, Zabeau *et al* developed AFLP, which is based on detection of genomic restriction fragments by PCR amplification. The fingerprints are produced, without any prior knowledge of sequence, using a limited set of generic primers. The number of fragments detected in a single reaction can be 'tuned' by selection of specific primer sets. AFLP is reliable since stringent reaction conditions are used for primer annealing. Since it is an ingenious combination of RFLP and PCR, (Saiki *et al* 1988, Ehrlich *et al* 1991) it is extremely useful in detection of polymorphism between closely related genotypes.

Application of Molecular Markers in Some Specific Areas of Plant Biology, with Special Reference to Rice

Molecular markers have been looked upon as tools for a large number of applications ranging from localization of a gene to improvement of plant varieties by marker assisted selection. They have also become extremely

popular for phylogenetic analysis adding new dimensions to the evolutionary theories. If we look at the history of the development of these markers, it is evident that they have been improved over the last two decades so as to provide easy, fast and automated assistance to the scientists and breeders. Genome analysis based on molecular markers has generated a vast amount of information and a number of databases are now available for its preservation and popularization. Though molecular markers have been used in many crops (Tanksley *et al* 1989; Patterson et al 1991; Winter and Kahl 1995; Gupta et al 1996; Mohan *et al* 1997a; Ribaut and Hoisington 1998; Kumar et al 1999) I wish to summarize the information mostly with respect to rice.

Genotyping of Cultivars

Genotyping of individuals has been one of the very first applications of DNA markers for effective management and utilization of germplasm. RFLP probes derived from nuclear (Wang and Tanksley 1989; Zhang *et al* 1992; Ishii *et al* 1995; Qian *et al* 1995; Oba *et al* 1996) and chloroplast DNA (Ishii and Tsunewaki 1991) have been used to differentiate between indica and japonica rice cultivars. However, the repetitive and arbitrary DNA markers are looked upon as markers of choice for genotyping of cultivars. Microsatellites like $(CT)_{10}$, $(GAA)_5$, $(AAGG)_4$, $(AAT)_6$ (Yanagisawa *et al* 1994), $(GATA)_4$, $(CAC)_5$ (Ramakrishna *et al* 1994, Gupta *et al* 1994b) and minisatellites (Broun and Tanksley 1993, Dallas *et al* 1993, Zhou and Gustafson 1995, Ramakrishna *et al* 1995) have been successfully employed in these analyses. Apart from being used as hybridization based markers, microsatellites have been useful for generation of STMS markers. The latter were useful in revealing polymorphisms within closely related cultivars of indica and japonica rice (Wu and Tanksley 1993, Zhao and Kochert 1993, Yang *et al* 1994 and Panaud *et al* 1996), upland rice accessions (Tanh *et al* 1999) and white milled and brown milled rice cultivars (Bligh *et al* 1999). Besides genotyping, microsatellites play an important role for protection of proprietary germplasm and have also been advantageous in pedigree analysis as they represent specific locus. The multi allelism of these markers facilitates comparative allelic variability detection across a wide range of germplasm and allows individuals to be ubiquitously genotyped, so that gene flow and paternity can be established.

Apart from microsatellites, Sano and Sano (1990) and Cordesse et al (1990) have used ribosomal DNA spacer length variants to investigate the genetic diversity in rice, while Fukuchi et al (1993) have demonstrated the utility of retrotransposons for fingerprinting of rice cultivars. RAPDs have also been employed in genetic diversity evaluation of *O.sativa* cultivars (Ko et al 1994; Virk et al 1995; Kim et al 1996) and upland and low land varieties (Yu and Nguyen 1994; Guimaraes, Borrero and Ospina-Rey, 1996). Recently, ISSR, RAPD and STMS markers have been shown to be useful in estimating genetic diversity within closely related Indian elite rice varieties (Davierwala *et al* 2000 (in press).

Genetic Diversity Analysis: Tapping the Existing Germplasm Collections for Crop Improvement Strategies

Diversity Analysis of Exotic Germplasm

Following domestication, the genetic variation in crop plants has continued to narrow due to continuous selection pressure for specific traits i.e. yield and / or disease resistance rendering them more vulnerable to disease and insect epidemics and jeopardizing the potential for sustained genetic improvement over the long term (Harlan 1987). The outbreak of southern corn leaf blight in 1970 is an excellent example which drastically reduced the corn yield in the US, and was attributed to extensive use of a single genetic male sterility factor which was unfortunately linked to the disease susceptibility (Ulstrop 1978). Thus it is extremely important to study the genetic composition of the germplasm of existing modern day cultivars in comparison with their ancestors and related species. This will not only give an opportunity of finding new and useful genes, as the accessions with most distinct DNA profiles are likely to contain greater number of novel alleles but will also provide information on their phylogenetic relationship. Such studies are now becoming more important as the foundation of crop based agriculture largely rests on the availability and knowledge of exotic germplasms. The latter is selected on the basis of two features like I] the exotic germplasm must possess a significant number of unique DNA polymorphisms (throughout the genome) relative to the modern day cultivars and II] each exotic genotype has to be genetically dissimilar (on

the basis of DNA profiling) (Brown and Kresovich 1996).

Since molecular markers have the advantage of sampling diversity directly at the genome level a database can be established for genetic diversity of germplasm collections. A wide spectrum of DNA markers, both specific, as well as arbitrary, have been used so far for DNA fingerprinting of various classes of germplasm (Winter and Kahl 1995, Callow 1997, Virk *et al* 1995, Chowdari *et al* 1998). Further studies with STMS markers may also provide an insight into the domestication in crop plants and provide criteria for enriching the genepool of crop plants (Powell *et al* 1996; Ramakrishna *et al* 1998; Udupa *et al* 1999). Besides the above examples, potential of ISSR markers has been exploited for diversity analysis of rice (Blair *et al* 1999).

Alien Introgression in Rice

The existence of genetic diversity enables crop plants to evolve and cope with environmental changes. In nature, the introgressions do occur spontaneously by chance, however, breeders are now attempting to enhance the rate of these by using modern biotechnological practices to overcome the crossability barriers between the wilds of different genomes and *O.sativa*. The wilds and land races usually have low yield capacity but their yield levels are highly stable. They may be regarded as natural composites of an array of resistance genes, and are well adapted so that no single race or biotype of plant pests can attack at the time of epidemic (Oka 1991). About 9,700 (12% of the total collection) species/ genotypes of rice in the currently existing collection in IRRI, have been found to contain useful genes for tolerance to several biotic and abiotic stresses (Oka 1991). Among these genes, a gene for resistance to grassy stunt virus of rice found in *O.nivara* (A genome) has been transferred to *O.sativa* by back cross breeding (Khush *et al* 1977). Similarly genes for several other traits have been transferred to *O.sativa* (A genome) like bacterial blight resistance from *O.longistaminata* (A genome), *O.officinalis* (C genome), *O.minuta* (BC genome), *O.latifolia* (CD genome), *O.australiensis* (E genome) *and O.brachyantha* (F genome), blast resistance from *O.minuta* (BC genome), cytoplasmic male sterility from *O.sativa f.spontanea, O.perennis and O.glumaepatula* all belonging to the A genome, increased elongation ability,

tolerance to acid and sulfate soils from *O.rufipogon* (A genome) etc (Brar and Khush 1997). Some of these introgressed genes have been mapped via linkage to molecular markers, which have been useful in tracing the introgression during backcross breeding (Brar and Khush 1997). Wild germplasm is an extremely precious material that holds a reserve of exotic traits that might be useful in future crop improvement plans and DNA markers will definitely continue to prove as important tools for assisting these introgression programs.

Phylogeny and Evolution

Most of the early theories of evolution are based on morphological and geographical variations between the organisms. However, the molecular biology tools hold a promise of providing a detailed information about the genetic structure of natural population, which has not been available in the past (Slatkin 1987). The DNA markers in particular are being used extensively for reconstructing phylogenies of various species and are speculated to provide path breaking information regarding the fine time scale on which closely related species have diverged and the genetic variations associated with species formation. The taxonomic classification is an essential step to determine whether the germplasm is a part of the primary, secondary or tertiary genepool of the system concerned. This is especially important in cases where morphological markers can prove to be inaccurate and misleading. A genuine example is provided by cultivars Azucena and PR 304 that have been classified as indicas using morphological characters, whereas they behave like japonicas upon crossing (Callow *et al* 1997). These genotyopes, are, however, revealed to be japonicas after RAPD analysis (Virk *et al* 1995). The phylogeny of rice has been reconstructed by several workers using different marker systems like RFLP (Wang *et al* 1992; Second and Wang 1992), RAPDs (Ishii *et al* 1996) and AFLPs (Aggarwal *et al* 1999).

Although DNA markers provide a valuable information regarding the evolution and phylogeny of various species, the trend is now shifting towards the use of ESTs (Expressed sequence tags) for such analysis (Mason-Gamer *et al* 1998; Deshpande *et al* 1998; Ge *et al* 1999). This is because in such studies one

actually looks at the evolution of functional genes. More details on phylogenetic organization of rice are given in Chapter 2.

Mapping and Tagging of Genes: Generating Tools for Marker Assisted Selection in Rice Breeding

Plant improvement, either by natural selection or through the efforts of breeders, has always relied upon creating, evaluating and selecting the right combination of alleles. Manipulation of a large number of genes is often required for the improvement of even the simplest of characteristics (Flavell 1995). With the use of molecular markers, it is easy to trace the valuable alleles in a segregating population and map them. These markers once mapped enable dissection of complex traits into component genetic units more precisely (Hayes 1993) thus providing the breeders with tools to manage them more efficiently in a breeding program.

The very first genome map in plants was reported in maize (Helentjaris *et al* 1986, Gardiner *et al* 1993). This was followed by development of framework maps in rice (McCouch *et al* 1988), *Arabidopsis* (Chang *et al* 1988, Nam *et al* 1989) etc using RFLP markers, which were further saturated by other DNA markers. Microsatellite markers, especially STMS have been extremely useful in this regard, as they follow clear Mendelian inheritance. The very first attempt to map microsatellites in plants was made by Zhao and Kochert (1993) in rice using $(GGC)_n$, followed by mapping of $(GA)_n$ and $(GT)_n$ by Tanksley *et al* (1995) and $(GA/AG)_n$, $(ATC)_{10}$ and $(ATT)_{14}$, by Panaud *et al* (1995). The existing map is being constantly updated using minisatellites, STMS and AFLPs (Cho *et al* 1996; Mackill *et al* 1996; Miyao *et al* 1996; Panaud *et al* 1996; Chen *et al* 1997; Maheshwaran *et al* 1997; Zhu *et al* 1999; Gustafson and Yano 2000; Temnykh *et al* 2000) so as to cover the entire genome. Besides all the above markers, recently, Cho et al (2000) have mapped 323 STMS markers, which are ESTs. Another important application of genetic linkage maps has been comparative mapping. In this strategy, the mapping information is used to predict the linkage relationships in between closely related or distant taxa. Apart from genome organization and evolution of the species under study, such information is also useful in speculating the crossability between them. In

between species comparisons of rice genomes, using RFLP markers revealed that gene order is highly conserved with the exception of some rearrangements, inversions and deletions (Jena *et al* 1994; Huang and Kochert 1994). However, genetic linkage map using microsatellites and STS markers for a cross between *O.glaberrima* x *O.sativa* shows comparable overall map length and good colinearity with intra-specific maps (Lereaux *et al* 2000) thereby implying that mapped markers in one system can be extrapolated to other species within a genus.

Mapped markers are now efficiently employed in tagging several individual traits that are extremely important for a breeding program which is designed for enhancing yield, disease resistance, stress tolerance and seed quality. A large number of monogenic and polygenic loci for various traits have been identified in rice, which are currently being exploited by the breeders and molecular biologists together, so as to make the dream of marker assisted selection come true. Several genes have been tagged in the last few years, the most prominent examples being resistance genes for blast like *Pi*-1*, Pi*-2 (t), *Pi*-5 (t), *Pi*-7 (t), *Piz*-5 *and Pita* which have been tagged using RFLPs (Yu *et al* 1991; Wang *et al* 1994; Hittalmani *et al* 1995, 2000), while Naqvi and Chattoo (1996) have identified 2 RAPDs tightly linked to *Pi*-10 and converted them into SCARs to facilitate marker assisted selection. Resistance genes have also been mapped and tagged for gall midge viz. *GM*-2 and *GM*-4 (Mohan *et al* 1994; Nair *et al* 1995; Mohan *et al* 1997b), brown plant hopper (Ishii *et al* 1994; Hirabayashi and Ogawa 1995) and stripe disease (Saito *et al* 2000). RFLP markers have also been useful in identifying a yield trait improving QTL from wild relative of cultivated rice *O.rufipogon* (Xiao *et al* 1996a). Some additional examples of gene tagging include the *S-a* locus for pollen sterility (Zhuang *et al* 1999), nuclear restoration gene, RF1 for fertility (Akagi *et al* 1996). Thermosensitive genic male sterile lines *tms*3 (t) and *tms*4 (t) (Wang *et al* 1995, 1998; Lang *et al* 1999; Dong *et al* 2000; Reddy *et al* 2000) and the photo period sensitive genic male sterility (PGMS) gene (Zhang *et al* 1994a). In addition, mapped markers along with YACs / BACs are now being employed to facilitate map based cloning of resistance genes. A classical example of map based cloning is that of the *Xa* 21 gene from *O.longistaminata* conferring resistance to

bacterial blight (Song *et al* 1995).

Marker Assisted Prediction of Hybrid Performance and Heterosis

Measures of Genetic Distance Based on Molecular Markers

In 1948, Melcot defined the co-ancestry coefficient 'f' as the probability that two homologous genes drawn at random, one from each of the two individuals, are identical by descent for the quantification of genetic similarity between two related individuals. The 'f' values have been used in some autogamous crops such as barley, oats, wheat and soybean to i] determine the genetic relationship between cultivars of a genepool adapted to a certain region ii] compute the relative genetic contributions of ancestral genotypes to released cultivars and iii] examine the level of genetic diversity of a given gene pool and monitor its changes over a time. The co-ancestry coefficient is an indirect measure of genetic similarity and its application is questionable due to unrealistic assumptions (absence of selection and genetic drift) underlying the calculation of 'f' values (Cox and Murphy 1990; Smith *et al* 1990) indicating need for more accurate assays.

Molecular markers have been extensively employed for genetic diversity estimations in studies targeted towards finding the relationship between genetic diversity and heterosis (Lee *et al* 1989; Melchinger *et al* 1990b; Smith *et al* 1990; Godshalk *et al* (1990); Dudley *et al* (1991); Boppenmaier *et al* (1993); Zhang *et al* 1994b, 1995, 1996). All molecular marker based assays generate a characteristic banding pattern for each individual which can be assigned a marker genotype and genetic distance (GD) can be calculated from the differences in allele frequencies at the marker loci. When the banding pattern is complex and the genotype cannot be determined directly, the GD is usually estimated based on the uncommon bands between the individuals being studied (Nei and Li 1979).

Crucial considerations for the application of molecular markers for grouping of genotypes are:

1] The mean genetic similarity among genotypes within germplasm groups and between germplasm groups.
2] The variation in genetic similarity of genotypes from the same as compared to different germplasm groups.
3] The establishment of simple and robust criteria that allow unambiguous assignment of genotypes (with unknown genetic background) to the respective germplasm groups.

Relationship of Genetic Distance with Hybrid Performance

Heterosis (hybrid vigor) is one of the most difficult and controversial genetic terms which was Introduced by Shull in 1914. It can be best defined as the superiority of F_1 hybrids in terms of performance, over some measure of performance of its parents. Although several scientific investigations have been undertaken, to understand this phenomenon, its genetic, biochemical and physiological bases remain largely unknown.

The successful development and utilization of the phenomenon of heterosis was observed in maize in about 1930 and this was a landmark in crop breeding (Lee 1995). It definitely provided impetus to plant breeders to explore commercial exploitation of hybrid vigor or heterosis in other crops. However, despite all the efforts made so far, there has not been much success in understanding the theory of functioning of this complex phenomenon. With the recent developments of recombinant DNA technology and the refinement of a number of molecular techniques, constant efforts have been made to advance our understanding of the genetic mechanism of heterosis.

Based on experimental analysis, Moll *et al* (1965) illustrated that midparent heterosis (heterosis) could be related to genetic divergence, a relationship that was supported by quantitative genetic theory (Falconer 1981). Distances computed from isozyme loci data were in some cases significantly correlated to heterosis, but, these correlations were genetically too low for the distances to be of practical predictive value (Peng *et al* 1988). This was probably due to insufficient genome coverage on account of low number of marker loci being used. With the availability of a wide range of molecular markers a larger genome area could now be covered. However, in several crops like maize

(Lee *et al* (1989); Godshalk *et al* 1990; Melchinger *et al* 1990a; b; Melchinger *et al* 1992; Smith *et al* 1990; Dudley *et al* 1991; Boppenmaier *et al* 1993), brassica (Diers *et al* 1996), rice (Zhang *et al* 1992,1994,1995,1996) and wheat (Martin *et al* 1995) it was observed that the predictions of hybrid performance highly depended on the germplasm involved in the study. It was also evident that the correlation between genetic distance and hybrid value increases with increasing relatedness in the germplasm under study. Melchinger *et al* (1992) have concluded that RFLP distances would not be useful for predicting the value of crosses between lines that belong to different heterotic groups in maize inbreds and this conclusion was supported by Boppenmaier *et al* (1992), while Arcade *et al* (1996) found significant correlation between GD and hybrid performance in larch. In conclusion, it appears that genetic similarity estimates based on either molecular markers or 'f' (co-ancestry coefficient) values have limited potential for correlation with hybrid performance in their existing form.

Heterosis: A Phenomenon Governed by QTL Associated with Yield and its Component Traits

Yield is an agronomically and economically important complex trait governed by QTLs associated with it and its component traits (grain number, grain weight, grains per panicle, spikelet fertility etc). All such loci cause continuous variation in the phenotype and are often accompanied by environmental effects. Determination of number, location and magnitude of effects through RFLP linkages of QTLs underlying yield component could elucidate the genetic basis of the trait leading to improved breeding and selection efficiency (Stuber *et al* 1992; Paterson *et al* 1995a). Stuber *et al* (1992) were able to detect QTL contributing to hybrid vigor in maize using molecular markers. The subsequent progress in DNA marker technology followed by the development of molecular linkage maps has made it possible to identify, map and measure the effects of genes underlying quantitative traits (Tanksley 1993, Dudley 1993).

Heterosis or hybrid vigor of any hybrid rice is estimated based on the final grain yield. Extensive attempts are being made to identify the basis of this phenomenon using mapped and tagged molecular markers. Xiao *et al* (1995) have shown that dominance is the major genetic basis of heterosis in rice by analyzing the QTLs using molecular markers. Recently Yu *et al* (1997) have

shown that epistasis is the genetic basis of heterosis in an elite rice hybrid as their studies show low correlation between marker heterozygosity and trait expression, indicating that the overall heterozygosity has made little contribution to heterosis. Thus the heterosis phenomenon is very complex and Identification of discrete mutations corresponding to QTLs offers a more rapid and efficient means to dissect quantitative traits (Paterson *et al* 1990; Dorweiler *et al* 1993). As more markers of this type are identified, predictions of hybrid performance might improve in comparison to their present status. This aspect has been covered in detail with more examples in chapter 3.

Future Prospects of Marker Assisted Selection (MAS) in Crop Breeding Programs

Plant breeding has been remarkably successful in the improvement of both qualitative and quantitative traits that affect the agronomic performance and consumer preferences (Martin 1998). Future improvement of crops will benefit from the isolation and characterization of functional genes that are either simply inherited or polygenic traits. Our inadequate knowledge about underlying biological processes, has impeded the breeding progress for quantitatively inherited traits such as environmental stress tolerance, yield, and even some simply inherited traits such as resistance to pathogens and insects. The isolation of economically important plant genes, has now been facilitated by the construction and application of genetic maps, transposon based gene tagging, protein- protein interaction cloning, and the development and analysis of large collections of cDNA sequences (Martin 1998).

However, even after identification of QTLs, it is extremely important to keep a track of these in the newer genotypes being generated and this may be difficult in cases where phenotypes are not very obvious. Identification of DNA markers linked to specific QTL offers the possibility of marker assisted selection for these traits. Genetic maps based on DNA markers have allowed the dissection of some quantitative traits into single component loci, which contribute substantial percentages of the phenotypic variation for a trait. This has been possible because QTLs can be localized very precisely near specific

DNA markers using nearly- isogenic lines (Grandillo *et al* 1996, Alpert and Tanksley 1996).

Molecular tags – a prerequisite for MAS (marker assisted selection) have been developed for many crop plants using different kinds of molecular markers. The success of marker assisted selection in plant breeding programs is dependent on some important factors like: i] Markers should co- segregate or be closely linked (1 cM or less is probably sufficient for MAS) with the desired trait. ii] An efficient means of screening large population for molecular markers should be available, like PCR based analysis (for the time being). iii] The screening technique should have reproducibility across laboratories, be economical and user friendly (Mohan *et al* 1997a).

It is evident that the development of DNA markers has revolutionized the construction of genetic maps in plants and their utilization in studies of plant evolution, systematics, and breeding. The utility of molecular markers will continue to increase as more and more genes are tagged. Technology for the utilization of molecular markers to their fullest efficiency is evolving rapidly and further advances in automation will surely change the present economics of MAS.

Genesis of Thesis and its Organization

In the present thesis, I have analyzed the rice genome using DNA markers and have studied their implications in phylogenetic and hybrid performance analysis. The Rockfeller Foundation which supports Rice Biotechnology Research world wide, has provided grant (RF 99001 # 715) to NCL for its project on " DNA markers in hybrid rice program" since 1994. The main emphasis of this program was focussed on DNA fingerprinting of rice and utility of molecular markers in Hybrid rice program.

As one of the aspects of my research, I have studied the genetic diversity and phylogenetic relationship in the genus *Oryza*, as revealed by inter simple sequence repeat polymorphism. This work has been carried out in collaboration with Dr. D. S. Brar (IRRI Philippines) and Dr. R. K. Aggarwal (CCMB Hyderabad, India).

Wild species of *Oryza* are an important reservoir of useful genes and can be exploited to broaden the existing narrow genetic base and enrich the existing varieties with desired agronomically important traits. To effectively utilize this wild species germplasm, it is essential to determine genetic diversity and identify a subset of alleles from the wild or exotic germplasm as this can help produce elite breeding lines. I have therefore, exploited the potential of Inter simple sequence repeat (ISSR) polymorphism as markers for genotype profiling and phylogeny of the members of genus *Oryza.*

In continuation with the above theme, I have further characterized the same germplasm with STMS markers containing $(CAC)_n$ repeat motifs, that I have developed by isolating expressing sequences from rice developing seed cDNA library. The genetic variability data of these markers may be useful in finding new alleles of useful genes which will definitely provide information about target gene sources in the wild germplasm that can be potentially used in future breeding programs for crop improvement and functional genomics.

In the second part of my work, I have studied the utility of DNA markers in the prediction of hybrid performance and heterosis in a three-line hybrid system. Heterosis is a complex phenomenon and molecular markers is one way scientist are using to understand more about the behavior of these quantitative traits, and use them for actual breeding programs. My work has been carried out in collaboration with Dr. B. L. Dhonukshe, Konkan Krishi Vidyapeeth, Dapoli, Maharashtra, India.

The CMS or three-line system has been found to be an effective way of developing hybrid varieties. Marker assisted selection of parents for such crosses are, however, uncommon in hybrid programs. Two cytoplasmic male sterile lines were crossed with fourteen restorer lines of rice widely grown in the western regions of Maharashtra, India to produce 28 F_1 hybrids. All these crosses were further evaluated for eight agronomically important traits, contributing to the yield potential, in a replicated field trial.

All the parental rice lines used in the above crosses were subjected to marker based variability studies using specific (RFLP and STMS markers linked to QTLs for yield and yield related traits) and random (ISSR) markers. The

genetic distance calculated on the basis of banding profiles was correlated with hybrid performance and heterosis. An attempt was also made to dissect each individual trait and correlate the genetic distance based on specific DNA markers linked to a particular trait and its individual field performance.

The entire thesis is organized as follows:

- **Chapter 1** contains review of literature in which I have discussed the different kinds of DNA marker and how they are used in research especially with reference to rice.
- **Chapter 2** is divided into two parts, Section I, consists of my findings on utility of ISSR matkers in fingerprinting genotypes of *Oryza* and revealing their phylogeny, while Section II, details the studies on variability of $(CAC)_n$ containing ESTs.
- **Chapter 3** describes my observations on the utility of DNA marker based GD in prediction of hybrid performance in a three line hybrid system.
- **Chapter 4** gives an overview of the thesis.

Chapter 2

The overall emphasis of the work embodied in this chapter is on studying the genetic diversity of the genus *oryza*, using microsatellite based DNA markers. In Section-I, I have studied the genetic diversity and phylogenetic relationship in the specific rice germplasm, using the Inter simple sequence repeat polymorphism. In section-II, the focus of my research is on identifying alleles of transcriptionally active, CAC microsatellite containing genes in the same germplasm.

Section I

Evaluation of Genetic Diversity and Phylogenetic Relationship in the Genus *Oryza* as Revealed by Inter Simple Sequence Repeat (ISSR) Polymorphism

The contents of this chapter have been published in Theoretical and Applied Genetics 2000; Volume 100, 1311-1320

Introduction

Rice is one of the agronomically and nutritionally important cereal crops and is the principal staple food in the developing countries. There has been an improvement in the yield of rice during green revolution and post green revolution period. However, a series of biotic and abiotic stresses continue to limit its productivity making it necessary to identify diverse sources of genes for tolerance to various stresses and broaden the rice gene pool. Wild species of *Oryza* are an important reservoir of useful genes and can be exploited to broaden the genetic base of rice and enrich the existing varieties with agronomically important traits.

The genus *Oryza* has two cultivated species, namely *O. sativa* (2n=24 AA) grown worldwide and *O.glaberrima* (2n=24 AA) cultivated in a limited area of western Africa. *Oryza* has 22 species (2n=24 or 48) representing AA, BB, CC, BBCC, CCDD, EE, FF, GG and HHJJ genomes (Vaughan 1989, Aggarwal *et al* 1997) grouped into four main species complexes, (1) Sativa, (2) Officinalis, (3) Ridleyi and (4) Meyeriama (Vaughan 1989). The species *O.brachyantha* (FF genome) does not fall into any of these complexes (Aggarwal *et al* 1999).

The rice germplasm comprises 76614 accessions of *O sativa,* 1255 accessions of *O.glaberrima* and 2779 accessions of wild species and their natural hybrids (Jackson 1997). To effectively utilize wild species germplasm, it is essential to determine genetic diversity and identify a subset of alleles from the wild or exotic germplasm and produce elite breeding lines that would contain only specific “wild QTL” or “gene”. Recently, *O.rufipogon* (2n=24 AA), one of the closest relatives of cultivated rice, despite its overall inferior appearance has been found to contain genes that can substantially increase the yield of rice (Xiao *et al* 1996a). A number of useful traits such as cytoplasmic male sterility, resistance to grassy stunt virus, bacterial blight, blast and brown plant hopper have been introgressed from wild species into cultivated rice (Brar and Khush 1997). Considering the large number of wild species germplasm available, it is difficult to choose which accessions should be used in order to maximize the chances of finding new and useful genes. Genetic diversity determined by various markers can be exploited in alien introgression from these accessions

into elite breeding lines. Morphological, isozymes and molecular markers such as RFLP, RAPD, AFLP and microsatellites have been used to determine genetic diversity and phylogenetic relationships in *Oryza* (Tateoka 1962, Morinaga 1964, Second 1982, Dally and Second 1990, Wang *et al* 1992 and Aggarwal *et al* 1999).

Our laboratory has been engaged in inter and intra species relationship analysis in genus *Oryza*. In previous reports, we have studied utility of various types of molecular markers like microsatellites and minisatellites to screen rice germplasm (Ramakrishna *et al* 1994, 1995, Gupta *et al* 1994b). Genetic variability at specific loci has also been analyzed in the wild germplasm using locus specific markers Rm 122 and *knotted*-1 homeobox (Ramakrishna *et al* 1998, Deshpande *et al* 1998). The ISSRs (inter simple sequence repeats) are the regions that lie within the microsatellite repeats and offer greater potential to determine intra genomic and inter genomic diversity compared to other arbitrary primers (Zietkicwicz *et al* 1994). Several properties of microsatellites like high variability among taxa, ubiquitous occurrence and high copy number in eukaryotic genomes (Weising *et al* 1998) make ISSRs extremely useful markers. They exhibit specificity of sequence tagged site markers, but need no sequence information for primer synthesis enjoying the advantage of random markers. In the present studies, I have exploited the potential of these markers in genotype profiling of genus *Oryza* representing all the nine genomes.

Materials and Method

Plant Materials

The material comprised 28 accessions of wild species belonging to 9 genomes, viz. *AA, BB, CC, EE, FF, GG, BBCC, CCDD* and *HHJJ*, 10 cultivars (indicas and japonicas), 1 landrace (indica) and three related genera (Table 1). Most of the samples were obtained from the Plant Breeding, Genetics and Biochemistry Division, IRRI, Philippines. The DNA was isolated following the method of Dellaporta *et al* (1983) and its quality and quantity were estimated both spectrophotometrically as well as visually by ethidium bromide staining on agarose gel.

PCR Amplification and Electrophoresis

A set of 100 anchored microsatellite primers was obtained from the University of British Columbia, Canada, while decamer arbitrary RAPD primers were from Operon Technologies, USA.

Amplification of DNA was performed using 10 mM Tris-HCl pH 9.0, 50 mM KCl, 1.5 mM $MgCl_2$, 0.25 mM dNTPs, 2% formamide, 0.2 μM primer, 0.5 mM spermidine, 0.8 U Taq DNA polymerase enzyme (Bangalore Genei, India) and 20 ng DNA per 25 μl reaction, in a Double Engine thermal cycler (M J Research, USA) for 45 cycles. After initial denaturation of 5 minutes at 94°C, each cycle comprised 1 min denaturation at 94°C, 45 seconds annealing at 49°C, 2-min extension at 72°C with a final extension for 5 minutes at 72°C at the end of 45 cycles. The annealing temperature was usually adjusted according to the Tm of the primer being used in the reaction.

Amplified products were mixed with bromophenol blue gel loading dye and were analyzed, by electrophoresis on varying percentage of agarose gel using 1x Tris Acetate EDTA buffer pH 8.0 at room temperature. In general, the quality of patterns generated differed from primer to primer. However, most of the patterns with extremely good polymorphism and useful information, were often accompanied with background smear, which was reduced using 2% formamide in the reaction. All the patterns generated were obtained at least three times to have a reproducible data.

Scoring of ISSR Data Points and Construction of Dendrogram

Each fragment that was amplified using ISSR primer, was treated as an unit character and scored as binary codes (1/0 = +/-). Only those bands that were reproducible and about 0.5mm apart were considered for scoring. Table 1 indicates the total number of data points scored for each sample being analyzed. The 1/0 matrix was used to calculate similarity / genetic distance using Dice coefficient. The resultant distance matrix was used to construct a phenogram based on unweighted pair-group method with arithmetic average (UPGMA) (Sokal and Michener 1958) using software packages NTSYS-PC 1.8 (F. J. Rohlf, State University of New York, Stony Brook, USA) and

TABLE 1: List of plant material used.

List of germplasm used	Accession	Genome	Origin	Total no. of bands
1. *O.glaberrima*	TOG 6216	*AA*	Africa Via Warda	55
2. *O.glaberrima*	TOG 6229	*AA*	Africa Via Warda	52
3. *O.longistaminata*	-	*AA*	Africa Via CRRI, India	53
4. *O.perennis*	104823	*AA*	Thailand	58
5. *O.rufipogon*	106424	*AA*	Vietnam	54
6. *O.rufipogon*	105908	*AA*	Thailand	59
7. *O.punctata* (2n)	105980	*BB*	Cameroon	57
8. *O.punctata* (4n)	100884	*BBCC*	India	66
9. *O.minuta*	101141	*BBCC*	Philippines	61
10.*O.minuta*	101125	*BBCC*	Philippines	57
11.*O.rhizomatis*	105432	*CC*	Sri Lanka	58
12.*O.officinalis*	100896	*CC*	Thailand	54
13.*O.officinalis*	101399	*CC*	Vietnam	50
14.*O.alta*	105143	*CCDD*	Guyana	72
15.*O.grandiglumis*	105669	*CCDD*	Brazil	58
16.*O.latifolia*	100167	*CCDD*	Costa Rica	70
17.*O.latifolia*	100965	*CCDD*	Costa Rica	64
18.*O.australiensis*	100882	*EE*	Australia via CRRI, India	66
19.*O.australiensis*	T-1434	*EE*	Australia via CRRI, India	62
20.*O.brachyantha*	101232	*FF*	Sierra Leone	59
21.*O.brachyantha*	B98-8025	*FF*	Africa	62
22.*O.granulata*	104986	*GG*	via CRRI India	52
23.*O.granulata*	106448	*GG*	Nepal	53
24.*O.longiglumis*	105148	*HHJJ*	Indonesia	46
25.*O.ridleyi*	100821	*HHJJ*	Thailand	49
26.*O.ridleyi*	101453	*HHJJ*	Malaysia	67
27.*O.nivara*	-	*AA*	India	61
28.*O.malampuzhaensis*	-	*BBCC*	India	51
O.sativa cultivars				
29.IR 36	Indica	*AA*	Cultivated variety	62
30.IR 64	Indica	*AA*	Cultivated variety	54
31.BG 90-2	Indica	*AA*	Cultivated variety	56
32.Basmati 370	Aromatic rice	*AA*	Cultivated variety	69
33.Azucena	Japonica	*AA*	Cultivated variety	62
34.Indrayani	Indica	*AA*	Cultivated variety	64
35.Ambemohr	Indica	*AA*	Cultivated variety	61
36.Intan	Indonesian	*AA*	Cultivated variety	64
37.Khadkya	Land race	*AA*	Maharashtra India	75
38.Waseas	Japonica	*AA*	Cultivated variety	67
39.Hakkoda	Japonica	*AA*	Cultivated variety	71
Related Genera				
40.*Porteresia coarctata*	Related genera	Unknown	Bangladesh	54
41.*Leersia*	Related genera	Unknown	Madagascar	48
42.*Rhynchoryza subulata*	Related genera	Unknown	Argentina	40

PHYLIP version 3.57 C' (J. Felsenstein, University of Washington, Seattle, USA), which was later used to infer phylogenetic relationship in the genus *Oryza*. Bootstrapping was carried out using the software program 'Win-boot' developed at IRRI to determine the confidence limits of UPGMA based dendrogram (Nelson and Yap 1996). Molecular weights of each of the potential specific bands were calculated using the software program "Seqid" (Rhoads and Roufa 1989).

Results

Identification of Useful ISSR Primers for Diversity Estimates in Genus *Oryza*

To investigate the utility of microsatellite directed DNA fingerprinting, complete set of 100 ISSR primers was attempted with a few representative *Oryza* genotypes. Finally 26 individual 3' anchored (ISSR) microsatellite primers, 2 (ISSR+ISSR) and 2 (ISSR + RAPD) primer combinations were used for amplification of all the genotypes. The microsatellite primers used included dinucleotide repeats anchored at the 3' end and some tri, tetra and penta nucleotide repeats. Of all the above primers used, 60% showed amplification in all the samples and only 40% produced polymorphic fingerprints. Mostly dark and prominent bands were scored, although bands of lower intensity but with high reproducibility were also included in analysis (Table 1).

The bands amplified using ISSR primers were in the range of 80bp to 4.5kb. Table 2 gives a description of the primers and fingerprint patterns using selected 30 primers. Of all the amplified profiles, 11 best and highly polymorphic patterns were selected for further analysis. A total of 302 bands were scored from the 11 patterns and 87 were observed to be present in either individual accession or species/genome (Table 3).

Fingerprint Patterns Revealing Putative Genome/ Species/ Cultivar Specific Markers

Dinucleotide repeats $(AG)_8$ and $(GA)_8$ with a number of anchors gave the best polymorphic and informative patterns. However, *O.granulata* accession 104986 originating from India did not show any amplification with $(GA)_8$ T but

showed amplification when (GA) repeat was anchored with C or A, indicating the possibility of (GA) repeat flanked by C or A and not by T in this accession. Other dinucleotides such as GT, TG, CA, and CT did not generate informative patterns. AT repeat, irrespective of its anchor, produced smears, which could be attributed to its abundance in the rice genome.

Figure 1 depicts fingerprints for various species of *Oryza* using dinucleotide repeat $(AG)_8$ YT. The primer revealed 4 potential individual accession specific bands one for *O.longistaminata* (Lane 3; 745 bp) and one each for *O.rhizomatis (*105432) (Lane 11; 970 bp), *O.alta* (105143) (Lane 14; 1000 bp), and *O.ridleyi* (101453) (Lane 26; 3600 bp). It also yielded 4 putative genome specific bands for the *F* genome (Lane 20, 21; 973 bp), *G* genome (Lane 22, 23; 324 bp), *BC* genome of *O.minuta* (Lane 9, 10, 2200 bp) and *HJ* genomes (Lane 24, 25, 26; 769 bp). One unique cultivated *O.sativa* specific band that occurred in indica, japonica, land race, and their wild progenitor *O.nivara* (Lane 27, 29 to 39; 821 bp) was an interesting feature of this profile. Apart from this, the primer also produced a band that was common to the *officinalis* complex, which included B, C, BC, and CD genomes (Lane 7 to 17; 440 bp).

The tri nucleotide repeats produced polymorphism and the patterns were often accompanied with bright background smears. The unanchored tetra nucleotide repeat $(GATA)_4$ detected extremely polymorphic and informative fingerprinting profile and yielded 6 putative accession /genome specific bands (Table 3). The primer $(GATA)_4$, when used in combination with RAPD primers although produced less number of bands, the amplified products were uniquely present in most of the samples being analyzed (figure not shown).

The penta nucleotide repeats (CTTCA) 3, generated informative pattern as depicted in Figure 2. This primer produced two potential genome specific markers for the *E* (Lane 18, 19; 2400 bp) and *G* genomes (Lane 22, 23; 370 bp) and two putative species-specific markers for related genera *Leersia* (Lane 41; 500 bp) and *R.subulata* (Lane 42; 2600 bp). It did not give any amplification in both the accessions of *O.officinalis* (*CC* genome) (Lane 12, 13) and one accession 100821 of *O.ridleyi* (*HHJJ*)(Lane 25) indicating that these repeats are probably distantly located in these genomes and spaced well beyond the capacity of amplification by Taq DNA polymerase.

Thus using various ISSR primers, 87 bands were identified, which can be converted into accession, genome and species specific markers (Table 3). The pattern that amplified maximum number of specific bands (19 bands) was the one amplified using $(GATA)_4$ in combination with a RAPD primer OPA 15, demonstrating the potential of primer combination system (microsatellite + RAPD). Although no two amplification patterns were exactly identical in general, the bands originating from same species were similar (Figure 1 and 2). The large number of unique bands obtained in the present analysis signifies the power of ISSR markers in fingerprinting and diversity analysis especially between closely related species. However, owing to the large number of accessions available for each of the species in the genus *Oryza*, it is important to screen the specificity of these bands over a large sample size before affirming their diagnostic capacity.

Inter and Intra Species Relationship

The ISSR derived fingerprint patterns of the 42 samples (Table 1) were used for cluster analysis and generation of dendrogram (Figure 3). This involved computation of a similarity coefficient matrix, which was used to calculate the genetic distance for all pairs of taxan units and then actual cluster analysis based on unweighted pair group method with arithmetic average (UPGMA). The confidence limits for using the Win-boot program (Nelson and Yap 1996) to perform UPGMA-based bootstrapping tested the groupings produced by this method. Two thousand bootstrap replications were carried out as suggested by Hedges (1992) and the values were obtained in terms of percentage of the number of times a group would be found in the bootstrap replications. This analysis in the form of a dendrogram (Figure 3) was used as the basis to infer phylogenetic relationships in the genus *Oryza.*

The bootstrap values seen at the nodes indicate that our data support phylogenetic inferences for groupings mostly at intra specific levels, i.e. samples carrying similar genomes and to a reasonable extent at inter specific levels, i.e. for part of *Latifolia* group species. However, the bootstrap values are rather low for rest of inter specific levels, indicating weaker inter specific relationships in the genus *Oryza* (as seen in Figure 3).

TABLE 2: Fingerprint patterns generated using the 30 ISSR primers

Primers used	**Repeat motif**	**Amplification pattern**	**Amplification of specific bands**
807	$(AG)_8T$	Good	Good
808	$(AG)_8C$	Good	Good
810	$(GA)_8T$	Good	Extremely good
811	$(GA)_8C$	Good	-
812	$(GA)_8A$	Smeared with bands	-
813	(CT)8T	Improper amplification	-
814	(CT)8A	Good	Average
815	(CT)8G	Smeared	
816	(CA)8T	Improper amplification	
817	(CA)8A	Not polymorphic	
818	(CA)8G	Not polymorphic	
819	(GT)8A	Not polymorphic	
821	(GT)8T	Smeared pattern	
829	(TG)8C	Smeared	
834	(AG)8YT	Good	Good
835	(AG)8YC	Good	Average
836	(AG)8YA	Not polymorphic	
841	(GA)8YC	Not very polymorphic	
847	(CA)8RC	Smeared	
848	(CA)8RG	Improper amplification	
850	(GT)8YC	Smeared with bands	
856	(AC)8YA	Smeared	
857	(AC)8YG	Smeared with bands	
868	(GAA)6	Smeared with bands	
872	(GATA)4	Good	Good
879	(CTTCA)3	Good	Average
834 + 847	(AG)8YT+(CA)8RC	Good but with smear	
834 + 807	(AG)8YT+(AG)8T	Good	Good
872 + opA15	(GATA)4+opA15	Good	Extremely good
836 + opA15	(AG)8YA + opA15	Not reproducible	

***TABLE 3*:** Putative genome/ species/ cultivar specific bands revealed through ISSR fingerprinting

Genome Complex	**Genome**	**Number of bands**
O.officinalis complex	B.C.BC	1 (834 440)
O.ridleyi complex	HHJJ	2 (807+834 520, 807 247)
Species / Genome	**Genome**	**Number of Bands**
O.nivara +all cultivars	AA	1 (834 821)
O.australiensis	EE	2 (872+opA15 194, 879 2400)
O.granulata	GG	4 (879 370, 834 324, 807 340, 872+opA15 1300)
O.minuta	BBCC	3 (835 270, 834 2200, 872 4400)
O.brachyantha	FF	7 (834 973, 810 1195,335, 807+834 1100, 872+opA15 550,444,808 630)
O.ridlevi	HHJJ	1 (834 769)
O.Latifolia	CCDD	3 (810 1636, 807 696, 872+opA15 639)
O.malampuzhaensis	BBCC	3 (872 1600, 980, 814 404)
O.glaberrima	AA	1 (872+opA15 380)
Accession	**Genome**	**Number of bands**
O.glaberrima TOG-6229	AA	1(810 1700)
O.longistaminata	AA	4(834 745, 872 1004, 807 363, 835 614)
O.perennis 104823	AA	1 (807+834 1300)
O.minuta 101141	BBCC	1 (872 + opA15 601)
O.rhizomatis 105432	CC	4 (872 + opA15 2200,1600,1400, 834 970)
O.officinalis 100896	CC	3 (872 2100, 807+834 2300, 872+ opA15 1360,)
O.officinalis 101399	CC	1 (872 + opA15 3000)
O.Latifolia 100167	CCDD	1 (810 750)
O.Lalitolia 100965	CCDD	1 (807+834 1000)
O.alta 105143	CCDD	7 (814 523,363, 834 1000, 810 1366,1021 , 807+834 2800 455)
O.grandiglumis 105669	CCDD	4 (810 880 785, 872+opA15 1460, 808 662)
O.australiensis 100882	EE	6 (808 328, 835 464, 807+834 3000, 807 98, 810 2090, 666)
O.australiensis T-1434	EE	1 (807 + opA15 979)
O.brachyantha B98-8025	FF	1 (835 2000)
O.Longiglumis 105148	HHJJ	1 (808 556)
O.redlevi 101453	HHJJ	2 (808 528, 834 3600)
Cultivars	**Genome**	**Number of bands**
IR-36	AA	2 (872 647, 807 + 834 436)
IR-64	AA	1 (872 1500)
Intan	AA	2 (807 1150,158)
Waseasahi	AA	1 (814 1600)
Hakkoda	AA	1 (807 305)
Related genera		
Porteresia coarctata	-	1(872+opA15 250)
Leersia	-	5 (814 1430,472 , 879 500, 810 511, 835 263,)
Rhynchoryza subulata	-	2 (879 2600, 872+ opA15 320)

* The Numbers in brackets indicate the UBC anchored microsatellite primer numbers while the subscripts indicate the molecular weights of the specific bands produced.

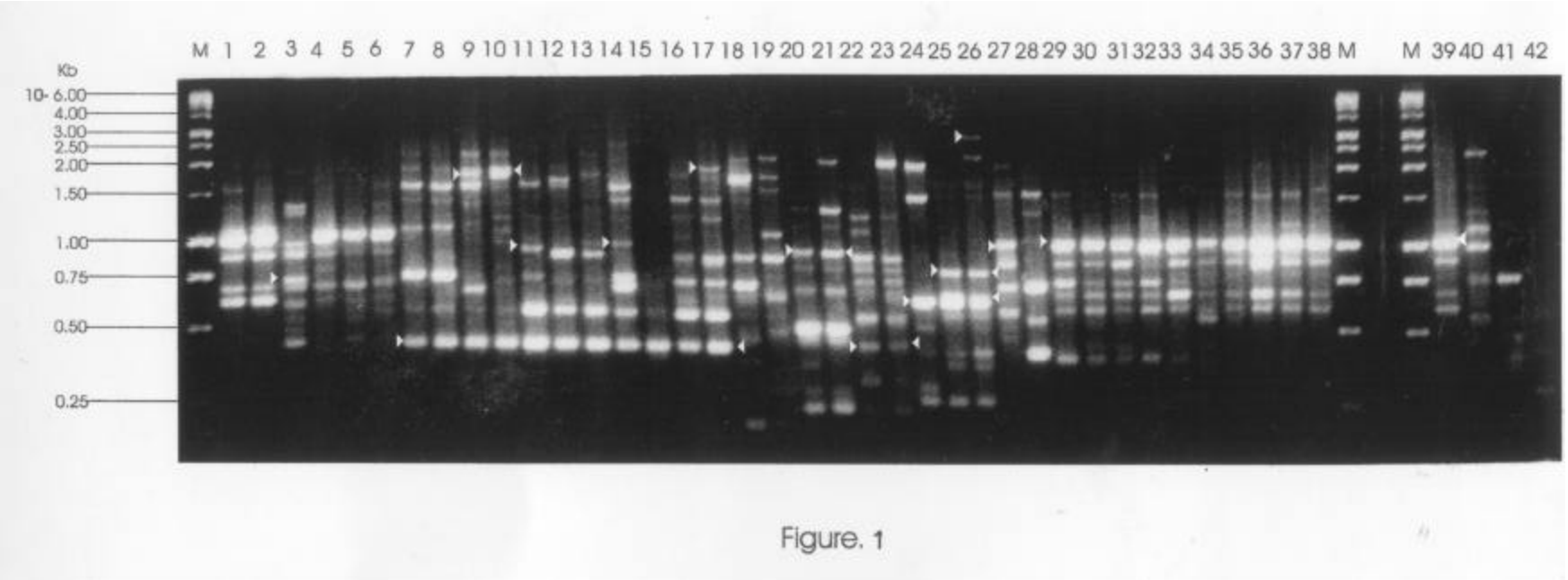

Figure. 1

Fgure 1: ISSR PCR fingerprints of 39 genotypes of *Oryza* and 3 related genera using 3' anchored $(AG)_8$ YT primer. The lane numbers are same as the sample numbers in Table 1. The lane marked M shows the 1-kb ladder being used as marker. The bands marked with white arrows are the specific bands amplified, which have been detailed in Table 3.

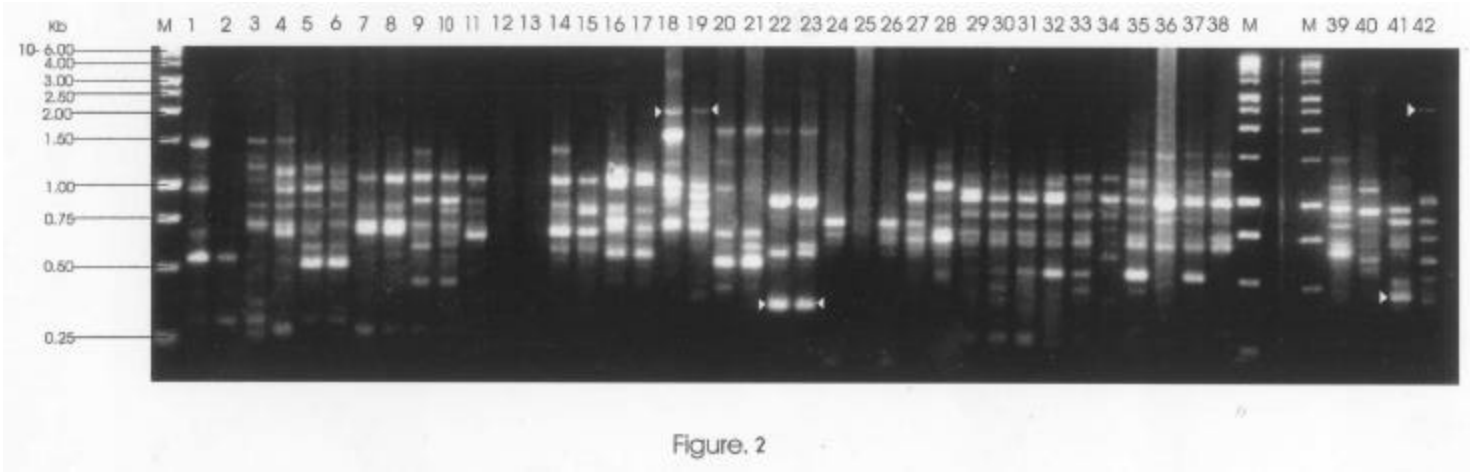

Figure. 2

Figure 2: ISSR PCR fingerprints of 39 genotypes of *Oryza* and 3 related genera using penta nucleotide $(CTTCA)_3$ primer. The lane numbers are same as the sample numbers in Table 1. The lane marked M shows the 1-kb ladder being used as marker. Note the lack of amplification in lane No. 12, 13 and 25. The bands marked with white arrows are the specific bands amplified, which have been detailed in Table 3.

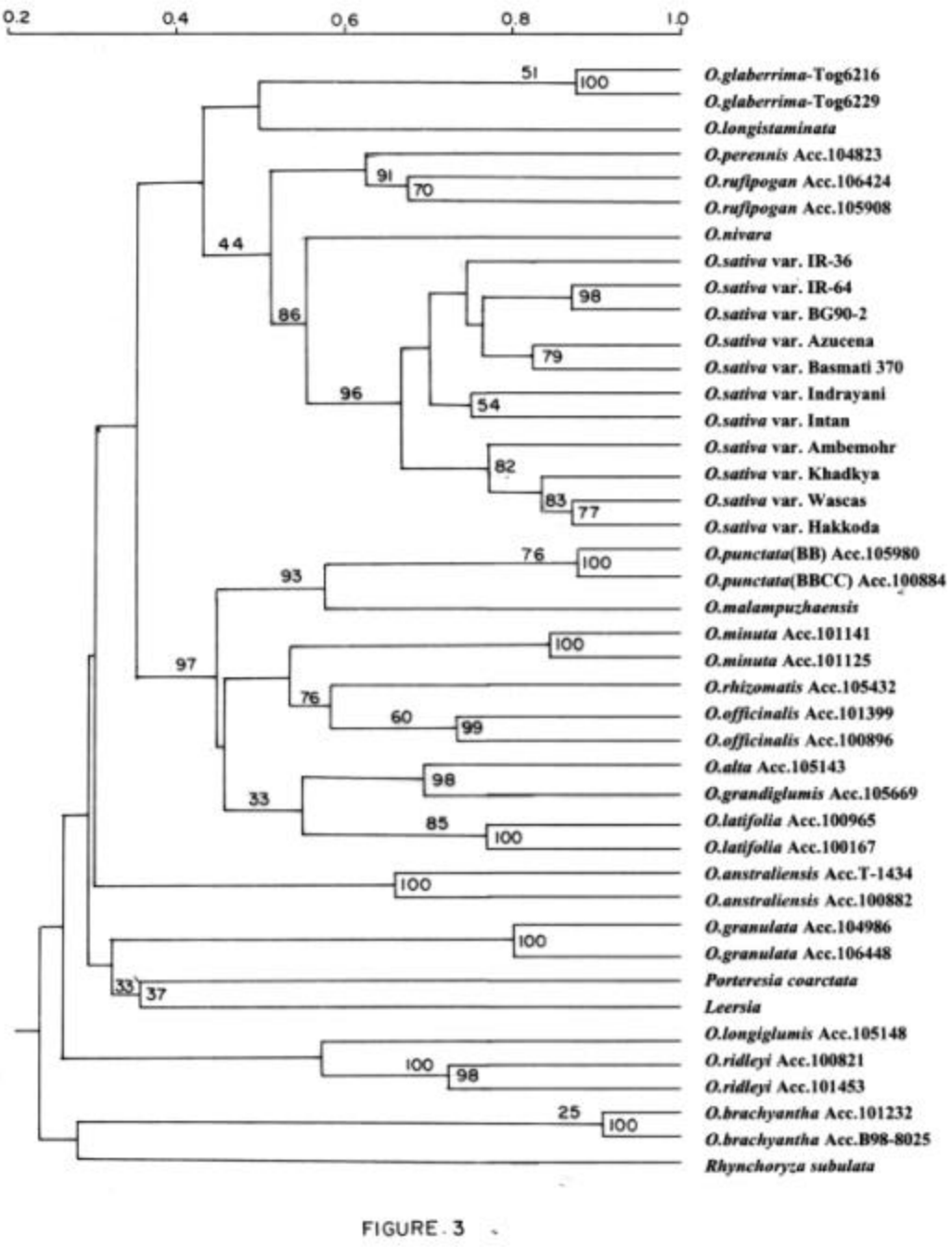

FIGURE 3

Figure 3: Dendrogram for the genus *Oryza* obtained using un-weighted pair group method with arithmetic average (UPGMA). The numbers at the forks indicate the confidence limits for the grouping of those species, which are to the right of that fork. This UPGMA tree indicates generic affiliations among members of genus *Oryza* based on ISSR polymorphism.

Accessions of the same species in most cases except *O.rufipogon* (70 %), show about 98-100% bootstrap values, while within cultivars of *O.sativa*, the bootstrap values are much lower except for BG90-2 and IR64 group where the value is as high as 98%.

Apart from these observations, three major groups are seen in the dendrogram, GroupI: *O.sativa complex* comprising *A* genome wild species and the cultivated rice varieties. GroupII: *O.officinalis* complex including BB, CC, BBCC, and CCDD genomes, and GroupIII: includes rest of the genomes i.e.: diploid genomes *EE, FF, GG* and the tetraploid *HHJJ* genome. Each of these four genomes maintains its independent status in the dendrogram. Besides the UPGMA analysis, parsimony analysis using Dollop (Phylip based program) was also attempted (Figure not shown). The groupings remained more or less the same and no major shift in the positions of the taxans was noted. Interestingly in this analysis, all the three related genera *(Leersia, Rhynchoryza and Porteresia*) grouped outside the *Oryza* complex without the need to re-root the tree.

Discussion

Genome Profiling Using ISSR Markers

In view of the large collection (80,000 accessions) of genetic resources of rice and importance of wild exotic germplasm in future breeding programs, it is essential to determine genetic diversity and phylogenetic relationships in *Oryza* species. With the use of DNA profiling, employing various molecular markers such as RFLP, DAF, RAPD and microsatellites, the genetic uniqueness of each accession can be determined and quantified (Brown and Kresivich 1996), however, PCR based markers are more suitable for large-scale analysis.

Among the PCR based markers, microsatellites are becoming more popular markers for genetic diversity and breeding research. Panaud *et al* (1996) have established a linkage map in rice using microsatellites. Distribution of microsatellites in the rice genome indicates that there are an estimated 1360 poly (GA) and 1230 poly (GT) microsatellites per rice genome, while the number of poly (AC) sites has been estimated to be 1000 with a genome size

of 0.45 * 10^9 bp (Panaud *et al* 1995, Wu and Tanksley 1993). The most frequent tetra nucleotide repeat reported to date in rice is poly (GATA) with about 270 poly (GATA) motifs in the entire rice genome (Panaud *et al* 1995). Our finding that the most polymorphic pictures have been obtained using poly (GA) or (AG) microsatellites irrespective of the anchors at the 3' end, and the most useful tetra nucleotide repeat is $(GATA)_4$ supports this microsatellite distribution. However, poly (GT) or (TG) and poly (CA) or (AC) microsatellites have not given good profiles in our analysis (Table 2) which may be due to distribution of these repeats in the rice genome beyond the range of amplification by Taq DNA polymerase.

The primers used in our analysis are anchored at their 3' end, to ensure that the annealing of the primer occurs only at the 3' end of the microsatellite motif thus obviating internal priming and smear formation. The anchor also allows only a subset of the targeted inter repeat regions to be amplified, thereby reducing the high number of PCR products expected from priming of dinucleotide inter-repeat regions to a set of about 10-50 easily resolvable bands. Pattern complexity can be tailored by applying different primer lengths and sequences (Zietkiewicz *et al* 1994).

Thus owing to their better specificity, and background information available on the distribution of repetitive DNA in the rice genome (Panaud *et al* 1995), 3' anchored microsatellites, were exploited in the present study. To the best of our knowledge, this is the first report where entire range of *Oryza* genomes and their related genera are compared using ISSR markers. The identification of inter and intra species polymorphism, obtained using ISSRs, opens a possibility of using these as multilocus DNA markers to generate species / genome specific profiles. The unique bands can be converted into STS markers. Such specific markers would be of great value to detect alien introgression and serve as DNA fingerprints for characterization of genetic resources of rice. Several duplicates could be eliminated and genetically diverse accessions from each species/ genome could be selected for use in breeding program.

Evolutionary Implications

Apart from establishing usefulness of ISSR markers for DNA profiling, the similarity/ diversity observed in the patterns of related species has enabled us to infer evolutionary relationships in the genus *Oryza*. The UPGMA based dendrogram obtained from the binary data deduced through fingerprints of the forty two samples being analyzed, adds a new dimension to the perspectives generated for the evolution of the genus *Oryza,* by other molecular markers like isozymes, RFLPs, AFLPs, chloroplast DNA probes as well as different morphological markers. The ISSR analysis revealed new understanding about the relationship at the inter species/ intra genomic levels with high bootstrap values, but is weak about the inferences related to higher level inter group relationships.

I] Evolution of the A Genome

The genus *Oryza* consists of nine genomes, amongst which only A genome has been domesticated over the years. The common rice *O.sativa*, and the African rice *O.glaberrima* are thought to be an example of parallel evolution in crop plants (Oka, 1991). *Oryza rufipogon* is considered to be the wild progenitor of Asian rice *O.sativa*, which shows a range of variation from perennial to annual types, while that of *O.glaberrima* is *O.barthii*, which is an annual grass, endemic to West Africa. In our analysis we find that *O.rufipogon* and *O.glaberrima* form two separate clusters. *O.longistaminata*, which is geographically known to be African rice, shows more similarity with *O.glaberrima* than with *O.rufipogon*. Similar to the earlier observations based on AFLP markers (Aggarwal *et al* 1999), among the A genomes, African forms are the most distant and they do not align with *O.rufipogon/ O.perennis*, but fall in *A* genome group of Asian species.

II] Wild Progenitors of Cultivars

O.rufipogon is distributed from Pakistan to China and Indonesia, and varies between perennial and annual types, which differ markedly in life history traits (Oka 1988). The variation between perennial and annual types is nearly continuous. It has been suggested that some intermediate perennial-annual populations, which regenerate by both ratoon and seed, are most likely to be

the immediate progenitors of cultivated rice. This is evident by their high evolutionary potential as shown by their rich genetic variability, a moderately high seed productivity and tolerance for habitat disturbance (Sano *et al* 1980, Oka 1988). The consensus tree obtained from our data using the ISSR markers supports this theory, as the cultivars form clear grouping with *O.nivara* the annual form of *O.rufipogon* and then cluster with the perennial forms including accessions of *O rufipogon* and *O.perennis*. The clustering of genotypes in the A genome (*O.sativa)* complex and the diversity between *O.nivara* and other accessions of *O.rufipogon* and *O.perennis* may support the proposition made by Shastry *et al* (1965) that *O.rufipogon* needs to be divided in to two categories.

III] Indica- Japonica Differentiation From Wild Progenitors

Differentiation of *O.sativa* in to two types i.e. indica and japonica is an interesting feature in the evolution of rice. There are two schools of thoughts regarding the evolution of these two plant types, monophyletic evolution and diphyletic evolution. A number of evidences are available in favor of each of this evolutionary path. The diphyletic path was proposed by Chinese workers (Oka 1988), and was supported by Wang *et al* (1992), while monophyletic evolution was supported by a number of evidences (Ting 1957,Oka and Morishima 1982, Wang 1984, Oka 1988, Glazmann 1987, Oka 1991). The indica japonica clusters, first join together and then group with *O.nivara* (annual form of *O.rufipogon*) before joining the cluster of *O.rufipogon/ O.perennis*, thus suggesting evolution of indica japonica rice to be monophyletic. Interestingly Figure 1 has yielded a single band that was shared by all the cultivars and their wild progenitor *O.nivara,* which seems to be missing in the other forms of the A genome, supporting the above observation to some extent.

IV] Evolution of Other Genomes

***O.officinalis* complex**

This is the second major complex in the genus *Oryza* and is the most diverged amongst all as it includes diploid *BB, CC* and *EE* genomes, and tetraploid *BBCC* and *CCDD* species (Vaughan *et al* 1989). However, the dendrogram (Figure 3) constructed with ISSRs, suggests that the *EE* genome may not be a

part of the Officinalis complex and has its own independent status in the genus *Oryza*. A putative *O.officinalis* complex specific band was observed in one of our fingerprint profiles (figure 1), obtained using primer $(AG)_8$ YT, which is present only in the four genome types i.e. *BB, CC, BBCC and CCDD*, and absent in the *EE* genome of *O australiensis*.

In the *officinalis* complex, the *BB, CC*, and *BBCC* genomes form one cluster while the *CCDD* genomes form a second cluster (*Latifolia* group). Within the first cluster, it is observed that *O.punctata* and *O.malampuzhensis* (both *BBCC* genomes) are closer to the *BB* genome than to the *CC* genome. The closer proximity of *O.punctata* (*BBCC*) to *O.malampuzhensis (BBCC)* is in agreement to previous observations made by Wang *et al* (1992).

The tetraploid species representing *CCDD* genome viz. *O.latifolia, O.alta*, and *O.grandiglumis*, form two groups within them as suggested by Wang *et al* (1992). *O.alta* and *O.grandiglumis* (belonging to Guyana and Brazil, respectively) are more closely related to each other than to *O.latifolia*, while the two accessions of *O.latifolia* (both from Costa Rica) cluster together. This observation is consistent with the one made by Wang *et al* (1992) using RFLPs and that made by Aggarwal *et al* (1997 and 1999) using whole genome hybridization and AFLP analysis, respectively. The diverse nature of the *CCDD* genome types may be due to the unique *DD* genome component, the origin of which is still a mystery. It has been widely proposed that the *D* genome originated from other diploid members of the genus *Oryza*. Our results however, are unable to provide any clear indication on this aspect, as the *CD* genome members are almost equidistant from them.

Other genomes

These include diploid genomes *O.granulata* (*GG* genome), *O.australiensis* (*EE* genome) and *O.brachyantha* (*FF* genome) and tetraploid genome *O.ridleyi* (*HHJJ* genome). Our data are not quite in agreement with the previous reports regarding the distribution of these genomes in the phenograms of genus *Oryza*. As suggested by Wang *et al* (1992), *O.brachyantha* (*FF* genome) stands as an independent species but has no affinity with A genome (*O.sativa*) as suggested by the former. However, our results indicate that *O.brachyantha* is the most

diverse genome and forms a separate group. This is in agreement with the AFLP analysis of *Oryza, which* also suggests *O brachyantha* to be the most divergent species (Aggarwal *et al* 1999). Similarly *EE* genome has often been grouped with the *officinalis* complex in a number of previous reports (Wang *et al* 1992) and has always been proposed to be a possible progenitor of the *D* genome, while, Tateoka (1962) proposes *EE* genome to be having its separate status. Our analysis supports the later view. The "G genome" maintained its individual status, which is in accordance with Aggarwal *et al* (1999) and Wang *et al* (1992). *HHJJ* genome is one of the newly designated genome like the G genome (Aggarwal *et al* 1997). The complex consists of *O.ridleyi* accessions and *O.longiglumis,* and shows a clear separate group formation in our cluster analysis thus supporting its separate biosystematic status in the genus *Oryza.*

The three related genera / species, *Porteresia coarctata*, *Rhynchoryza subulata*, and *Leersia* align with different *Oryza* groups in distance based UPGMA analysis, but with absolutely non-significant bootstrap values, however in parsimony analysis all the three species lie out side the *Oryza* complex.

In summary the ISSR markers provide a powerful tool for generation of potential fingerprinting diagnostic markers for genomes/ species/ cultivars. Also phylogenetic analysis on the basis of ISSR derived phenogram supports the polyphyletic evolution in the genus *Oryza*, wherein multiple lineages underwent independent divergence after separation from a common ancestor.

Section II

Detection of Variable Alleles of Transcriptionally Active Microsatellite $(CAC)_n$ Containing Genes in the Genus *Oryza*

The contents of this chapter have been communicated to 'Genome'

Introduction

In the wake of the development in genetic studies and an increasing amount of sequencing database becoming available, it has been amply emphasized that even single base changes in the genetic makeup of genes can change an individual's response to stimulus. Hence more and more STMS and SNP (single nucleotide polymorphism) markers especially in the coding sequences for major economic traits are being used and made available for analysis of expressed sequence tags in comparative genomics of related species. Such analysis would be important to identify superior breeding germplasm possessing variable alleles for the traits being studied. In this approach, parental material can be identified to use in conventional breeding programs without involving transgenic methods.

The STMS marker system is based on variation in the number of microsatellite DNA repeat units at a given locus. These markers can be generated using specific primers flanking the repeat motifs, which are designed from the sequence data of a specific locus in either coding or non-coding regions of the genome. Being PCR based, these markers are easier for studying the variation at a specific microsatellite locus. The level of heterozygosity is usually high since many alleles of a locus exist in a population due to variation in the number of tandem repeats (VNTR loci) in a given repeat motif (Weber and May 1989). The STMS have been extensively used for construction of saturated molecular maps in rice (Zhao and Kochert 1992, Tanksley *et al* 1995, Panaud *et al* 1996). These markers have proved to be extremely useful in establishment of linkage to agronomically important traits (Yu *et al* 1994), cultivar identification, pedigree analysis (Rongwen *et al* 1995, Sant *et al* 1999) and in understanding evolutionary processes and estimation of gene introgression in plant breeding programs (Hillel *et al* 1990).

In rice, (*Oryza sativa L.*), microsatellite markers are distributed throughout the genome and can detect a high level of allelic diversity in cultivated varieties and distantly related species (McCouch *et al* 1997). The rice genome is estimated to contain about 5700-10,000 microsatellite sequences consisting of different di-, tri-, and tetra- nucleotide repeat units (McCouch *et al* 1997). About 312 of

these estimated markers have been mapped (Temnykh *et al* 2000) and are being used in basic and applied genetic studies.

Although microsatellites are usually considered to be transcriptionally inactive, there are a few reports including our own observations, where poly-A^+ transcripts have been found to contain sequences homologous to the simple tandem repeats like $(CAC)_n$ in rice (Gupta *et al* 1994b) and a few other plants (Gortner *et al* 1996). The STMS markers from such coding sequences (ESTs) are not only useful for the saturation of linkage maps, but being functional genes they can provide dual information on the mapping position of the genes and the variability of alleles existing in the species being studied. However, only limited information is available in rice on characterization of such transcripts and on the use of STMS markers generated from expressed sequences for variability analysis and in molecular systematics of *Oryza*.

In order to study the use of STMS markers in rice germplasm analysis, I used $(CAC)_n$ microsatellite containing STMS markers with the wild and cultivated rice germplasm detailed in the previous section. The dendrograms generated from the data points scored based on these STMS alleles, are expected to elucidate the phylogeny of sets of homologous expressing sequences in *Oryza*. This information will contribute to a newly developing database on plant molecular systematics using low copy number genes, while the variable alleles themselves will reflect on the different variants of the genes being studied.

Materials and Methods

Screening of cDNA Library

An endosperm cDNA library in a λ Zap vector (a kind gift from Prof. T. W. Okita, Washington State University, USA) was screened, according to the instructions given in the Stratagene λ zap kit manual, using a $(CAC)_5$ end labeled oligonucleotide probe. The probe labeling was done as described by Ramakrishna *et al* (1994). Positive phage plaques were *invivo* excised into pBluescript phagemid SK (-) and transformed into *E. coli* strain XL1blue. The transformants were cultured and plasmids containing the inserts were isolated using the method described by Burnboim and Dolly (1979). To estimate the

size of the inserts, the isolated plasmids were digested using the restriction enzymes flanking the inserts. To further cross check that the inserts contain the $(CAC)_n$ microsatellite motifs, they were again hybridized with the oligonucleotide probe $(CAC)_5$.

Sequencing of the cDNA Clones and Primer Synthesis

In order to generate STMS markers, the putative clones were sequenced by Sanger's dideoxy method using sequenase version 2.0 sequencing kit (US Biochemical Corp). All these sequences have been submitted to the gene bank database and there accession numbers are given in Table 1. The oligonucleotide primers were synthesized from the nucleotide sequence flanking the microsatellite repeat motif in each clone (Table1).

Plant Material and DNA Extraction

The material used for locus specific amplifications comprised 28 accessions of wild species belonging to 9 genomes, viz. AA, BB, CC, EE, FF, GG, BBCC, CCDD and HHJJ, 10 cultivars (indicas and japonicas), 1 landrace (indica) and three related genera, (Table 1 of section I). In addition to these, four out-grouped members, viz. *Triticum durum* landrace Narsimgarh III (tetraploid wheat *AABB* genome), *Zea diploperensis* (wild relative of maize), *Hordeum marinum* (wild species of Barley), and *Avena vaviloviana* (wild species of oat) were also used. The seed material for respective samples of wild species of *Triticum durum* land race Narsimgarh III, *Zea diploperensis*, *Hordeum marinum*, and *Avena vaviloviana* were obtained from Agharkar Research Institute, Pune, India; Indian Agricultural Research Institute, New Delhi, India; and USDA, ARS, National Small Grains Collection, Aberdeen, USA, respectively. The DNA was isolated from the leaf tissue as described in Section I.

Polymerase Chain Reaction (PCR) Amplification and Analysis

The STMS primers were custom synthesized from Operon Technologies and the primer synthesis facility at the National Chemical Laboratory. Total genomic DNA from the samples was subjected to PCR amplification, using the designed primers (Table 1). Each PCR reaction was carried out in a volume of 10µl, using a Platinum *pfx* DNA polymerase PCR amplification kit for GC rich

templates from Gibco BRL. Each reaction contained 50ng of DNA, 0.15μM of each primer, 0.25mMdNTPs, 1X Pfx amplification buffer, 1mM $MgSO_4$, 1x enhancer solution, 1.25μCi α [^{32}p] dATP and 0.3 units of platinum *pfx* DNA polymerase. DNA amplifications were performed in a DNA Engine thermal cycler (M J Research, USA) for 35 cycles. After initial denaturation for 5 minutes at 94°C, the following 35 cycles comprised 15 seconds denaturation at 94°C, 30 seconds annealing at 58 to 62°C and 1 min extension at 68 °C. At the end of the cycles, a final extension for 5 minutes at 68°C was given, as per the instruction manual. The annealing temperature was usually adjusted according to the Tm of the primer being used in the reaction. Amplification products were separated on a 6% polyacrylamide denaturing gel containing 7 M urea and 0.5x TBE buffer (pH 8) and electrophoresed at 60 W constant power. One of the primer pairs showed amplifications in high molecular weight range (254bp to 1.2kb) and hence for this particular pair, the amplification was carried out in a 25μl reaction without adding radioactively labeled α [^{32}p] dATP. The amplified products of this reaction were loaded on a 2% agarose gel and the bands were resolved by electrophoresis in TAE buffer.

Scoring of STMS Amplification Data Points and Construction of Dendrogram

Each fragment that was amplified using the primer pairs was treated as a unit character and scored as binary codes (1/0 = +/-). Only those bands that could be amplified in replicate PCR's were considered for scoring. The 1/0 matrix was used to calculate the genetic distance using Dice coefficient, using the program Win-dist available along with the software Win-boot for bootstrapping, developed at IRRI (Nelson and Yap 1996). Later, Win-boot was used for both, construction of a phenogram based on unweighted pair-group method with arithmetic means (UPGMA) algorithm (Sokal and Michener 1958) and bootstrapping to determine the confidence limits of UPGMA based dendrogram.

***TABLE 1*:** Characterization of $(CAC)_n$ containing ESTs

CAC Clones, their gene bank accession no and homology	**Microsatellite motif being amplified**	**Primers used**	**Expected band size**
15 CAC (AF –263543) 94% homologous to cyp2	$G(GAG)_2(GTG)_4TTAGTC$ $(T)_5ATGAGTTC(GTG)TC$ $(GTG)TTG(GTG)AGATG$ AG	Pmb1: 5'AAGCCGGTCGTCATC GCC3' Pmb2: 5'CGAATCGATCCAAGA ACTC3'	211 bp
19 CAC (AF- 263542) 96% homologous to cyp2	$(TC)_2T(TC)_2ACCAA(CAC)G$ $A(CAC)GAACTCCTT(A)_4GA$ $CTAA(CAC)_4(CTC)_2CTTATT$ TA	Pmb3: 5'AGCACCGCCAAGCG GGTTGT3' Pmb4: 5'TCGCAGATCCAACTC CACCGAATC3'	238 bp
20 CAC (AF- 263544) 86% homologous to cyp2	$(CCAA)_2ACCAT(G)_3CGAT$ $(CT)_3(CAC)_3GA(CAC)GAAC$ $TC(T)_3(A)_4GACTAA(CAC)$ $(CTC)_2CTTA(T)_3A(G)_3ACT$ $(CGA)_2CTCC(GAC)_2$	Pmb5: 5'CAGAAGGATCGCAGA TCCAACTC3' Pmb6: 5'AGCTTTCATAAGATCT GTGC3'	209 bp
23 CAC (AF- 264730) Unknown gene	$(A)_5TAATTTAAGCAA(GA)_2A$ $GG(CA)_2(GATA)_2(CAC)CAA$ $CCTATCTATCTAC(CAC)_6C$ $TCGTGGGCTTGAGC$	Pmb7 5'ATGAGAGCTTGATGA ACTGAT3' Pmb8 5'ACTCATCAGCTAGGC ATATGT3'	235 bp
25 CAC (AF- 264731) Leucine zipper homeodomain containing gene	$CTGAT(CTG)_2GCCG(TCC)_2$ $(TCG)_3TCCGCGTGTTG$ $(GTG)_5$	Pmb9: 5'AGTTCCACATCCACC AGCAGCAGC3' Pmb10: 5'TGGACAGCAGCGGG GAGTCGTAC3'	254 bp
28 CAC (AF- 264732) Unknown gene	$CTCC(T)_2C(T)_7C(T)_2C(T)_5C$ $(T)_6C(T)_6CACC(T)_3C(T)_4$	Pmb11: 5'GTACTACCAACTCTC TACG3' Pmb12: 5'AATGATATGGATGGA AGCGG3'	180bp

Results

Identification, Characterization and Homology Search of ESTs Containing Microsatellite Sequence $(CAC)_n$

cDNA library from the developing seed (endosperm) of rice cultivar IR36 consisting of three genomes equivalent plaques was screened using $(CAC)_5$ oligonucleotide probe. Eighteen putative clones were identified out of which only 9 were found to be positive after tertiary screening (Figure 1). These nine clones were further characterized by dideoxy nucleotide sequencing where two clones did not show any continuous stretches of repeat motif and were thus deleted from further studies. Two other clones were found to be redundant and hence one of them was also eliminated. Among the remaining six clones, five clones were useful for primer synthesis flanking the $(CAC)_n$ microsatellite motif while one had the $(CAC)_n$ repeat immediately attached to the poly A^+ tail. However, this clone had additional stretches of $(T)_n$ mononucleotide repeats. Furthermore, none of the clones analyzed in the present study showed long stretches of $(CAC)_n$ repeat motifs (Table 1) like other di, tri or tetra nucleotide repeats reported earlier in rice (Panaud et al 1996). Most of them appeared to be dispersed short stretches and were present as part of a complex of small repeats. This observation is in line with the earlier reports (See Sertedaki and Lindsay 1996) where in $(CAC)_n$ motifs have been found to be organized in both minisatellites and microsatellites. Minisatellites are highly polymorphic but not present in the gene sequences while microsatellite sequences exhibit low heterozygosity and are often found in gene sequences. In either case, they do not form continuous large repeat units and are usually present as interrupted repeats or as part of some other complex repeats (Sertedaki and Lindsay 1996).

Sequences of six EST clones were subjected to homology search using the EMBL data- base. Three clones viz.15CAC, 19CAC and 20CAC revealed 96%, 86%, and 94% homology respectively to a previously reported *cyp 2* gene from rice leaf and root tissue (Buchholz *et al* 1994) suggesting them to be members belonging to the cyclophilin gene family. Within themselves, the clone 15 CAC is 89% Similar to 19CAC and 87% similar to 20CAC,

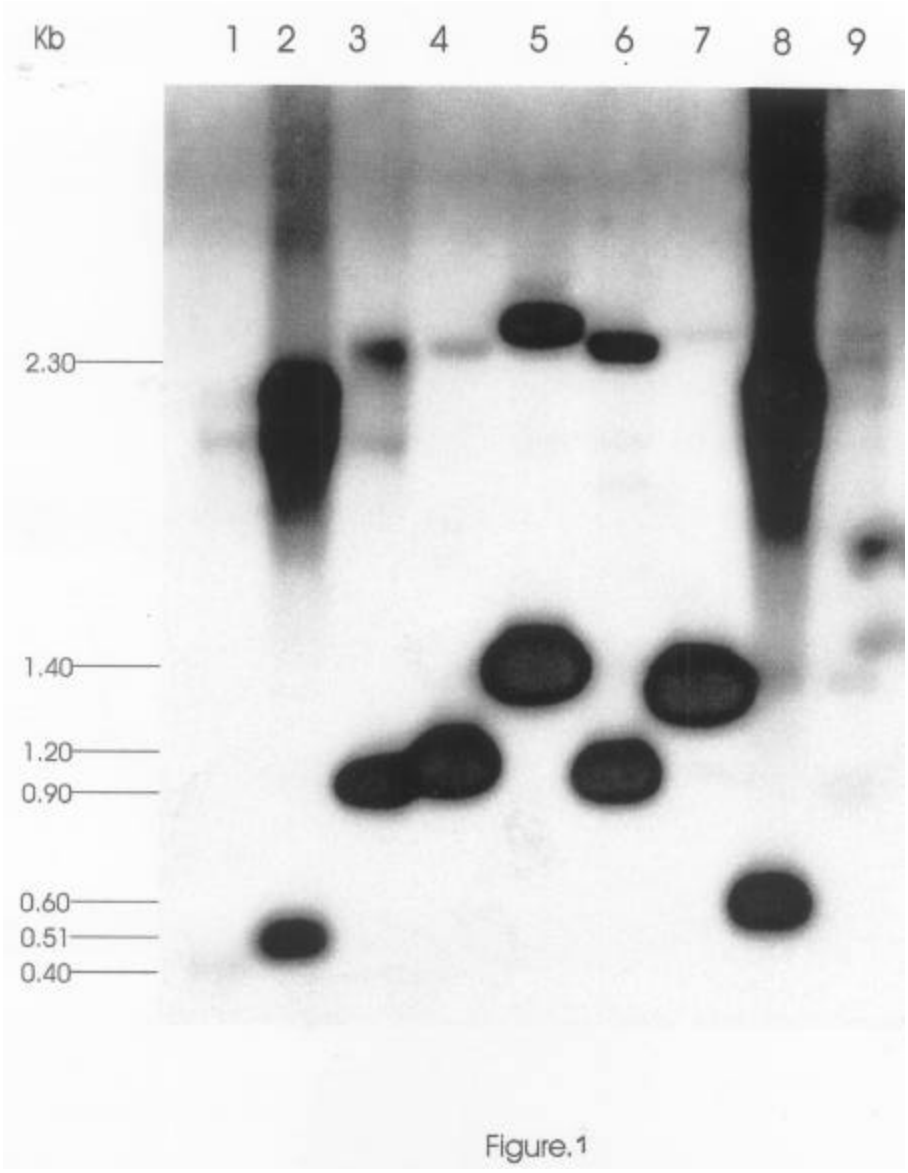

Figure.1

***Figure 1*:** Hybridization of oligonucleotide probe $(CAC)_5$ to *Kpn* I and *Xba* I double digest of the nine putative $(CAC)_n$ containing clones isolated from rice endosperm cDNA library to confirm that they were true positives. Lanes 1 to 9 are (1) 3 CAC, (2) 15 CAC, (3) 19 CAC, (4) 20 CAC, (5) 23 CAC, (6) 24 CAC, (7) 25 CAC,(8) 28 CAC and (9) 29 CAC. Band sizes depicted to the left, were determined using the molecular markers λ*Hind* III and ƒ*X Hae* III.

while 19 CAC is 81% similar to 20 CAC. The microsatellite motifs in all the three clones showed considerable similarities (Table 1) with minor changes in these regions which altered their reading frames. Clone 25 CAC showed the presence of a highly conserved leucine zipper homeodomain in its sequence. Apart from the homeodomain region, this cDNA sequence did not show homology with any of the genes reported in the database and the repeat motif did not lie in the homeodomain. The remaining two clones (23 and 28 CAC) did not show homology to any of the sequences available in the data bank.

Allelic Variation of the Genes in the Genus *Oryza*

Oligonucleotide primers flanking the repeat motifs for all the six cDNA, sequences were synthesized from the sequence data (Table 1). The amplification patterns with individual primer pair for respective clones are discussed bellow. One representative figure of amplification profile using primers PMB 3 and 4 (Figure 2) for clone 19 CAC has also been discussed.

Amplification pattern for clone 15CAC: Of the 5 alleles amplified using the primer pair PMB1 and 2 among the 46 DNA samples, a variant allele was found to amplify in *O.longistaminata* (214bp). Apart from this, an allele of 220bp was found to amplify in *O.longistaminata* and *O.nivara*, while a band of 208bp was seen in *O.alta (*105143*), O.grandiglumis (105669)* and *Zea diploperensis*. The *HHJJ* genome members *O.longiglumis* and both accessions of *O.ridleyi* amplified a unique allele of 223bp. Null alleles were observed in a number of samples, viz. both the accessions of *O.brachyantha, O.granulata,* cultivated rice varieties IR64, Azucena, Indrayani, Ambemohr, related genera *Porteresia Coarctata,* and *Leersia,* and an out group member *Triticum durum* Narsimgarh III. All the remaining samples showed presence of the 211bp expected band.

Amplification pattern for clone 19CAC: Figure 2, depicts the banding pattern amplified by the primers PMB 3 and 4. Usually the bands amplified as doublets i.e. a band along with a shadow band, except in a few cases. All the cultivated varieties (Lane 29-39) along with *O.nivara* (Lane 27), other A genome members (Lane 1 to 6), *O.punctata* (2n and 4n) (Lane 6-7) and *O.minuta* (101125) (Lane 10) were found to amplify the expected 238bp allele of this locus. A specific allele of 229bp was observed in *O.officinalis* (100896*)* (Lane

12). *W*hile null alleles were detected in *O.minuta* (101141) (Lane 9), *O.brachyantha* (101232) (Lane 20), *O.ridleyi* (101453) (Lane 26), and an out group member *T.durum* land race Narsimgarh III (Lane 44). Some of the variant alleles amplified in the members of the *Officinalis* complex (*BB, CC, BBCC, CCDD* and *EE* genomes) and *P.coarctata* (Lane 40) included products of size 232 and 235bp.

Amplification profile for 20CAC: Primers PMB 5 and 6 generated from the clone 20 CAC amplified six bands. Among these, one allele (212bp) was found to be common to *O.longistaminata*, *O.nivara* (the wild progenitor of cultivated rice) and all the cultivated varieties of rice. This allele was surprisingly absent in all other wild rice species and out-groups. Variable alleles were detected for this locus in the HHJJ genome members, viz. *O.ridleyi (*1008211 and 101453*)* (194bp). An allele of 197bp was amplified in *O.grandiglumis* (105669) (CCDD genome), *O.officinalis* (100896), *O.granulata* (104986), members of *HHJJ* genome (see Table1 Section I) and *Avena vaviloviana*, while no amplification was found in *O.malampuzhensis* and *O.granulata* (106448). The expected band of 209bp was seen to be amplifying only in the members of *AA* (including cultivated varieties)*, BB,* and *BBCC* genome while a major band of 203bp was observed in all wild accessions of all the genomes of *Oryza* being analyzed (Table1 Section I). Interestingly, this band was absent in the cultivated varieties of *AA* genome and their wild progenitor *O.nivara*. However, as mentioned above these cultivated varieties and *O.nivara* seem to share a band of 212bp, with *O.longistaminata* whereas *O.longistaminata* showed two alleles of 203 and 212 bp but did not show the presence of the 209bp expected band. Thus, an interesting distribution of this gene is observed in the AA genome suggesting a possible introgression of genes from *O.longistaminata* to cultivated rice during evolution. Still another variant of size 200bp was observed in the members of *CC, EE, FF,* and *CCDD* genomes.

The lack of amplification in certain samples, for all the primer pairs of cyclophilin gene family members, could be possibly attributed to mutations in the primer binding regions in their respective DNA's, as the cyclophilin genes have been reported to show lack of introns (Buchholz *et al* 1994).

Amplification pattern for 25CAC: The primers for the leucine zipper homeodomain containing gene PMB 9 and 10 showed a lot of variation (almost 28 alleles were observed), in its repetitive microsatellite motif. Among the germplasm used in our study, profound variation was observed with the band size ranging from 1.2 kb to 254 bp, while the expected band size was only 254 bp. Based on the current information, it is difficult to explain such tremendous variation in the band size and also the presence of multiple bands in all the samples. One possible explanation could be the presence of introns within the loci being amplified, which can only be confirmed by sequence comparison with the genomic clone (which is currently not available). Owing to the importance of bZIP genes in seed development, it would be interesting to analyze this clone in future studies. Surprisingly, these primers did not show amplification in a number of samples viz. *O.officinalis* (101399), *O.malampuzhensis,* both the accessions of *O.latifolia, O.australiensis, O.brachyantha,* cultivated variety *Azucena,* and the related genera *Leersia.* This can once again be attributed to either mutation in primer binding region of their DNA or presence of extremely large introns within the locus being amplified.

Amplification profiles for clone 23CAC and 28 CAC: For the clone 28 CAC, the $(CAC)_n$ motif was found to be present immediately before the poly A tail and hence flanking primers were difficult to synthesize. However, this clone contained stretch of mononucleotide $(T)_n$ repeat and therefore primers flanking this repeat motif were synthesized (Table 1). However, the primers for the clones 23 CAC and 28 CAC did not directly provide any useful information as their sequences do not show homology to any existing sequences in the gene bank. Primers PMB 9 and 10 for the clone 23 CAC produced 3 alleles in the size range of 150 to 235 bp, with the majority of samples showing the presence of the expected band of 235 base pairs and a null allele being observed only in *P.coarctata* and *R.subulata.* The primers PMB 11 and 12 produced monomorphic patterns and were not useful in this study.

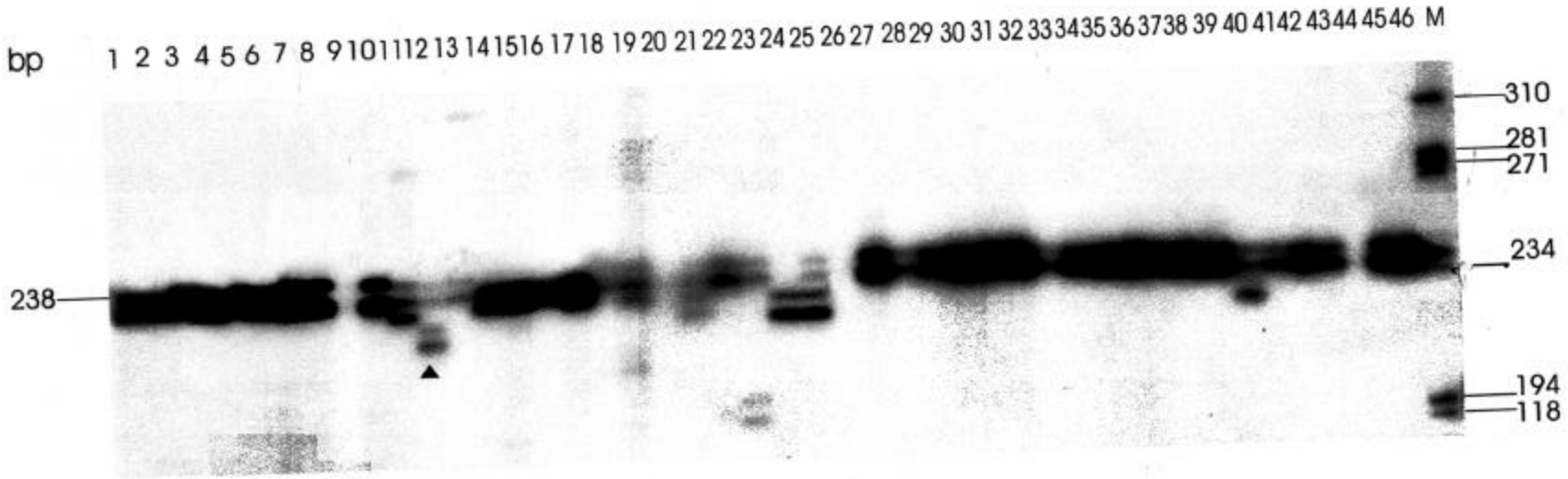

Figure. 2

Figure 2: Amplification pattern for clone 19CAC, using primers PMB 3 and 4. Lane M: marker *fX Hae*III (Biogenei), Lane 1 to 42: 28 accessions of wild species belonging to the 9 genomes of genus *Oryza*, 10 cultivars and 3 related genera as listed in Table 1 of Section I, Lane43: *Zea diploperensis* (wild relative of maize), Lane 44: *Triticum durum* land race Narsimgarh III (tetraploid wheat *AABB* genome), Lane 45: *Hordeum Marinum* (wild species of barley) and Lane 46: *Avena vaviloviana* (wild species of Oat).

Construction of Dendrograms for Phylogenetic Analysis

In order to study the phylogenetic relationship in the genus *Oryza,* dendrograms were constructed involving computation of genetic distance matrix (using the dice coefficient), based on the data generated by cyclophilin like genes and leucine zipper homeodomain containing gene, respectively (Figures 3 and 4). Only those samples in which amplification was obtained for all three primer pairs for cyclophilin were used for the construction of phylogeny (Figure 3) to avoid any inappropriate measurement, as lack of amplification does not necessarily imply a complete absence of the gene in a particular sample. The matrices were used to calculate the genetic distance and then the actual cluster analysis was done based on these values by unweighted pair group method with arithmetic average (UPGMA). The confidence limits for using the Win-boot program (Nelson and Yap 1996) to perform UPGMA-based bootstrapping tested the groupings produced by this method. In general, the consensus tree indicated that many of the species grouped together according to their genome types especially in the tree obtained using the cyclophilin gene family sequences (Figure 3). However, the bootstrap values were not adequately high indicating the uncertainty in the precise placement of the individual samples especially at the inter-genomic levels. A remarkable observation in Figure 4 was that all the cultivated varieties of rice grouped together first and then grouped with their wild progenitor *O.nivara*, as based on ISSR markers in section I. Another interesting observation was the grouping of *EE* genomes with the *CCDD* genome members namely *O.grandiglumis and O.alta*. This suggested that cyclophilin like gene sequences may be useful in future studies to clarify the ambiguous relationship among genome types and might provide additional support to the previous understanding that the *EE* genome is more closely related to the *DD* genome that gave rise to the *CCDD* genome.

The dendrogram (Figure 4) obtained using the genetic profiles with primers PMB 9 and 10 designed from clone 25 CAC appears to be confusing with regards to the true phylogeny of the genus *Oryza* and the low bootstrap values render this tree extremely unstable. This may be because of the amplification

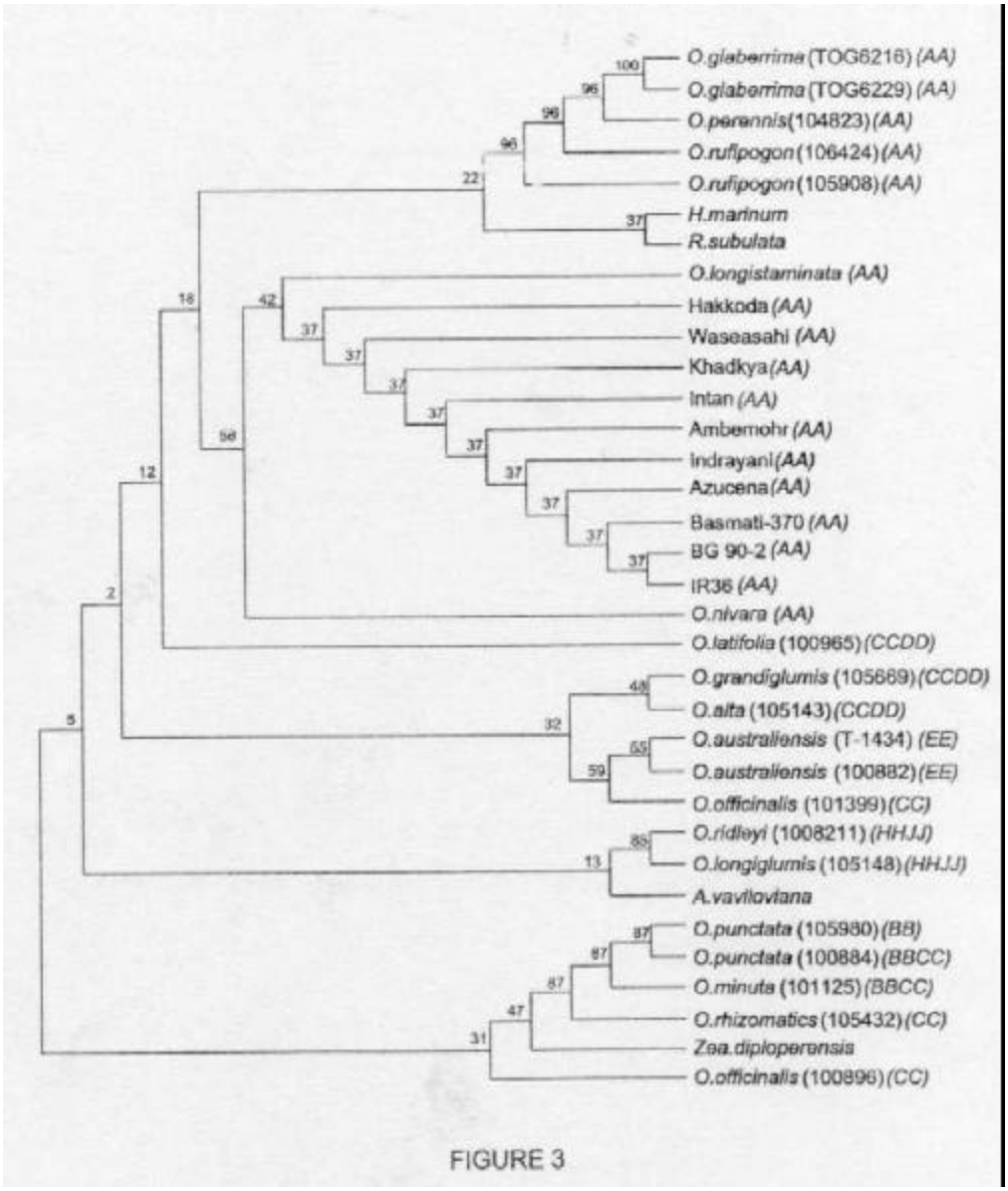

FIGURE 3

Figure 3: Consensus tree based on the locus specific amplification using primers flanking the microsatellite motif of cyclophilin like gene sequences viz. 15 CAC, 19CAC and 20CAC. The numbers at the forks show the percentage of times the groups consisting of the species to the right of that fork occurred.

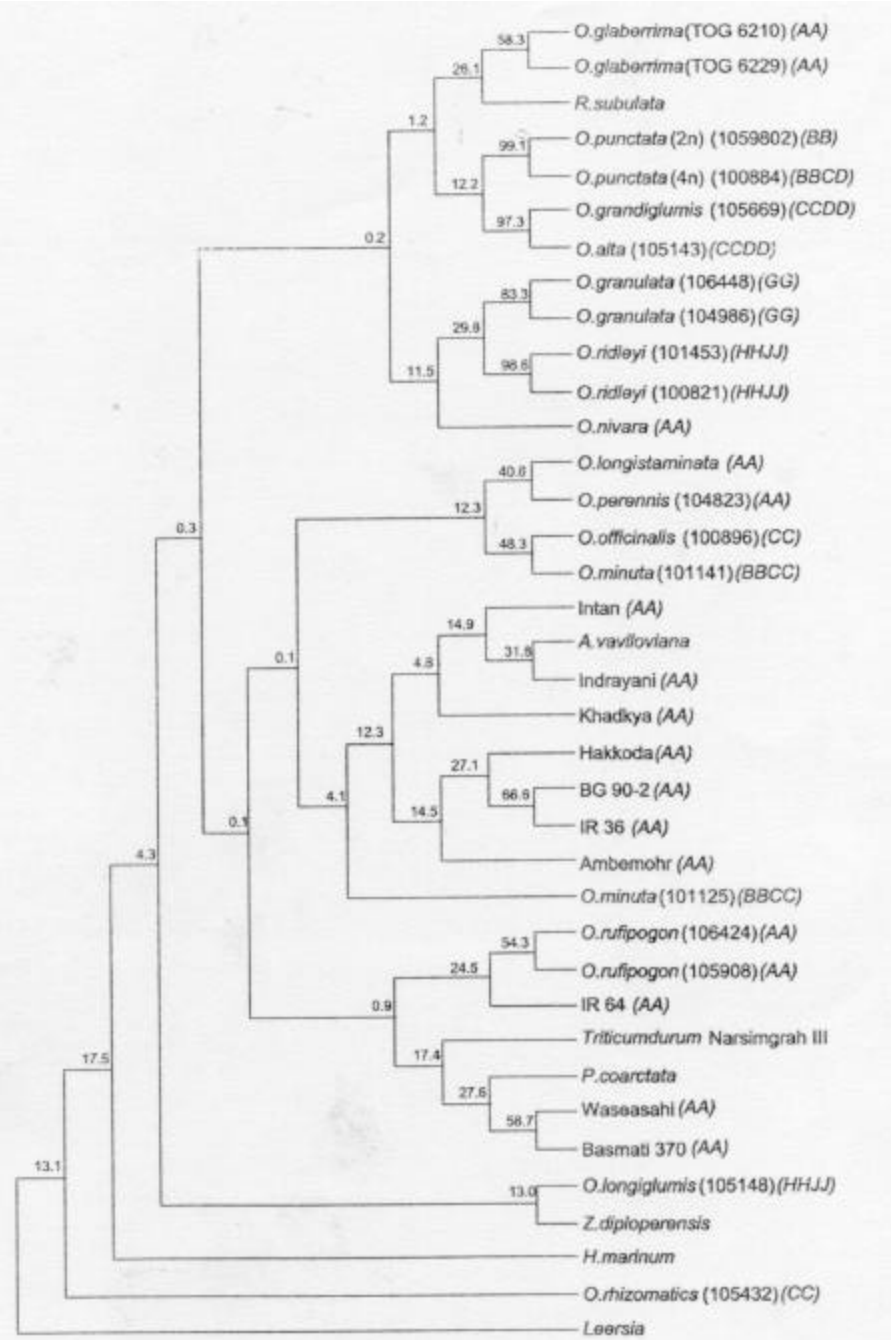

FIGURE 4

***Figure 4*:** Consensus tree based on the locus specific amplification using primers flanking the microsatellite motif $(CAC)_n$ of leucine Zipper homeodomain containing gene. The numbers at the forks show the percentage of times the groups consisting of the species to the right of that fork occurred.

of large number of bands due to possible intronic variation reducing the specificity of this marker.

Discussion

Functional Importance of the Genes Homologous to ESTs under Present Study

The homology search of the ESTs under our study showed three clones homologous to cyclophilin gene. Cyclophilins (*Cyp*) are reported to be highly conserved throughout evolution and appear to be ubiquitous in nature where variants of the gene have been found in mammals, insects, plants and fungi (Buchholz *et al* 1994). The proteins encoded by these genes have peptidyl-prolyl cis-trans isomerase activity that catalyses rotation of X- pro peptide bonds. It is proposed that cyclophilins are encoded by a small (6-10) member gene family, encoding abundant proteins with a common domain for CsA (Cyclosporin A) binding and enzyme activity, surrounded by unique domains involved in organelle and membrane targeting (see Marivet *et al* 1994, Saito *et al* 1999). *Cyp* genes are stress responsive and their expression increases during germination, suggesting their role in plant defense mechanism and plant development in maize and beans (Marivet *et al* 1992, 1994). Thus, in view of these properties it is evident that cyclophilins are not only involved in protein folding *in vivo* and are required for proper maturation, but they also play a role in trafficking and complex formation.

The remaining three clones when analyzed for homology showed two of them to be novel as they did not show any significant homology with existing gene sequences reported in the gene bank while the third clone showed the presence of leucine zipper homeodomain in its sequence. Leucine Zipper homeodomain containing genes are known to be the DNA binding genes which are crucial for development associated functions. Many plant basic leucine zipper (bZIP) proteins have been isolated to date, several of which have been shown to play an important role in seed specific gene expression. Recently, two members of plant bZIP family have been identified in rice viz. RITA1 (Izawa *et al* 1994) which has the function of transcriptional activator and REB

(Nakase *et al* 1997) which binds the α globulin promoter enhancing its expression thus emphasizing the role of leucine zipper containing genes in seed development.

Although the wild relatives of cultivated crops are low yielding than modern cultivars, it has been amply suggested that they may contain genes that make them more rugged to both biotic as well as abiotic stresses. Comparative genome analysis of candidate genes is now a popular approach to identify variant alleles of genes existing in nature. These variants can probably help to identify genetic predisposition of different individuals based on their genetic profiles. The variation in these genes can be studied by identifying SNP's in these genes, or by analyzing the microsatellite motifs contained in their sequences, as microsatellites are known to be hotspots of variation (Weber and May 1989). Hence, recognizing the importance of genes like cyclophilin and leucine zipper containing genes, we focussed our study on analyzing variability of these genes especially in their microsatellite motif region among the various members of the genus *Oryza*, a few related genera and some out-group members.

Low Copy Genes as Phylogenetic Tools

Phylogenetic data are critical for interpreting evolutionary relationships. It is usually assumed that phylogeny based on a single locus is a good approximation of the phylogeny of the organism (Kellog *et al* 1998). However, there are many examples in which this may not be true if introgressions or deletions have occurred. Such variations can be detected by comparison of a specific gene over a large number of related and unrelated taxa but these phylogenies may not bring out the true classification of the genus in question. This may be a possible reason why low copy number genes have not been extensively used in plant molecular systematics (see Kelloggs *et al* 1998).

Most of the molecular systematic studies have been based on either chloroplast genes or nuclear ribosomal DNA, while research on low-copy genes, especially in plant molecular systematics is in its infancy (Kellogg 1998). Recently, Ge *et al* (1999) have used the chloroplast gene (*mat*k) and alcohol dehydrogenase (*Adh*1 and *Adh*2) for phylogeny of rice genome, to determine

the origin of allotetraploid species. Some of the other functional genes found to be good candidates for phylogenetic reconstruction are phytochrome genes (Mathew and Sharrock 1996, Mathew *et al* 1995), small heat shock proteins (Waters *et al* 1996), nuclear gene *waxy* (Mason- Gamer and Kellogg 1996), and phosphogluco isomerase (Gottlieb and Ford 1996).

Being ubiquitous in nature, previously known cyclophilin genes have been studied for construction of their phylogeny and evolution across the prokaryotes and eukaryotes suggesting their polyphyletic origin (Chou *et al* 1997). However, they have not been studied so far to reflect the phylogeny of any genus in particular. The data points scored on the basis of alleles of the cyclophilin gene family members would possibly reflect upon the phylogeny of the genus *Oryza*.

In the present study, the dendrogram based on cyclophilin like gene family members showed a few interesting features; however, there were a few anomalies in this dendrogram. The genomes did not fall in precise inter-genomic order according to the classical phylogenetic classification, as has been suggested previously by several workers, using different marker systems (as discussed in section I of this chapter). All these studies showed a few variations with the change in the marker system, however the basic body of the tree remained more or less the same (Tateoka 1962, Morinaga 1964, Second 1982, Dally and Second 1990, Wang et al 1992, and Aggarwal et al 1999).

Thus the dendrogram generated based on cyclophilin like sequences (Figure 3) represented the phylogenetic distribution of these genes and not the genus as a whole. This is in line with the observation made by Kellogg (1998), that discrepancies are found among different gene histories, within closely related species or genera. Similar observations have also been made in case of wheat (Kellogg et al 1996), thus supporting the hypothesis that any single gene tree may be true with reference to that particular gene but may be misleading about the evolution of that organism.

In view of the large collection (100,000 accessions) of rice germplasm available, and the importance of wild exotic germplasm in future breeding programs, it is essential to determine genetic diversity and phylogenetic

relationships in *Oryza* species. Such an assessment of the relative's identity using gene sequences would provide a plethora of information, since these genome sequences will not only find the relationship within different members but will also provide information of a set of alleles that might be different and important for introgression programs. Although the phylogenetic relationships of the genus *Oryza* were not appropriately revealed in our studies, *O.longistaminata*, *O.officinalis (*100896*), O.ridleyi (*both the accessions*), O.longiglumis* (105148*), O.grandiglumis (*105669*), Zea diploperensis* and *Avena vaviloviana*, were found to contain variants of alleles of cyclophilin homologous genes. These results also indicated that all the cultivated rice samples showed similar amplification patterns with each other. It would be interesting to use the markers generated in these study and trace variant alleles in genome introgression programs, and to precisely determine the functional attributes of such markers in enhancing tolerance to biotic and abiotic stresses.

In summary, the genetic variability data generated in my study may not be of direct use for construction of phylogenies of the genus *Oryza,* in its present form. More direct sequence data needs to be generated from all the accessions to understand how the genes being studied have evolved during the evolution of rice. The preliminary phylogenetic analysis of cyclophilin genes, however, suggests the potential of these low copy nuclear gene sequences for future studies on evolution of allopolyploids. The STMS markers generated from the members of the cyclophilin gene family, would be useful in hunting for new genes playing important roles in the development processes and stress tolerance in rice and can be valuable target genes for functional genomics and introgression breeding.

Chapter 3

The emphasis of the work embodied in this chapter is on studying the utility of DNA markers in selecting parental lines for hybrid programs in a three- line hybrid system of rice

PREDICTION OF HYBRID PERFORMANCE AND HETEROSIS FOR A THREE LINE HYBRID SYSTEM IN RICE USING DNA MARKERS

The contents of this chapter have been communicated as a full paper to Biochemical Genetics.

Introduction

Many of the existing rice (*Oryza sativa L.*) hybrid cultivars are based on cytoplasmic male sterility and good combining restorers. The CMS or the three- line system has been found to be an effective way of developing hybrid varieties and will probably continue to play an important role in hybrid development for the next decade. Successful use of CMS lines in breeding hybrids depends on various factors such as the stability and adaptability of the CMS lines across different environments, the relative ease of fertility restoration, genetic diversity between the CMS and restorer parents, their out crossing potential, and Its combining ability (Virmani 1994).

In any hybrid program, a large number of crosses are made and only a few good hybrids are finally selected. This process is extremely labor intensive, time consuming and tedious, and hence alternative methods are explored for analyzing the potential material. The level of genetic diversity between the two parents being used for crossing has been proposed as a possible predictor of F_1 performance (Zhang *et al* 1994b). D^2 statistics by Mahalanobis (1936) and combining ability analysis by Griffing (1956) have been extensively used in predicting the hybrid performance on the basis of morphological traits. However, these methods require extensive field tests and crossing and hence there is a need to determine genetic distance at the molecular level and predict the hybrid performance (Melchinger *et al* 1990b). The results obtained using isozyme analysis by Li *et al* (1982) and Deng and Wang (1984) have revealed that distances computed from isozyme analysis are some times significantly correlated to heterosis. However, these correlations are generally too low for the distances to be of practical predictive value (Peng *et al* 1988; Peng *et al* 1991).

With the advent of molecular marker technology, DNA markers are now available and can be potentially used for the prediction of heterosis prior to actual fieldwork. The development of molecular linkage maps has made it possible to detect and analyze the loci underlying heterosis (Xiao *et al* 1995). It is envisaged that the use of such markers may have potential to estimate hybrid performance and heterosis. An essential assumption underlying this is

that a strong linear correlation exists between heterozygosity and heterosis (Zhang *et al* 1996). A number of efforts have been made in various crop plants to investigate the relationship between DNA marker based genotype variation of the parents in a hybrid program and hybrid vigor, with varying results. For example, in *Zea mays L.*, strong correlation between genetic distance and heterosis was observed by Lee *et al* (1989) and Smith *et al* (1990), while Godshalk *et al* (1990) and Dudley *et al* (1991) observed low correlation between marker distance and yield. Studies by Melchinger *et al* (1990a) and Boppenmaier *et al* (1993) showed that the correlation depended on the origin of the lines. Smith *et al* (1990) observed that the relationship between parental genetic distance and F_1 performance became significant with an increase in the sample size as well as the number of markers. Stuber *et al* (1992) found a significant relationship between parental heterozygosity and hybrid yield when the number of parental lines was increased. Studies in *Triticum aestivum L.* (Barbosa- Neto *et al* 1996, Martin *et al* 1995), *Glycine max (L.) Merr.* (Cerna *et al* 1997), and *Brassica napus L.* (Diers *et al* 1996), indicated that marker based genetic distance was not of any predictive value for hybrid performance. On the contrary, Arcade *et al* (1996) found significant correlations between genetic distance and hybrid performance in Larch.

In our laboratory attempts have been made to estimate correlations between the genetic distance based on random markers and heterosis in crops like *Pennisetum glaucum* and *Cicer arietinum*, where DNA markers have failed to be of any practical utility (Chowdari *et al* 1998, Sant *et al* 1999). However, Zhang *et al* (1996) have found that the correlations calculated using specific heterozygosity based on the small number of positive markers (markers linked to QTLs responsible for heterotic yield response) are more significant than those obtained using markers for general heterozygosity. While Bernardo (1992) carried out computer simulation to investigate complex but more realistic genetic model involving incomplete coverage of QTL by molecular markers. He suggested that effective prediction of hybrid performance based on marker heretozygosity is possible only if (i) dominance effects are strong, (ii) allele frequencies at individual loci in parental inbreds are negatively correlated, (iii) trait heritability is high, (iv) average parental allele frequencies vary only within

a narrow range, (v) at least 30-50% of the QTL are linked to molecular markers and (vi) not more than 20-30% of the molecular markers are randomly dispersed or unlinked to QTL. In case of rice, previous reports regarding marker- assisted selection for prediction of hybrid performance and heterosis have produced contrasting results (Saghai Maroof *et al* 1997, Yu *et al* 1997, Xiao *et al* 1996b, Zhang *et al* 1996, Zhang *et al* 1995, and Zhang *et al* 1994b). The utility of DNA markers in assigning inbreds into heterotic groups and quantifying genetic similarities between related lines is more or less proven. However, a general consensus view cannot be derived from all these reports regarding the utility of molecular markers in prediction of complicated phenomena, such as heterosis and more and more studies need to be carried out before their utility can be established.

In the present study, I have investigated the genetic diversity between the wild abortive (WA) type cytoplasmic male sterile lines (IR 58025A and IR 62829A) and their compatible 14 restorer lines, using both random (ISSR) as well as specific (RFLP and STMS) markers linked to 8 different yield related traits. Further, an attempt has been made to correlate the genetic distances of the genotypes with hybrid performance and heterosis. I have also dissected each individual trait and correlated the genetic distance based on specific DNA markers linked to a particular trait and its individual field performance.

Materials and Methods

Plant Material:

Two CMS lines (IR 58025A and IR 62829A) and 14 restorer lines (Rtn 1, Rtn 2, Rtn 24, Rtn 68, Rtn 73, Rtn 711, Kjt 1, Kjt 3, Kjt 14-7, Kjt 184, Kjt 487, Pnl 1, Pnl 2 and Plg 1) of rice were used in the present study. The restorers were selected so as to cover a range of characters required in the various terrenes of the Konkan region like very early/ early/ mid late/ late and very late flowering, bold grain/ long slender grain/ short bold grain/ very fine grain/ suitable for kharland/ resistant to BLB and gall midge etc. These restorer lines were well adapted to Konkan region and gave the higher yields specifically in this agroclimatic zone as compared to rest of the rice zones in India.

Each of the two CMS lines was crossed with 14 restorer lines to produce 28 F_1 hybrids. Crosses were made at the Konkan Agricultural University at Dapoli, Maharashtra. The hybrids along with their parents were grown in a randomized block design with three replications. During transplantation a space of 20 cm and 15 cm was left between the rows and in between the plants in each row respectively. The data was recorded using 15 randomly selected plants excluding side plants per replication per hybrid. These plants were tagged for recording detailed observations for 8 traits viz. plant height, panicle length, grains per panicle, unfilled grains per panicle, spikelet fertility, days to maturity, 100 grain weight and grain yield. Heterosis (mid-parent heterosis) and heterobeltiosis (better parent heterosis) for each cross were computed as deviations from mid parent and better parent values, respectively.

Leaf tissue of the parental lines grown in the fields at Konkan Agricultural University was harvested and DNA was extracted from young leaves using the hexa- decyl- trimethyl ammonium bromide (CTAB) method described by Rogers and Bendich (1988).

DNA Markers Analysis

Linked Markers for Specific Heterozygosity Analysis

A] RFLP (Restriction Fragment Length Polymorphism) Analysis

In these assays, DNA samples of the two CMS and 14 restorer lines used in the 28 crosses, were digested singly with six restriction enzymes (*Dra I, Bam HI, Hind III, Xba I, Eco R I* and *Eco R V*). These restriction enzyme digestion blots were further probed with 20 RFLP probes that are linked to 8 characters that directly or indirectly contribute to the yield of the rice crop (Table 1). Some of the probes that were linked to a common trait but positioned on different chromosomes were pooled during hybridization, to save on number of blots required and also time. This was possible as the aim was only to correlate the genetic distance between the parental lines based on fingerprinting patterns with the field performance of the F1 hybrids, and no F2 analysis was necessary. The Southern hybridization of the probes was carried out using the fluorescein gene images labeling and detection system from Amersham Life Science. The procedure for hybridization and stringency washes was

according to the instructions given by the manufacturer with the slight modification of increasing the stringency of washes.

B] STMS (Sequence Tagged Microsatellite Sites) Analysis

This was carried out using STMS markers linked to QTLs for contributing traits being studied based on their map positions (Table 1). Each 25µl reaction consisted of 200µM each of dATP, dTTP, dCTP and dGTP, 50 ng of primer, 50ng of genomic DNA, 1X reaction buffer and 0.6 units of Taq DNA polymerase. The reaction cycle consisted of denaturation at 94°C for 1 min followed by annealing at 52 to 60 °C 1 min (annealing temperature was modified according to the Tm of each primer pair being used) and extension at 72°C for 2 min. Thirty-five such cycles were performed after a prior denaturation at 94°C for 3 minutes. The final reaction was followed by an extension at 72°C for 5 minutes. Amplified products were separated on a 10 % polyacrylamide gel in 1X TBE buffer and were visualized by ethidium bromide staining.

Random Markers for General Heterozygosity Analysis

ISSR (Inter simple sequence repeat polymorphism) analysis was performed using 49 anchored microsatellite primers, representing di, tri, tetra and penta nucleotide repeats, obtained from University of British Columbia. PCR reactions were performed in volumes of 25 µl consisting of 100 µM each of dATP, dTTP, dCTP and dGTP, 200µM of primer, 20ng of genomic DNA, 1X reaction buffer and 0.8 units of Taq DNA polymerase. Each reaction cycle consisted of denaturation at 94°C for 1 min followed by 1 min annealing at 50°C and 2 min extension at 72°C. The annealing temperature was modified according to the Tm of each primer used for ISSR analysis. Forty-five such cycles were performed after a prior denaturation at 94°C for 3 minutes. The 45^{th} reaction was followed by a final extension at 72°C for 5 minutes. Amplified products were separated on a 1.2 to 1.4 % agarose gel in 1X TAE buffer and were visualized by ethidium bromide staining.

***TABLE 1*:** List of RFLP probes linked to or lying within the QTLs for yield and its component traits as identified by Xaio *et al* (1995) and STMS markers identified based on their map position, using the map of McCouch *et al* (1997).

Name of the mapped marker	Chromosome No. and marker type	Linked QTLs	Location of the marker
RG 480	5 (RFLP)	Plant height	Within the locus
RZ 599	2 (RFLP)	Plant height	Within the locus
RZ 495	5 (RFLP)	Plant height	Flanking the locus
RG 662	9 (RFLP)	Panicle length	Within the locus
RG 667	9 (RFLP)	Panicle length	Within the locus
RG 214	4 (RFLP)	Grains per plant, grains per panicle	Within the locus
RG 143 RZ 296	4 (RFLP) 5 (RFLP)	Grains per plant, grains per panicle Grains per plant, grains per panicle	Flanking the locus Within the locus
RZ 590	4 (RFLP)	Grains per plant, grains per panicle	Within the locus
RZ 556	5 (RFLP)	Grains per plant, grains per panicle and spikelets per plant	Within the locus
RG 864	4 (RFLP)	Grains per plant, grains per panicle, days to maturity and 1000 grain weight	Within the locus
RZ 638	11 (RFLP)	Grains per plant, grains per panicle, spikelets per plant and grain yield	Flanking the locus
RZ 390	5 (RFLP)	Spikelets per plant	Flanking the locus
RG 528	7 (RFLP)	Percent seed set	Within the locus
RG 417	7 (RFLP)	Percent seed set	Within the locus
RZ 828	6 (RFLP)	Percent seed set	Within the locus
RG 333	8 (RFLP)	Days to maturity, grain yield and 1000 grain weight	Flanking the locus
RG 256	2 (RFLP)	1000 grain weight and grain yield	Within the locus
RG 4	7 (RFLP)	1000 grain weight	Within the locus
RM 255	4 (STMS)	Grains per plant, grains per panicle, days to maturity and 1000 grain weight	Within the locus
RM 252	4 (STMS)	Grains per plant, grains per panicle, days to maturity and 1000 grain weight	Within the locus
RM249	5 (STMS)	1000 grain weight, grains per plant and grains per panicle	Within the locus
RM 241	4 (STMS)	Days to maturity, grains per plant, grains per panicle, 1000 grain weight, panicles per plant	Within the locus
RM25	8 (STMS)	Days to maturity, grain yield and 1000 grain weight	Flanking the locus
RM30	6 (STMS)	Percent seed set	Flanking the locus
RM38 RM39	8 (STMS) 5 (STMS)	Days to maturity, grain yield and 1000 grain weight Grains per plant, grains per panicle and 1000 grain weight	Flanking the locus Within the locus
RM 70	7 (STMS)	Days to maturity	Flanking the locus
RM 227	3 (STMS)	Days to maturity	Flanking the locus
RM 214	7 (STMS)	Plant height	Flanking the locus

Statistical Analysis

Similarity index values for ISSR and RFLP+ STMS patterns were calculated for all the possible pair wise comparisons, using Nei's coefficient (Nei 1987) with the help of the DOS based software program TAXAN (Swartz 1989) for general and specific heterozygosity, respectively. The genetic distances between the two CMS lines (fixed female parent) and the fourteen restorers were calculated from these similarity values. Correlation between genetic distance and field performance were established using the windows-based software package Microsoft Excel. The spreadsheets were prepared and used for both calculating the correlations and to plot graphs of genetic diversity against field performance to cross check the fidelity of the correlations for actual field applications

Results

A total of 16 rice genotypes including two CMS lines and 14 restorer lines adapted to western coastal regions of Maharashtra state (Konkan area), were subjected to molecular analysis using markers linked to eight QTLs for yield and its component traits (Table 1). These lines were also subjected to random marker (ISSR-PCR) analysis, to detect DNA polymorphism. Based on the similarity and dissimilarity of bands obtained in these DNA profiles, the genetic distances between the parental lines involved in all the twenty-eight crosses were calculated and their correlation with the field performance of their respective F_1 hybrids was estimated.

Field Performance of the Hybrids Along With Their Parents

The field data of 14 restorer lines used for crossing along with the maintainer lines of the CMS lines viz. IR 58025 B and IR 62829B, and the 28 hybrids, for grain yield and its component characters are presented in Tables 2a and 2b respectively. Performance of any cross is normally estimated on the basis of the final grain yield, which in turn is affected by several traits contributing to yield. In the crosses mentioned above, several traits individually showed very good performance and yet the grain yield was low. For example, the high heterosis observed for the character like unfilled grains per panicle masked the

positive and significant performance of trait like panicle length ultimately resulting in poor grain yield. On an average, hybrids from the CMS line IR 58025A performed better than those with CMS IR 62829A as a parent. Though CMS IR 62829A was not significantly more similar to the restorer lines than CMS IR 58025A (Figure 2 and 3). Among all 28 crosses, the best performing combination was CMS IR 58025A and Rtn 68 with respect to grain yield per plant (53.96 g), while the poorest performance was observed in the cross CMS IR 62829 A and kjt 184 (4.6g) (Table 2b). The yield of other hybrids ranged from 7.0 to 32.63 grams per plant.

DNA Polymorphism between the Parental Lines

RFLP probes and STMS markers were selected based on previous results that were found to be linked to or lying within the QTLs for yield and its component traits as identified by Xiao *et al* (1995) and McCouch *et al* (1997) (Table 1). STMS markers showed the highest percentage, 85% polymorphism, followed by ISSRs 79% and RFLPs 74%.

Markers linked to specific yield related traits were used to study the polymorphism levels in the parents. RFLP probes were hybridized with the genomic DNA digests of the parental lines and most of these were highly polymorphic with at least one of the six enzymes used. While of the 11 STMS markers used, four primer pairs produced monomorphic patterns and the remaining seven were polymorphic. The RFLP and STMS markers that demonstrated single locus variation were scored as co-dominant markers, while those showing multiple bands were scored as dominant markers.

Figure 1 depicts the RFLPs obtained when rice genomic probes, RG 864 and RG333, were hybridized to parental DNAs of the two CMS lines and their 14 restorers, digested by the restriction enzyme *Hind III*. These probes are linked to the QTL for grains per panicle, grains per plant, days to maturity, grain yield and 1000 grain weight. They are located on two different chromosomes where RG864 has been reported to be within the locus on chromosome 4, while RG333 is present flanking the QTL locus on chromosome 8.

Table 2a Field performance of the parental lines for yield and its component characters

Line/ Varieties	Plant Height (cm)	Days to maturity	Panicle length (cm)	Grains per Panicle	Spikelet fertility (%)	100 grain weight (grams)	Grain yield per plant (grams)
Lines							
IR 58025B	52.20	114.66	20.50	201.30	85.40	2.20	13.33
IR 62829B	57.20	109.00	11.00	128.00	88.40	2.20	16.23
Restorers							
RTN 73	61.96	88.00	20.30	92.06	81.61	2.43	14.76
KJT1	69.43	86.00	20.96	93.26	77.85	2.06	5.83
KJT184	71.33	102.33	19.86	90.53	72.50	1.90	6.00
KJT487	65.33	108.33	20.96	235.20	83.70	0.93	30.60
RTN711	73.40	103.00	25.23	163.06	81.90	2.13	22.56
RTN24	83.63	103.00	24.63	200.06	79.10	1.30	14.10
RTN1	79.80	101.66	20.43	124.66	82.20	2.90	20.56
PNL1	72.33	110.66	24.40	165.73	87.30	2.60	18.76
PNL2	73.26	109.66	25.10	194.20	88.31	1.86	19.13
KJT3	71.86	114.33	21.43	132.73	90.17	1.90	25.06
PLG1	68.93	125.00	26.63	266.46	90.08	1.40	23.26
RTN2	86.53	137.00	24.96	136.66	88.39	2.46	22.93
KJT14-7	85.96	130.33	24.90	142.33	85.00	1.96	12.90
RTN68	81.60	138.00	22.86	133.06	87.50	2.46	27.80

Table 2b. Field performance of the hybrids for yield and its component characters.

Hybrid	Plant height (cm)	Days to maturity	Panicle length (cm)	Grains per panicle	Spikelet fertility (%)	100 grain weight (g)	Grain yield per plant (g)	Unfilled grains per panicle
IR 58025A X RTN 73	65.5	102.33	21.13	109.73	69.43	2.10	12.30	33.2
IR 58025A X KJT 1	68.5	106.33	24.13	139.73	63.94	2.3	13	42.5
IR 58025A X KJT 184	79.16	108.66	27.20	177.06	74.95	2.26	8.00	40.51
IR 58025A X KJT 487	72.56	108.33	26.00	176.86	66.20	1.70	14.56	57.71
IR 58025A X RTN 711	71.36	110.33	25.36	176.26	78.26	2.20	13.70	38.89
IR 58025A X RTN 24	74.50	108.00	26.10	205.80	77.17	2.13	17.70	44.2
IR 58025A X RTN 1	79.43	116.33	26.93	209.73	83.44	2.43	29.73	36.29
IR 58025A X PNL 1	74.63	116.56	27.26	200.93	77.97	2.50	32.63	46.71
IR 58025A X PNL 2	65.20	120.00	25.10	197.43	48.71	2.26	10.86	117.9
IR 58025A X KJT 3	74.83	112.00	26.66	172.00	87.49	2.53	27.93	21.3
IR 58025A X PLGH 1	77.90	122.33	28.30	211.66	72.99	2.26	19.86	54.11
IR 58025A X RTN 2	81.73	123.33	26.00	199.33	76.28	2.60	19.60	46.90
IR 58025A X KJT 14-7	87.16	125.00	29.03	248.26	69.67	2.26	24.10	75.82
IR 58025A X RTN 68	87.13	121.33	28.00	282.40	87.70	2.23	53.96	33.79
IR 62829A X RTN 73	66.13	102.00	21.96	96.40	77.56	2.20	9.80	21.91
IR 62829A X KJT 1	70.56	106.66	21.30	112.86	65.73	2.20	10.73	34.2
IR 62829A X KJT 184	64.90	105.66	21.96	95.26	54.34	1.76	4.60	44.32
IR 62829A X KJT 487	59.33	106.33	19.06	100.53	78.23	1.80	8.06	21.6
IR 62829A X RTN 711	63.83	106.00	18.73	77.53	79.72	1.76	7.06	15.2
IR 62829A X RTN 24	74.60	123.00	24.03	177.33	66.99	2.80	20.20	60.11
IR 62829A X RTN 1	67.40	123.33	21.16	134.80	72.66	2.10	7.00	58.13
IR 62829A X PNL 1	65.83	107.66	18.80	83.96	76.62	2.30	9.60	18.08
IR 62829A X PNL 2	68.26	114.66	25.03	162.46	88.69	2.50	32.60	18.77
IR 62829A X KJT 3	64.03	107.33	15.00	97.80	78.35	1.96	15.43	21.4
IR 62829A X PLGH 1	74.80	118.33	22.93	161.56	65.74	2.40	14.43	57.60
IR 62829A X RTN 2	64.26	106.00	20.06	104.60	83.12	2.20	12.36	17.6
IR 62829A X KJT 14-7	77.66	122.00	27.10	220.26	69.72	2.26	30.40	65.87
IR 62829A X RTN 68	76.06	106.66	23.50	161.60	81.20	2.13	26.40	30.5
S.E + / -	1.766	0.887	0.644	11.970	4.532	0.059	2.387	4.25
C. D at 5 %	4.969	2.496	1.812	33.680	12.752	0.166	6.716	11.9
C.D at 1%	6.592	3.310	2.402	44.682	16.917	0.219	8.910	15.725

Figure 2 shows a pattern obtained when rice genomic probes RG256 (chromosome 2), RG333 (chromosome 8) and RG638 (chromosome 11) linked to the QTL for grain yield, were hybridized to genomic DNAs of the 2 CMS and 14 restorer parental lines, digested using the restriction enzyme *Dra I.* Of the three probes being used, RZ256 has been reported to be lying within the locus of QTL for grain yield, while RG333 and RZ638 have been shown to lie in the regions flanking the QTLs on their respective chromosomes.

After carrying out the specific markers analysis, ISSR markers were used for spanning the genome randomly, as in my earlier study, Chapter 2 (Section I) they were found to be polymorphic and informative. Of the 49 ISSR markers being amplified, 25 primers showed polymorphic patterns, generating 239 amplified products of which 51 were monomorphic. Among the 25 primers, 17 polymorphic bands were produced by the primer $(GATA)_4$. Though dinucleotide repeats are more frequent than tetra nucleotide repeats in the rice genome, this observation could probably be attributed to the distribution of these repeats in the genome, which promotes their better amplification.

ISSR markers were informative and produced a total of 29 putative variety specific bands. Seven of these bands were amplified for restorer line Rtn2, while three bands each were amplified for Rtn1, Kjt1 and Kjt14-7. Two bands each were amplified for Kjt3, Kjt184, Kjt487 and Plg1, whereas 1 band each was obtained for CMS 62829A, Rtn24, Rtn711, Pnl1 and Pnl2. Two primers, UBC 880 and 888, produced one band each, which specifically amplified in the two CMS lines. Primer 887 produced one band that was specifically amplified in all the restorer lines but did not amplify in the two CMS lines. Figure 3 represents a PCR amplification profile obtained using anchored primer ISSR 856 $[(CA)_8YA]$. As is evident in this figure, the primer amplifies 9 bands (dominant alleles differentiated on the basis of their band size) out of which 6 are polymorphic.

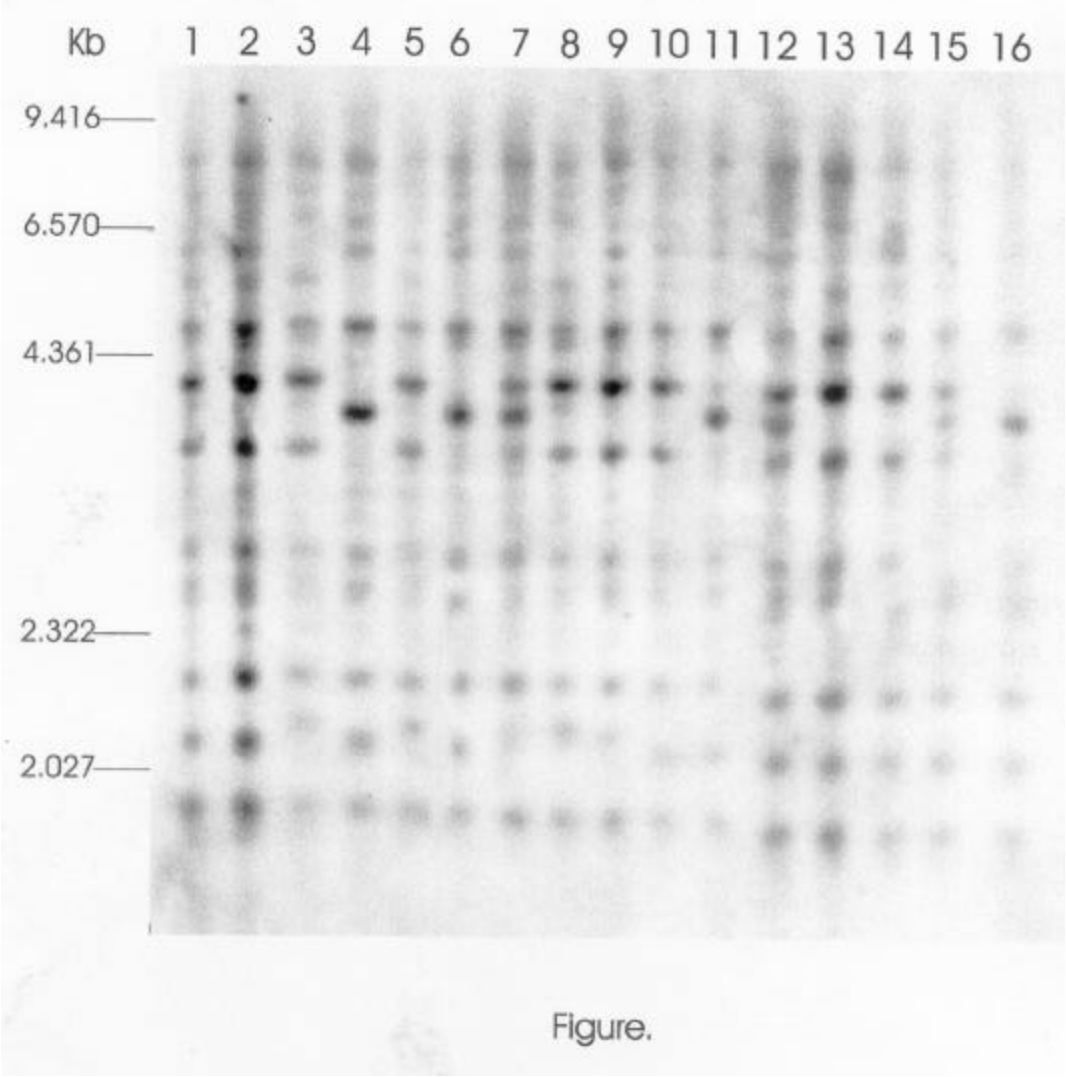

Figure 1: Hybridization patterns of *Hind III* digests of rice parental DNAs detected by rice genomic probes RG 864 and RG333 linked to QTLs for grains per panicle, grains per plant, days to maturity, grain yield and 100 grain weight, located on chromosome 4 and 8 respectively. Lanes 1to 16 are (1) CMS IR 58025 A, (2) CMS IR 62829 A, (3) RTN 1, (4) RTN2, (5) RTN 24, (6) RTN 68, (7) RTN 73, (8) RTN 711, (9) KJT 1, (10) KJT 3, (11) KJT 14-7, (12) KJT 184, (13) KJT 487, (14) PNL 1, (15) PNL 2 and (16) PALGH 1 respectively. DNA marker is λ *Hind III* digest.

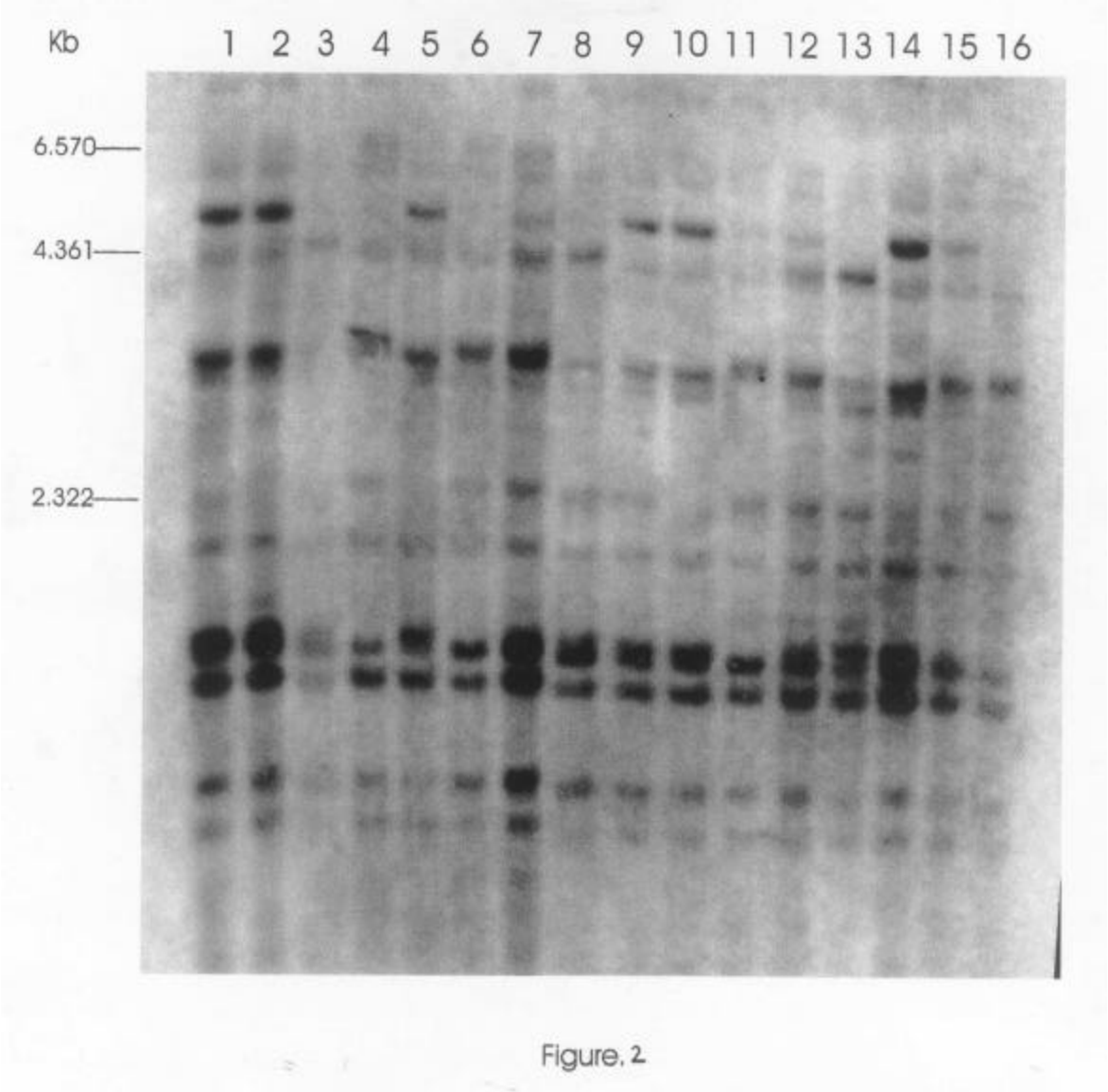

Figure. 2

Figure 2: Hybridization patterns of Dra *I* digests of rice parental DNAs detected by rice genomic probes RG 256, RG333 and RZ 638 linked to QTLs for grain yield, located on chromosome 2, 8 and 11 respectively. Lanes 1to 16 are (1) CMS IR 58025 A, (2) CMS IR 62829 A, (3) RTN 1, (4) RTN2, (5) RTN 24, (6) RTN 68, (7) RTN 73, (8) RTN 711, (9) KJT 1, (10) KJT 3, (11) KJT 14-7, (12) KJT 184, (13) KJT 487, (14) PNL 1, (15) PNL 2 and (16) PALGH 1 respectively. DNA marker is λ *Hind III* digest.

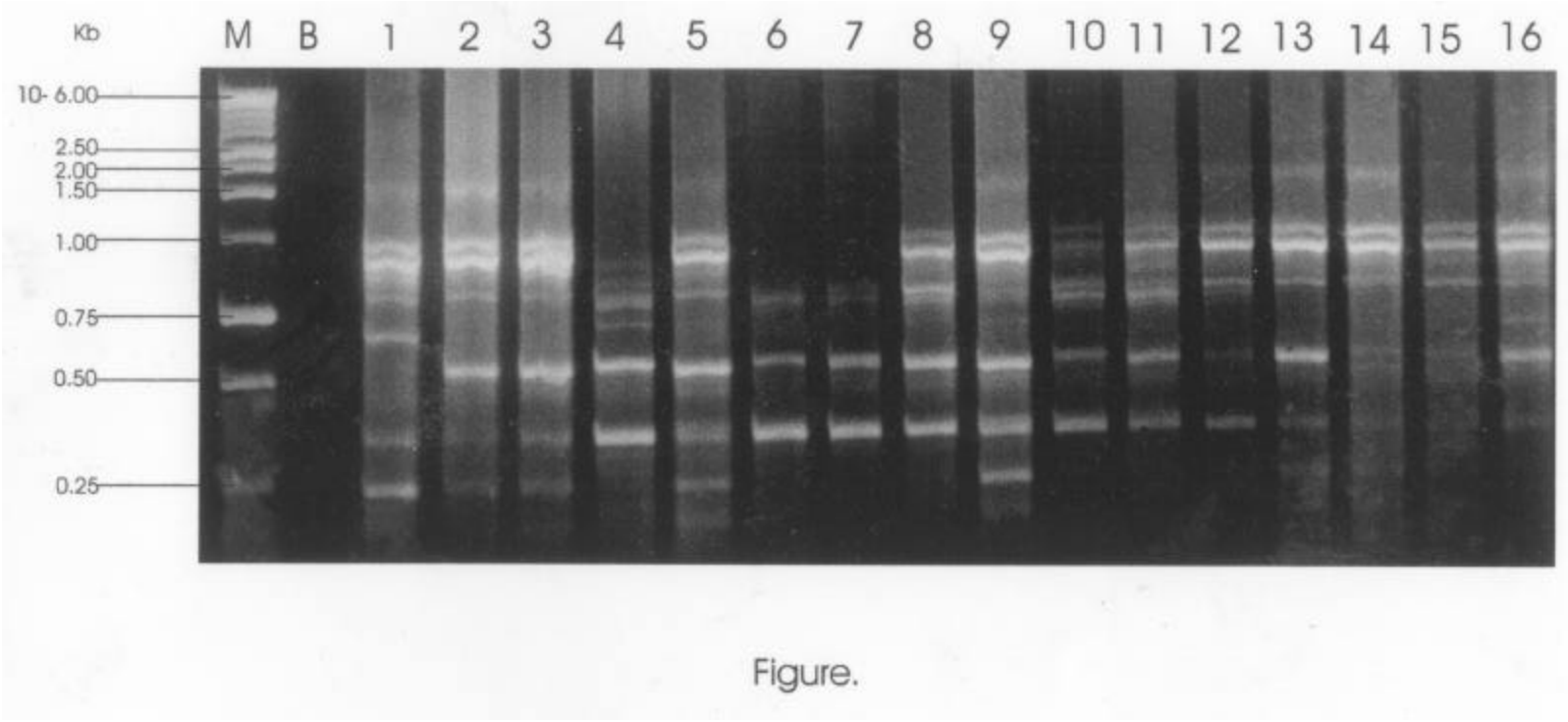

Figure 3: PCR amplification profile obtained using anchored microsatellite primer ISSR 856 [$(AC)_8$ YA]. Lane M: 1 Kb ladder (Stratagene), Lane B: Control and lanes 1to 16 are (1) CMS IR 58025 A, (2) CMS IR 62829 A, (3) RTN 1, (4) RTN2, (5) RTN 24, (6) RTN 68, (7) RTN 73, (8) RTN 711, (9) KJT 1, (10) KJT 3, (11) KJT 14-7, (12) KJT 184, (13) KJT 487, (14) PNL 1, (15) PNL 2 and (16) PALGH 1.

w

The genetic distance range for these markers was between 0.204 to 0.348, for RFLP + STMS markers and 0.180 to 0.293 for ISSR markers, for the CMS IR 58025A in combination with set of restorer parents. Similarly the genetic distance ranged between 0.174 to 0.341 and 0.205 to 0.305 respectively, for the CMS IR 62829A and restorer lines. The dendrograms based on these similarity indices produced two different patterns of clustering for the 16 parental genotypes as noted in the two dendrograms (Figure 4 and 5). Remarkably the CMS line IR 58025A appears to be more close to CMS IR 62829A in both the dendrograms.

Hybrid Performance and Heterosis Based on Specific Heterozygosity

Genetic distances were calculated for all the parental lines, based on markers linked to each of the eight traits (Table 1) separately. The correlation between these GD estimates and the field performance for the trait on which the GD QTL estimates were based were then computed. Genetic distances were also calculated based on all the RFLP and STMS markers used (Table 1) and the correlation between these estimates and field performance of each trait were estimated in the correlation values, in each of these analyses (Table 3 and 4).

100- grain weight showed positive significant correlation with GD based only on markers linked to 100- grain weight QTL for heterosis and heterbeltiosis, in CMS IR 58025A crosses (Table 3). On the other hand, they were negatively correlated for hybrid performance and heterbeltiosis in CMS IR 62829A crosses.

Most significant and positive correlation for hybrid performance as well as heterosis and heterbeltiosis, was observed in the CMS IR 62829A crosses when the field values for grain yield were regressed on the genetic distance calculated using probes that were linked or lie within the QTLs for grain yield (Table 3). The banding patterns with these probes were exactly identical for CMS IR 58025A and 62829A (Figure2), thereby implying that both of them share the same genetic distance from each of the 14 restorers. Only in the CMS 29 A crosses,

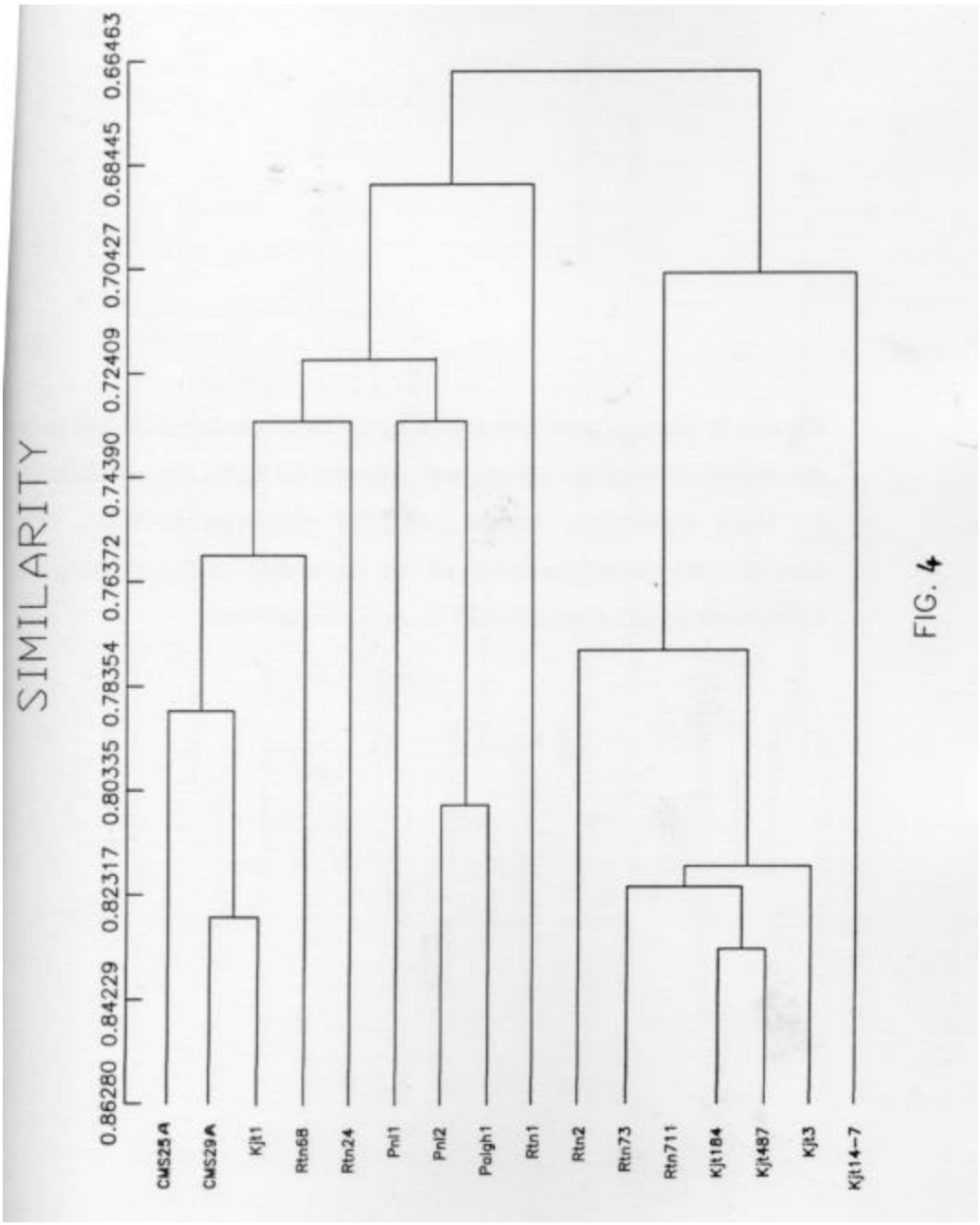

Figure 4: Dendrogram generated by UPGMA cluster analysis (TAXAN) using Nie's coefficient. The genetic similarity was determined on the basis of banding profiles generated by Restriction Fragment Length Polymorphism and Sequence Tagged Microsatellite Site markers linked to QTLs contributing to yield and its component traits. The dendrogram shows clustering of the 16 parental lines including 14 restorer lines and two CMS lines 58025 A and 62829 A.

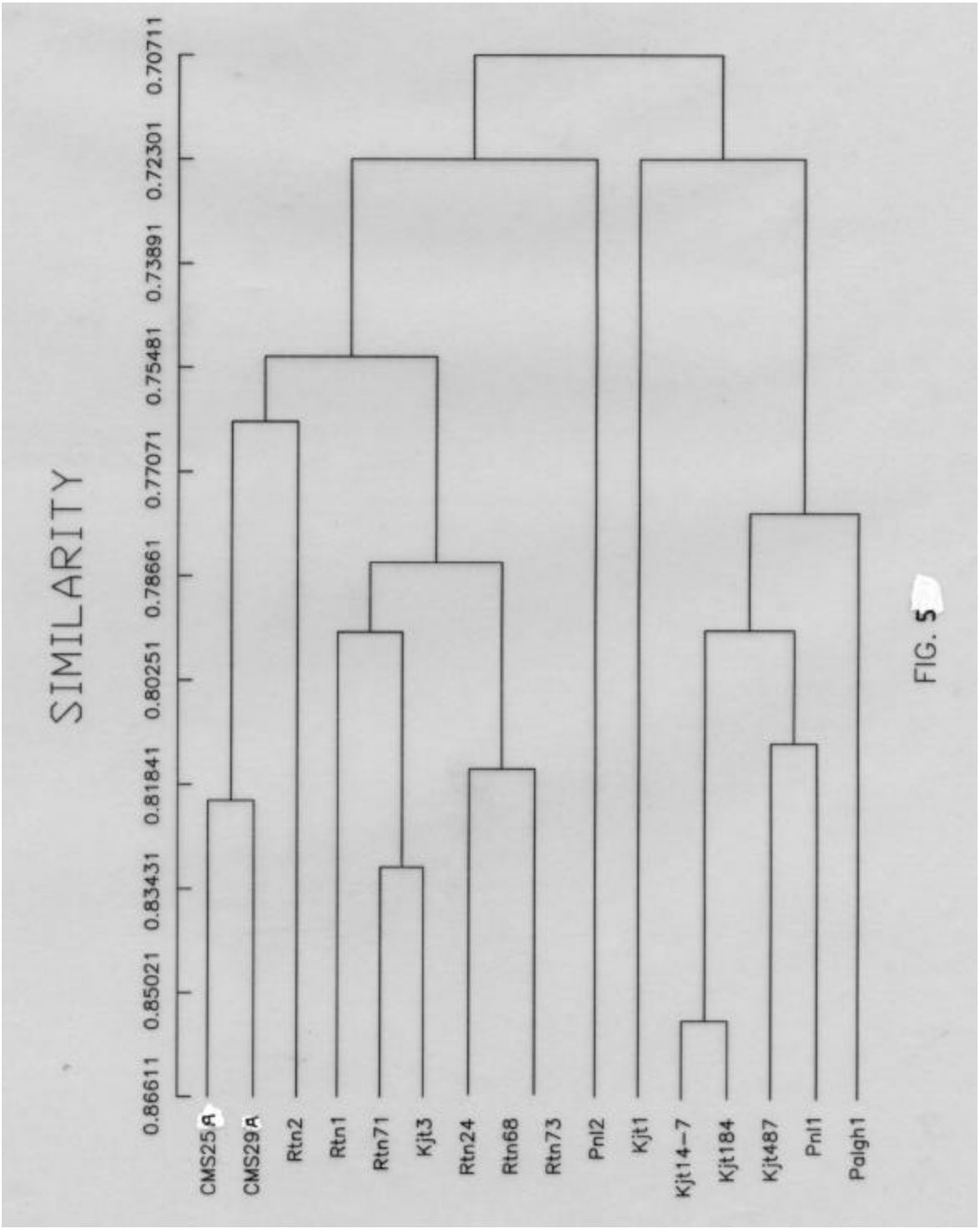

Figure 5: Dendrogram generated by UPGMA cluster analysis using the software program Taxan, with genetic similarity values deduced by Nie's coefficient, based on ISSR genotype profiles. The dendrogram shows clustering of the 16 parental lines including two CMS lines 58025 A and 62829 A, and 14 restorers.

however, was QTL associated genetic distance significantly related to hybrid performance or heterosis.

When the genetic distances between the parental lines based on all the specific heterozygosity marker was used for regression analysis (Table 4), many traits showed positive and significant correlations especially for hybrid performance. In general, the CMS IR 58025A x restorer crosses showed more significant correlation values than CMS62829A x restorer crosses. Almost all the traits like plant height, panicle length, grains per panicle, unfilled grains per panicle and days to maturity, showed positive and significant correlations (P = 0.01) in CMS IR 58025A and CMS IR 62829A crosses. Heterosis for traits like unfilled grains, grains per panicle, panicle length and days to maturity was observed to be significant, while heterbeltiosis was observed to be significant for traits like plant height, grains per panicle and unfilled grains per panicle. In the case of CMS IR 62829A combinations with different restorers, heterosis as well as heterobeltiosis was observed only in the case of unfilled grains per panicle, but there was no correlation with grain yield.

Genetic Distance and Regression Analysis Based on Random Markers

The genetic distances were also calculated based on DNA profiles obtained using ISSR PCR approach and used for regression analysis for performance and heterosis (Table 5). Most of the correlations derived from the ISSR marker analysis (Table 5), indicate that crosses involving the line CMS IR 58025A as one of the parents, showed slightly better correlations than those involving CMS IR 62829A. Only trait which showed positive and significant correlation for hybrid performance as well as mid- parent (P + 0.01) and heterbeltiosis (P = 0.05) was unfilled grains per panicle in the case of the CMS IR 58025A crosses. The traits that showed significant but negative correlations in the case of CMS IR 62829A crosses were panicle length and grains per panicle (P = 0.05). Other traits such as spikelet fertility and grain yield showed negative correlation but the P values

Table 3: Correlations of genetic distance between the parental lines for a trait, based on specific heterozygosity of subset of marker loci linked to QTLs affecting that trait, with hybrid performance, heterosis and heterobeltiosis

Character	**Hybrid performance**		**Heterosis**		**Heterobeltiosis**	
	CMS25A	**CMS29A**	**CMS25A**	**CMS29A**	**CMS25A**	**CMS29A**
Plant Height	-0.256	-0.184	-0.162	-0.020	-0.257	-0.320
Panicle Length	+0.233	+0.143	+0.418**	+0.268	+0.437**	+ 0.371*
Grains per panicle	+0.128	+0.256	+0.020	-0.210	+0.111	+0.212
Spikelet fertility	- 0.060	+0.129	- 0.108	+0.007	-0.271	+ 0.109
Days to maturity	+0.369*	+0.379*	+0.277	+0.376*	-0.137	+ 0.362*
100 grain weight	- 0.125	-0.390*	+0.379*	-0.268	+ 0.429**	-0.382*
Grain Yield	+0.024	+0.640**	+0.003	+0.599**	+0.038	+0.628**

CMS25A: Cytoplasmic Male Sterile line IR 58025A

CMS29A: Cytoplasmic Male Sterile line IR 62829A

*P = 0.05 ** P = 0.01

Note: In this table, the trait Unfilled grains per panicle is not included, as individual markers linked to this character have not been analyzed in our study. However, the field data of this trait was correlated with the GD based on all the markers linked to different traits together, in Table 4.

Table 4: Correlations between genetic distance based on all 20 RFLP loci and 7 polymorphic STMS markers linked to or lying within QTLs affecting yield and hybrid performance, heterosis and heterobeltiosis.

Character	**Hybrid performance**		**Heterosis**		**Heterobeltiosis**	
	CMS25A	**CMS29A**	**CMS25A**	**CMS29A**	**CMS25A**	**CMS29A**
Plant Height	+0.465**	+0.192	+0.286	-0.205	+0.465**	+0.314
Panicle Length	+0.659**	+0.287	+0.342*	+0.165	+0.067	+0.159
Grains per panicle	+0.613**	+0.425**	+0.360*	+0.258	+0.468**	+0.285
Unfilled grains per panicle	+0.466**	+0.506**	+0.428**	+0.511**	+0.445**	+0.498**
Spikelet fertility	-0.077	+0.012	-0.159	-0.053	+0.086	+0.012
Days to maturity	+0.807**	+0.672**	+0.329*	+0.128	-0.121	+0.052
100- grain weight	+0.294	+0.164	+0.161	+0.076	+0.176	+0.206
Grain yield	+0.25	+0.102	+0.244	+0.026	+0.323	+0.069

CMS25A: Cytoplasmic Male Sterile line IR 58025A

CMS29A: Cytoplasmic Male Sterile line IR 62829A

*P = 0.05 ** P = 0.01

Table 5: Correlations of genetic distance with hybrid performance and heterosis over mid-parent and better parent based on 188 Inter simple sequence repeat polymorphism loci.

Character	**Hybrid performance**		**Heterosis**		**Heterobeltiosis**	
	CMS25A	**CMS29A**	**CMS25A**	**CMS29A**	**CMS25A**	**CMS29A**
Plant Height	-0.213	+0.012	-0.155	+0.106	-0.214	+0.038
Panicle Length	+0.102	+0.383*	+0.017	+0.194	-0.038	+0.159
Grains per panicle	+0.151	+0.256	-0.203	-0.348*	-0.197	+0.145
Unfilled grains per panicle	+0.526**	+0.096	+0.482**	-0.009	+0.377*	+0.074
Spikelet fertility	-0.627**	+0.005	-0.575**	-0.010	-0.068	+0.017
Days to maturity	+0.038	+0.273	+0.289	+0.108	+0.308	-0.103
100- grain weight	-0.137	+0.042	+0.207	+0.285	+0.071	+0.306
Grain yield	-0.519**	+0.211	-0.327*	+0.238	-0.218	+0.256

CMS25A: Cytoplasmic Male Sterile line IR 58025A

CMS29A: cytoplasmic male sterile line IR 62829A

*P = 0.05

** P = 0.01

were 0.01 and 0.05, respectively for heterosis and hence, were significant. Thus indicating that only a few correlations are significant, mostly in CMS 25 A crosses, while the remaining represent an unfavorable correlation between GD and Hybrid performance and / or heterosis.

Genetic Distance and Performance/ Heterosis/ Heterobeltiosis: Do Significant Correlations Signify Predictive Values?

In order to check the fidelity of the correlation values and their applicability for field use, it was important to confirm that the positive significant correlation with field performance implied increase in hybrid performance or heterosis, with increase in the genetic distance. Among all the eight characters analyzed (Table 4), whenever a positive significant correlation was observed, it was expected that, as the genetic distance would increase, a better hybrid performance, heterosis and heterobeltiosis would be observed for that particular trait. Hence, the values of field performance for each trait were plotted against the average genetic distance obtained using specific heterozygosity markers. Figures 6 to 13 represent the plots drawn for all the eight traits being analyzed viz. Plant height, panicle length, grains per panicle, unfilled grains per panicle, spike let fertility, days to maturity, and 100-grain weight and grain yield against average genetic distance between the parental lines respectively. Interestingly the profiles showed more or less similar trend for all the eight traits analyzed and it became evident that even when the correlations were significant, they were not of practical utility. The best and comparatively poor performers (among the 14 restorers) were observed to share the same or very close genetic distance from the fixed CMS parental lines IR 58025A and IR 62829A. Varieties like Plg1 and Rtn2, were the most distant restorer lines from CMS lines IR 58025 A and IR 62829A, respectively, and hence, were expected to perform best, especially for the traits where correlation was observed to be positive and significant, but they did not perform as expected. Similarly the restorer lines that shared same or more or less similar genetic distances from the CMS parents, did not exhibit the same levels of performance or heterosis or heterobeltiosis. Only on rare occasions, did two different restorers

sharing similar genetic distance with CMS lines show comparable performance. For example, Pnl2 and Rtn 711, which shared more or less the same genetic distance from CMS IR 62829A and Matched in their performance, heterosis and heterobeltiosis for the trait unfilled grains per panicle.

The poor predictive power of marker based diversity estimates can be illustrated by the example of Rtn 68 which had a genetic distance of 0.244 only from CMS IR 58025 A, and therefore this cross was not expected to produce a better than average hybrid. Never the less this hybrid was the highest yielding (Table 2b) and exhibited more heterosis for the traits, panicle length, grains per panicle, and days to maturity, than many other crosses in which the parents were genetically more divergent. (Figures 6-13).

Discussion

Selection of Germplasm for Developing Hybrid Rice

One of the major breakthroughs in the marked increase in national grain yield and cropping intensity of rice was the development of F1 hybrids in the Peoples Republic of China during the late 1970s. CMS lines were later bred at International Rice Research Institute in 1988-89 (IR 58025 A and IR 62829 A) and in India (Punjab male sterile lines) in 1990 (Virmani 1994) where constant efforts were made to increase yield through hybrid rice technology. A large number of CMS lines are now being developed, all over the world with IR 58025A and IR62829A, as the most popular lines which are well adapted to various environments, and also share the same cytoplasm (WA type) but different nuclear factors (Virmani 1994). These lines were found to be the most suitable lines for hybrid program in the region where the crosses for my studies were carried out. I therefore, focussed on these two lines in our analysis and used them for developing hybrid rice by crossing with fourteen well-adapted local varieties as restorers. The field work was carried out in the Konkan region of Maharashtra, which has peculiar agroclimatic conditions including very heavy rainfall and a great deal of variation in the soil type, ranging from saline soils of seashore to very light

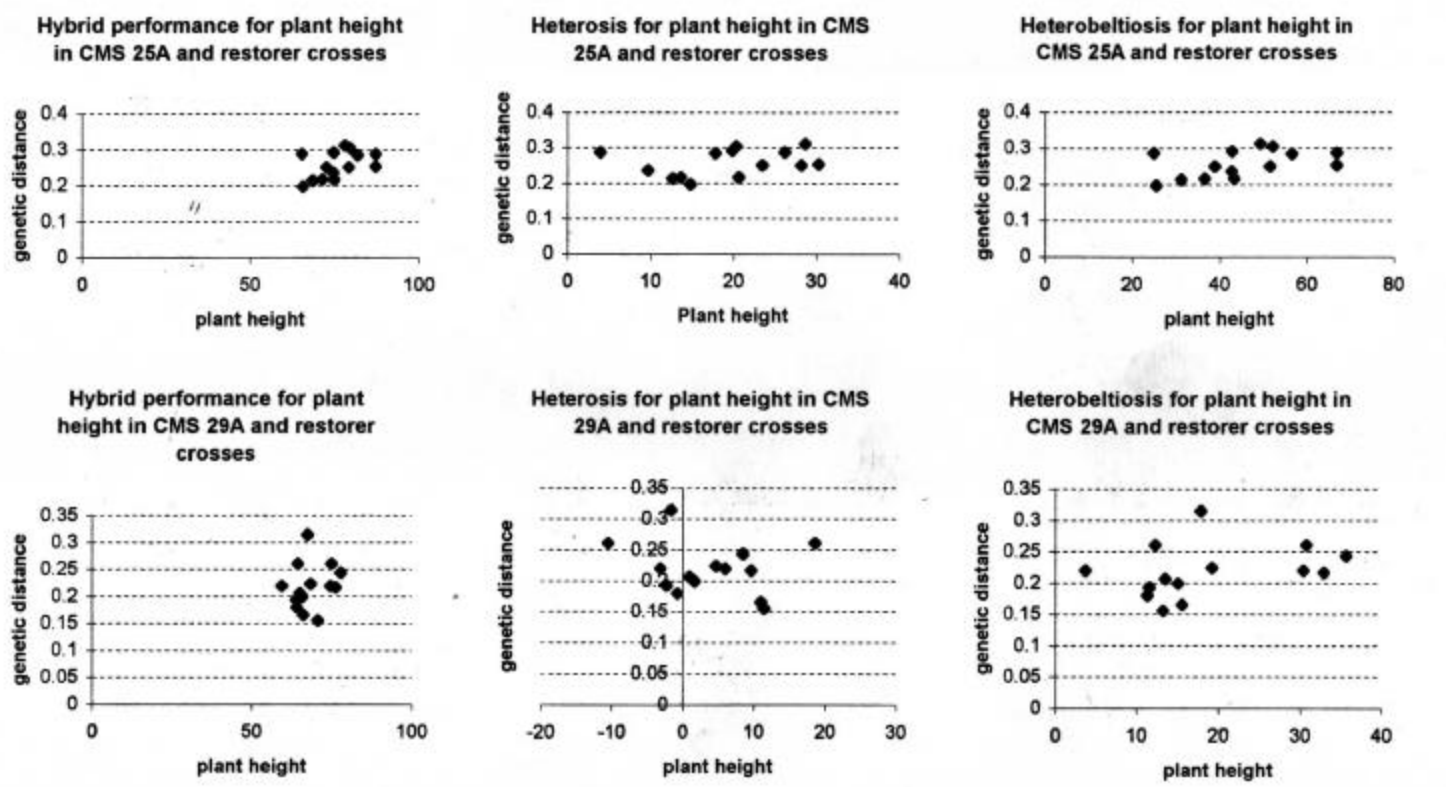

Figure 6

Graphs showing the relationship between genetic distance and hybrid performance, heterosis and heterobeltiosis in CMS IR 58025 A and CMS IR 62829 A crosses with the 14 restorer lines respectively. Note the variation in Hybrid performance, heterosis and heterobeltiosis, observed at the same genetic distance between the parental lines, for plant height,

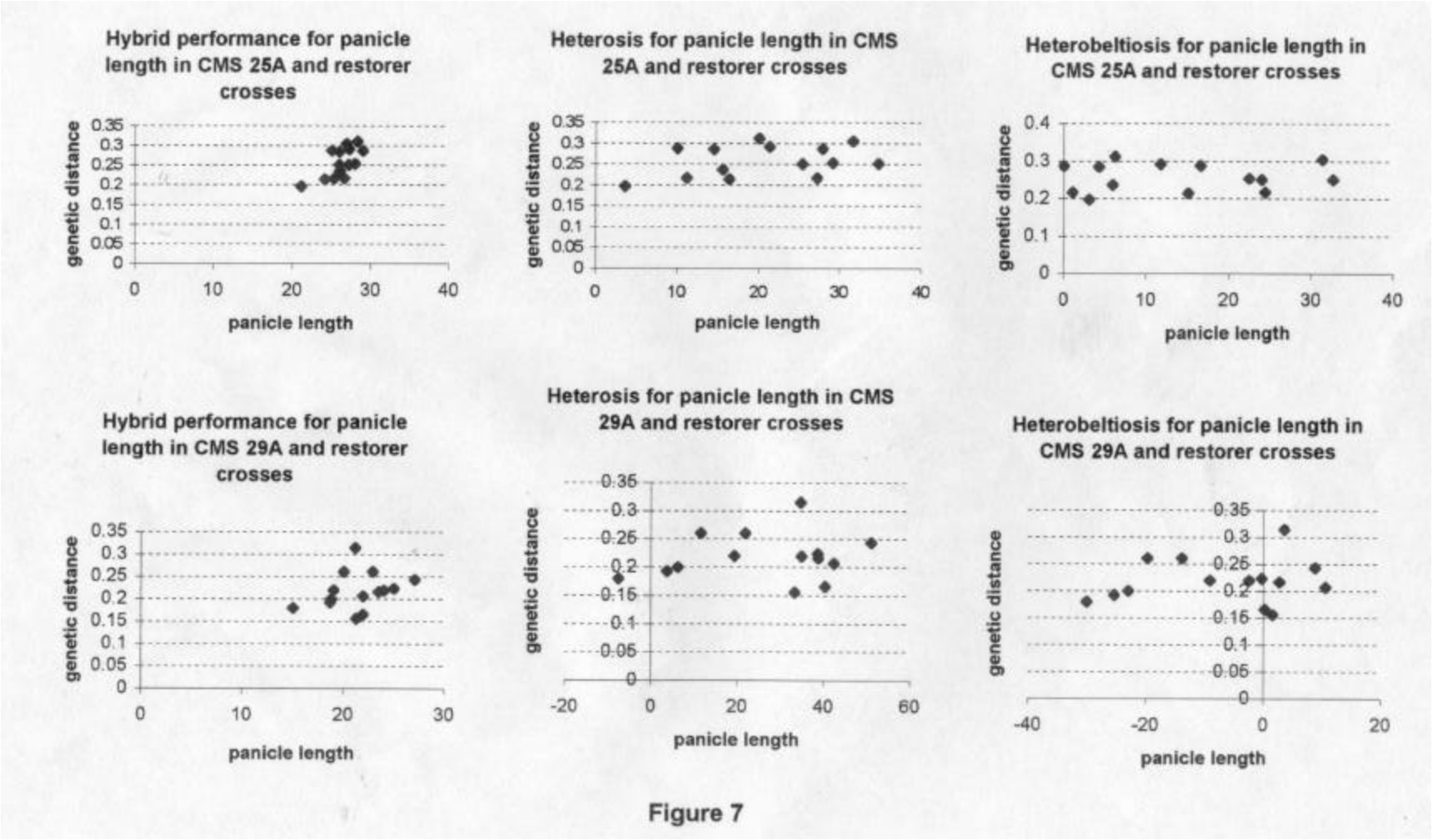

Figure 7

Graphs showing the relationship between genetic distance and hybrid performance, heterosis and heterobeltiosis in CMS IR 58025 A and CMS IR 62829 A crosses with the 14 restorer lines respectively. Note the variation in Hybrid performance, heterosis and heterobeltiosis, observed at the same genetic distance between the parental lines, for Panicle length

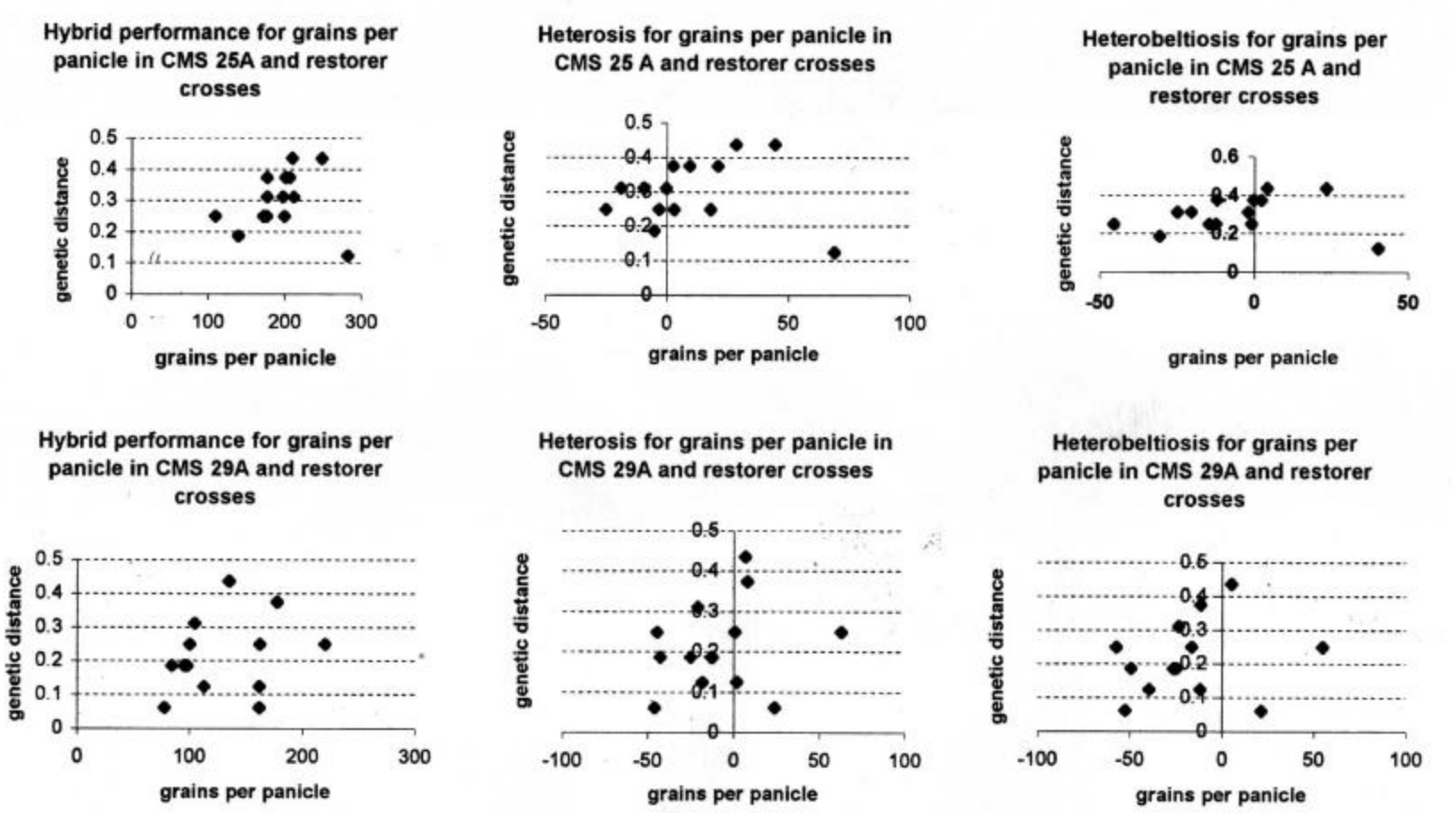

Figure 8

Graphs showing the relationship between genetic distance and hybrid performance, heterosis and heterobeltiosis in CMS IR 58025 A and CMS IR 62829 A crosses with the 14 restorer lines respectively. Note the variation in Hybrid performance, heterosis and heterobeltiosis, observed at the same genetic distance between the parental lines, for grains per panicle

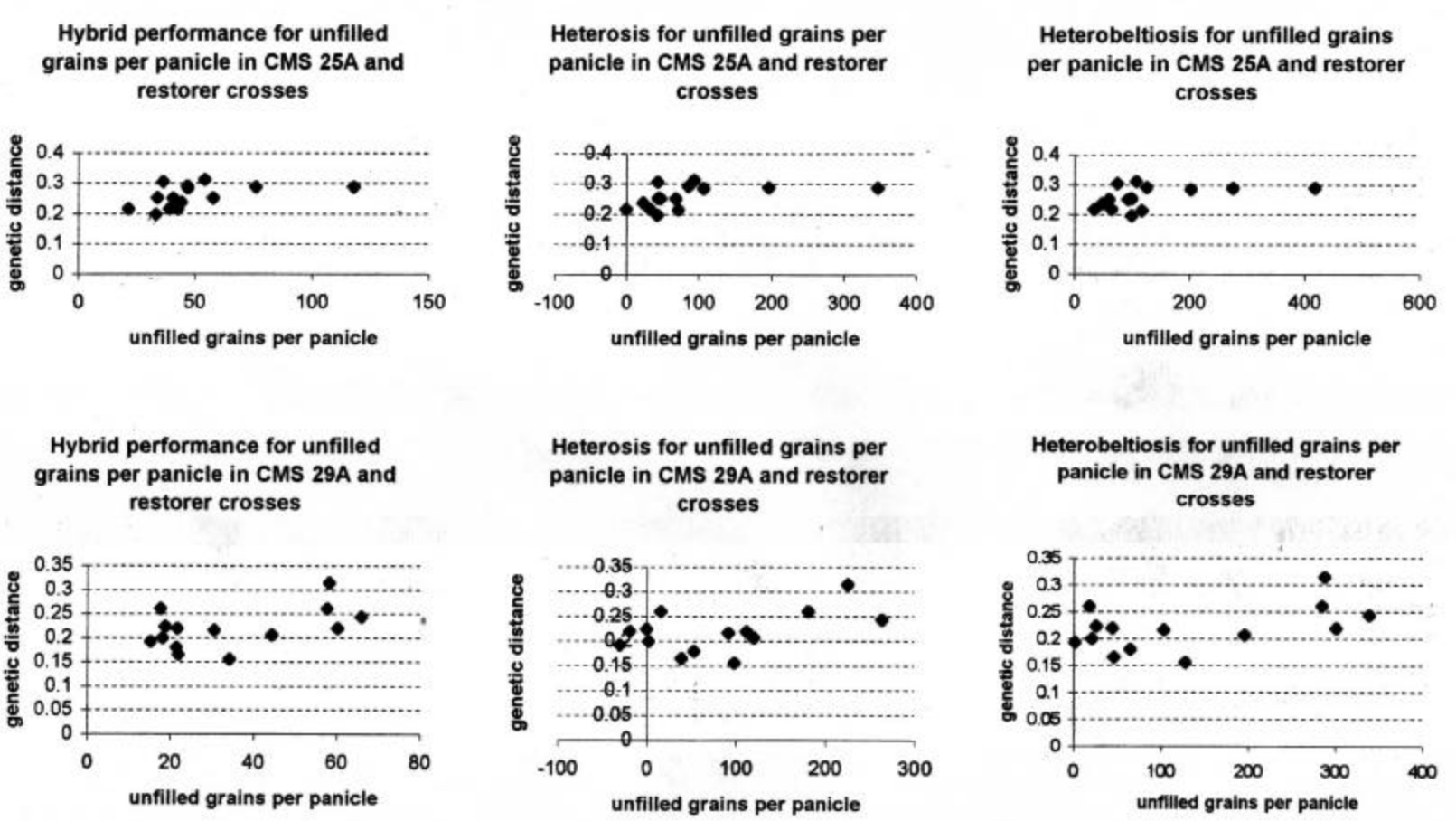

Figure 9

Graphs showing the relationship between genetic distance and hybrid performance, heterosis and heterobeltiosis in CMS IR 58025 A and CMS IR 62829 A crosses with the 14 restorer lines respectively. Note the variation in Hybrid performance, heterosis and heterobeltiosis, observed at the same genetic distance between the parental lines, for unfilled grains per panicle

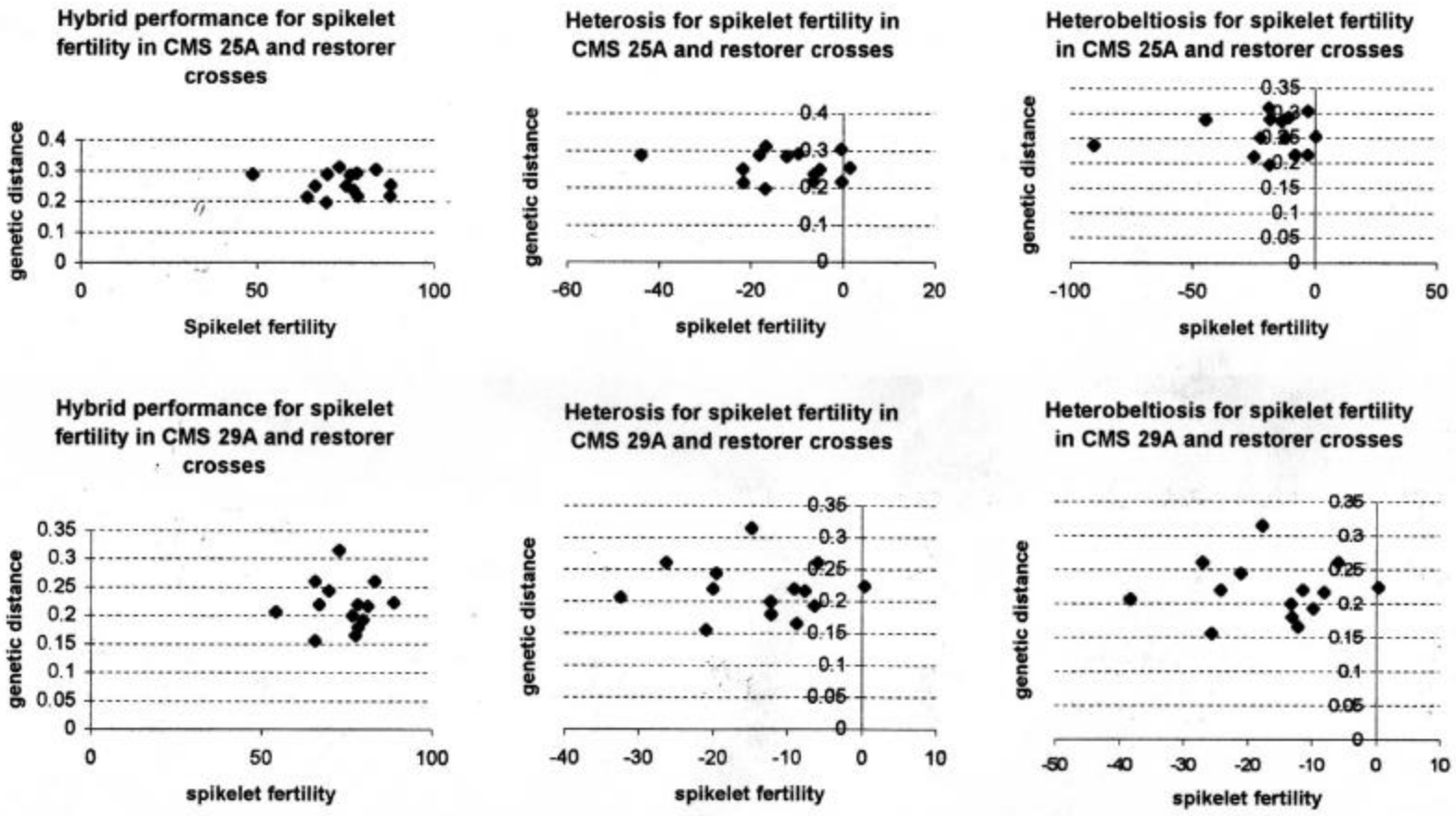

Figure 10

Graphs showing the relationship between genetic distance and hybrid performance, heterosis and heterobeltiosis in CMS IR 58025 A and CMS IR 62829 A crosses with the 14 restorer lines respectively. Note the variation in Hybrid performance, heterosis and heterobeltiosis, observed at the same genetic distance between the parental lines, for spikelet fertility

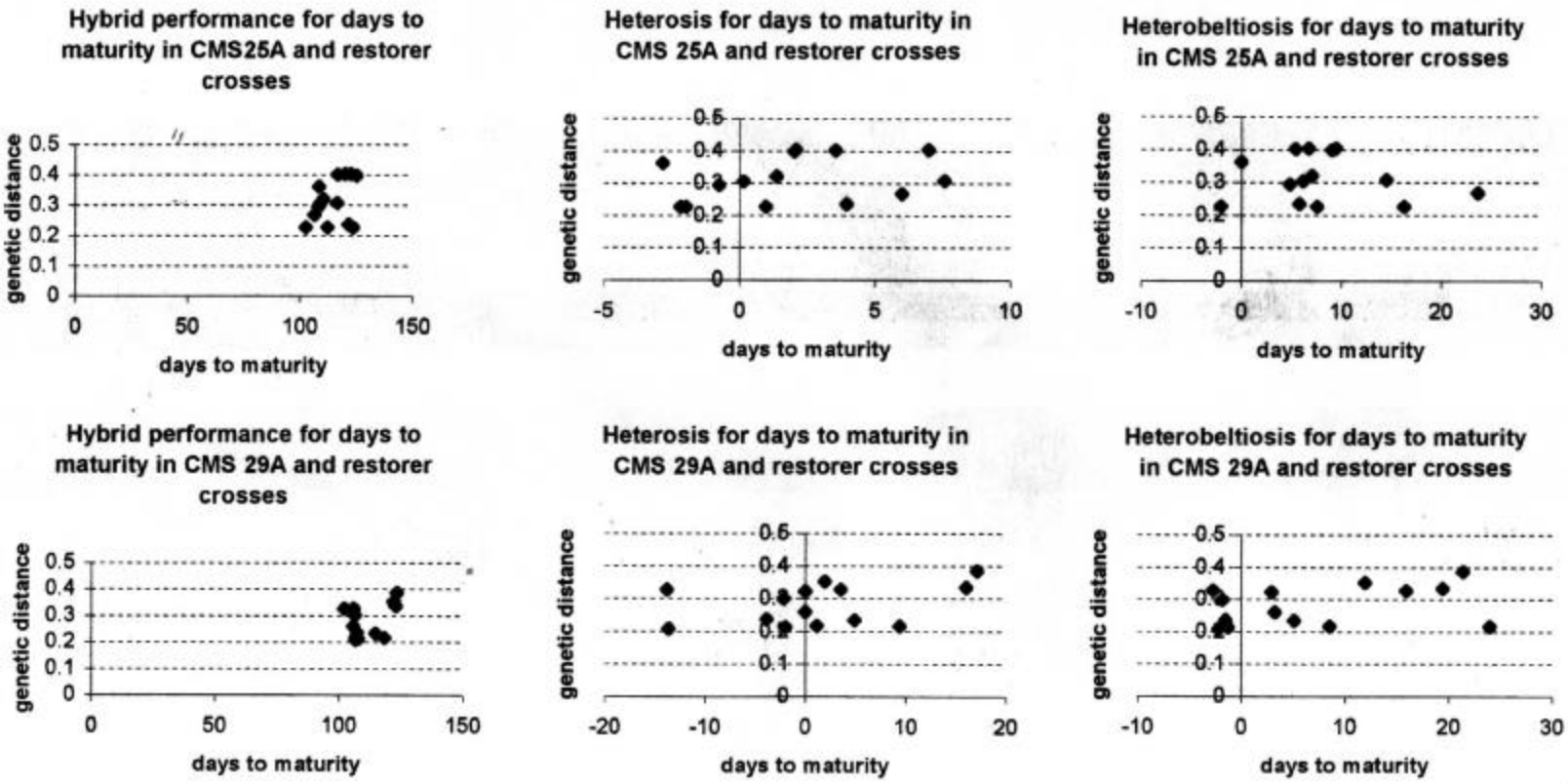

Figure 11

Graphs showing the relationship between genetic distance and hybrid performance, heterosis and heterobeltiosis in CMS IR 58025 A and CMS IR 62829 A crosses with the 14 restorer lines respectively. Note the variation in Hybrid performance, heterosis and heterobeltiosis, observed at the same genetic distance between the parental lines, for days to maturity

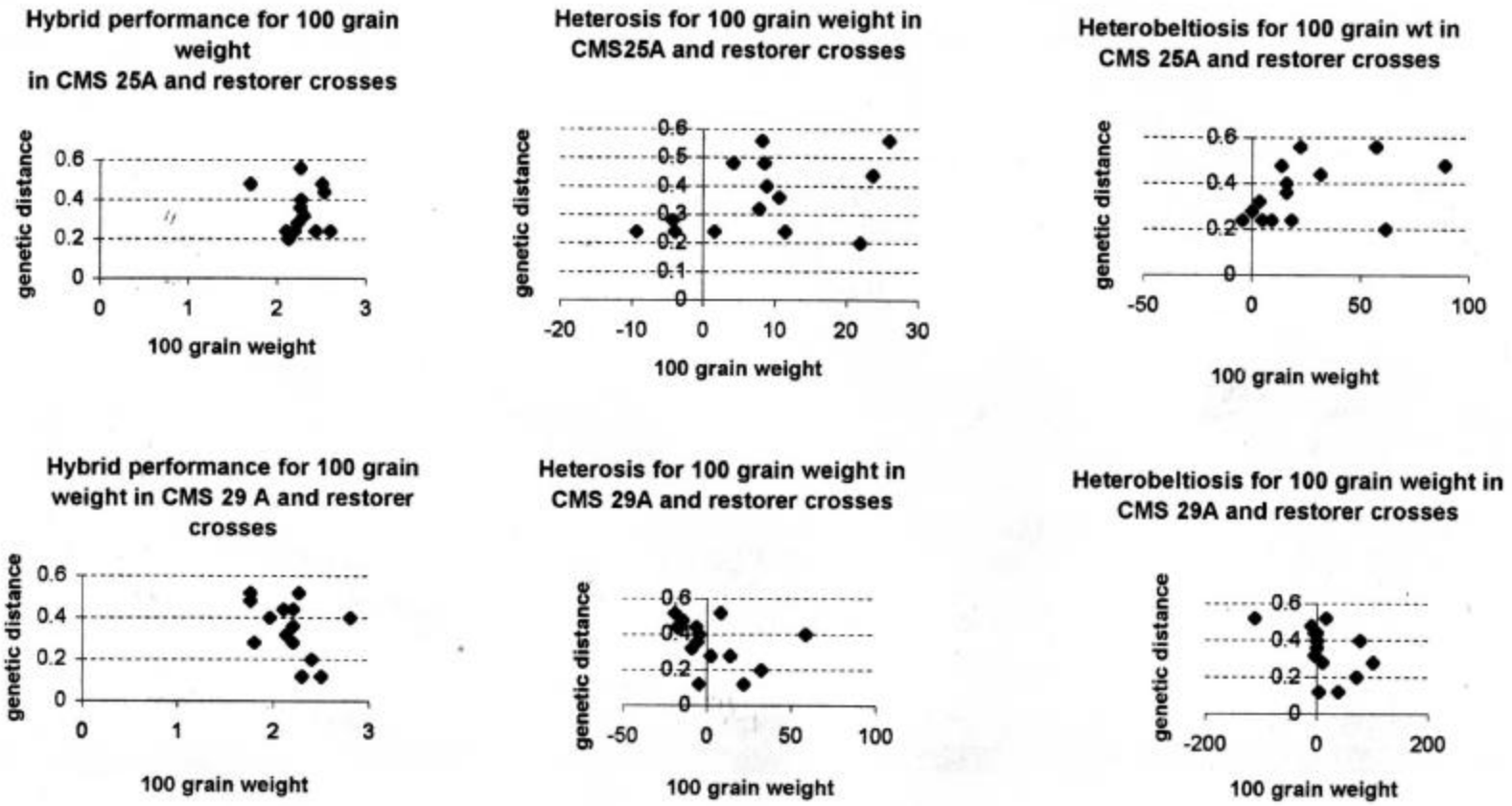

Figure 12

Graphs showing the relationship between genetic distance and hybrid performance, heterosis and heterobeltiosis in CMS IR 58025 A and CMS IR 62829 A crosses with the 14 restorer lines respectively. Note the variation in Hybrid performance, heterosis and heterobeltiosis, observed at the same genetic distance between the parental lines, for 100 grain weight

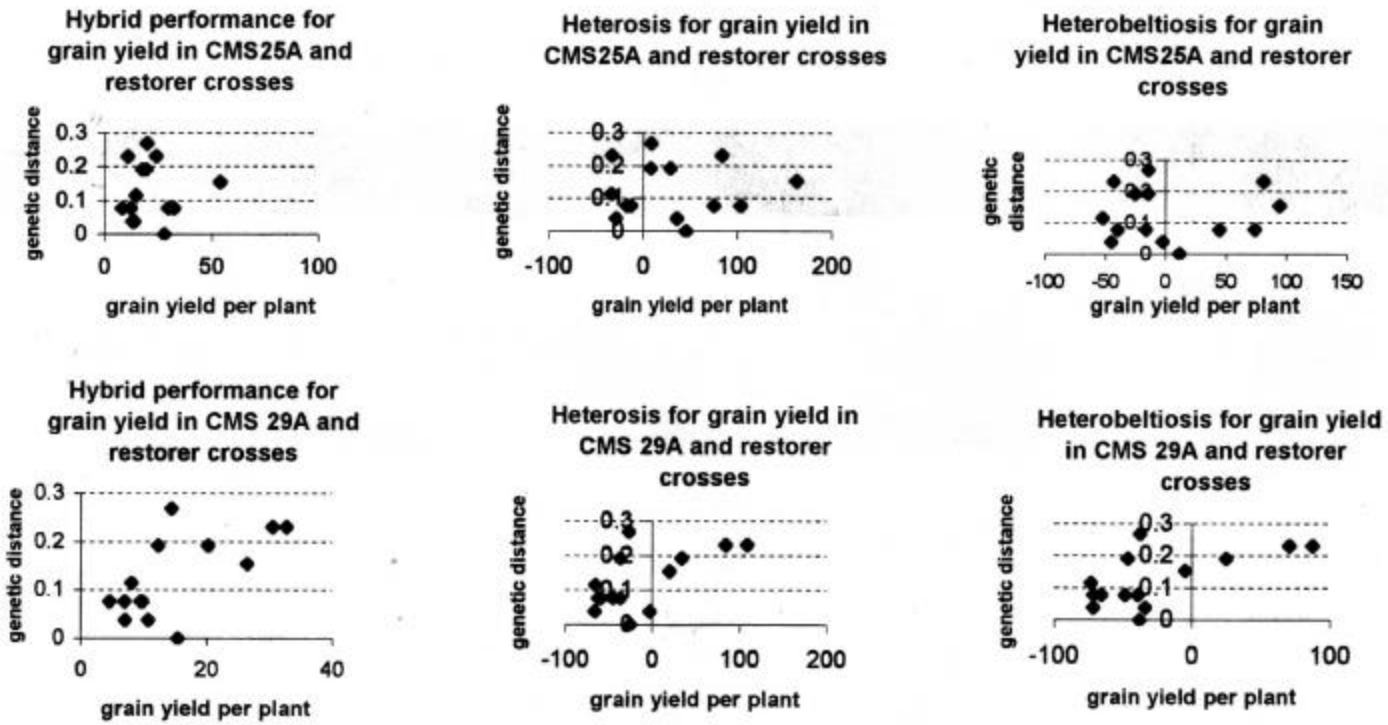

Figure 13

Graphs showing the relationship between genetic distance and hybrid performance, heterosis and heterobeltiosis in CMS IR 58025 A and CMS IR 62829 A crosses with the 14 restorer lines respectively. Note the variation in Hybrid performance, heterosis and heterobeltiosis, observed at the same genetic distance between the parental lines, for grain yield per plant

soils on hill slopes, necessitating the development of very early, early, mid late, late and very late maturing rice varieties. Fourteen well- adapted local varieties, which were found to be suitable as restorers in the test crosses, were used in the present study for developing hybrid rice for these regions.

Impact of DNA Markers on Breeding Programs

Based on the genetic diversity calculated from fingerprinting data, plant material can be classified into different genetic pools, which is extremely useful for the identification of the most appropriate parental lines to be crossed (Ribaut and Hoisington 1998). It has been proposed that, for crop species in which heterosis is exploited through the production of hybrid cultivars, DNA marker -based genetic distance can be useful for the prediction of yields of crosses between lines from the same germplasm group (Melchinger *et al* 1992). When combined with the phenotypic data, DNA marker- associated effects evaluated in early generation test crosses can be used efficiently to predict later generation test cross performance (Eathington 1997). However, there are a number of limitations on the utility of the DNA markers in prediction of complex QTLs like yield because epistasis seems to play a major role in the expression of these traits (Yu *et al* 1997). The favorable gene combination with less GD may show more heterosis than the unfavorable gene combination with more GD due to epistatic interactions among various alleles of complex QTLs and QTLs for their component traits. Continuous efforts are therefore required to discover the utility of molecular markers in the prediction of hybrid performance and heterosis.

Use of DNA Markers in Hybrid Rice Programs

It is evident that a better understanding of the genetic and molecular basis of heterosis is required to relate genotypic data to heterosis (Charcosset 1992). Zhang *et al* (1995) reported that the relationship between molecular marker heterozygosity and heterosis was variable, depending on the genetic material and the diversity of rice germplasm being used. Xiao *et al* (1996) observed that genetic distance measures based on random markers like RAPD and SSRs were useful for predicting the yield potential and heterosis of intra- subspecies hybrids,

but not inter- subspecies hybrids. Zhang *et al* (1994) reported that correlations based on specific heterozygosity were large for yield heterosis and hence might have practical utility for the prediction of heterosis. Saghai Maroof *et al* (1997) showed that marker heterozygosity could be highly correlated with rough rice yield and head rice yield and was also significantly correlated with heterosis of these two traits. This suggested that molecular markers might be potentially useful for predicting the performance of inter- subspecific hybrids, although more indica-tropical japonica hybrids are needed to confirm this relationship. In general, it is well accepted that the association of marker- based genetic distance with hybrid performance and heterosis depends on the type of crop, the germplasm being studied (Melchinger *et al* 1992) and also on the type of marker being used for the analysis (Parsons *et al* 1997).

Salient Features of My Analysis

Not many previous attempts have been made, where a three-line hybrid system of Indian rice genotypes has been used to correlate the genetic distance derived from DNA markers with hybrid performance and heterosis. Since the chromosomal locations of molecular markers influence the patterns of relationships revealed in diversity studies (Parsons *et al* 1997), we employed both RFLP and STMS markers representing the specific regions linked to traits of interest and ISSR markers spanning the genome randomly. The ISSR markers were not quite informative, in terms of performance prediction, indicating that these random markers are not of much predictive value for hybrid performance and heterosis. Some previous reports (Glaszmann 1987, Wang and Tanksley 1989 and Yu and Nguyen 1994) have indicated that the intra-species relationships derived using isozymes, RAPD and RFLP data are all similar. However, my results do not show much similarity in the clustering of the genotypes based on random or specifically linked RFLP/ STMS markers. This can be attributed to the fact that we used only those RFLP/ STMS markers that are linked to QTLs contributing to yield, spanning specific regions on different chromosomes and not covering the entire genome, as has been done by previous workers.

Need to Identify More Loci Playing Role in Heterosis

The use of specific heterozygosity markers produced positive correlation in a number of traits especially for hybrid performance. This was true when the markers for all the traits were considered together rather than the individual trait, except for grain yield. This may be due to the role of epistasis in the genetic basis of heterosis as proposed by Yu *et al* (1997) and as explained earlier in this chapter. When analyzed individually, the grain yield showed extremely positive and significant correlations in the CMS IR 62829A and restore crosses. However, the lack of significant correlations for grain yield in CMS IR 58025A and restorer crosses is difficult to explain as the banding patterns produced by markers linked to this trait were exactly identical for both CMS IR 58025A and CMS IR 62829A (observe the patterns in Lane No. 1 and 2 of Figure 2). It thus appears that there is a need to identify more loci, in this particular germplasm that may be playing a role in the heterosis phenomenon, which cannot be detected by the markers being analyzed in the present data. More so because the two CMS lines studied share the same cytoplasmic background (WA type), but, their nuclear backgrounds are quite different. Additional evidence to support the view of presence of additional loci contributing to yield lies in the fact that crosses having the same genetic distance showed contrasting values for hybrid performance and for mid- parent heterosis and heterbeltiosis (Figure 6-13).

Strategies for Meaningful Use of Markers in Breeding

Gimelfarb and Lande (1995) reported that the markers identified to be lying within a locus linked to a particular QTL may not necessarily be most tightly linked to the QTLs controlling the trait. Hence the additive effects associated with the markers estimated by the regression may not accurately reflect the contributions to the trait by the QTLs. Therefore, QTLs identified in a population may not be suitable for analysis of another unrelated population. The germplasm being analyzed in our study is different than the 9024 (indica) and LH422 (japonica) population that was used by Xiao *et al* (1995) for the QTLs used in the present study. Based on the observations in my analysis, It can be concluded that markers spanning the entire

genome or appearing to be linked to specific QTLs in the genome map may not always be useful in the selection of parental lines for hybrid program. It is therefore necessary to validate the markers identified from one population into few more populations to eliminate genotype effect on the traits and to identify markers, which are not affected by such genotype effects. Thus an in depth analysis of genotype and environmental interactions needs to be undertaken, for meaningful use of molecular markers in marker assisted breeding program.

Thesis Overview

Food Security: A Challenging Task

The world population is expected to grow from 6 billion to 8 billion in the next 20 years, with much of the increase coming in rice dependent nations. To feed the growing population, the grain production needs to be increased by 40 % according to the statistical data, from International Rice Research Institute. Though the green revolution has brought about a significant self-sufficiency in the production of food grains, food security remains an alarming concern due to difficulties in facing the challenge of meeting the ever-increasing demand for food. Since conventional breeding practices alone cannot be enough to meet the challenge of increasing food production, efforts are being made all over the world to integrate modern molecular approaches with breeding. Genetic engineering and DNA marker technology are two well-established approaches for increasing quality and productivity of crop plants. The DNA marker based strategies in particular involve the use of different types of DNA markers with various applications in relation to plant breeding as described in chapter I of this thesis. One more important feature of this approach is that no ethical issues are raised when it is used for breeding in crop improvement.

Exploitation of DNA Markers in Rice Genome Analysis

In my thesis, I have exploited the potential of DNA markers to study the variation and phylogenetic organization of specific rice germplasm and in prediction of hybrid performance, in order to assist selection of parents in a three- line hybrid system. The Specific Highlights can be summarized as Follows:

Genetic Diversity and Phylogenetic Analysis of the Genus *Oryza* Based on ISSR Markers.

This analysis has revealed the potential of these markers in understanding the evolutionary relationships of *Oryza* and to produce diagnostic fingerprints for

various species, genomes and also for individual accessions / cultivars. Nearly 87 putative genome / species specific molecular markers have been identified for 8 of the 9 genomes of *Oryza*.

Detection of Variable Alleles of $(CAC)_n$ Containing Genes in *Oryza*

This information can be very useful for finding new alleles especially of genes like cyclophilin because these genes play important role in the life cycle of rice plants like development, stress response and other house keeping functions, thus providing target gene sources for future introgression breeding programs for crop improvement.

Prediction of Heterotic Combination in Three Line Hybrid System

My data indicates that ISSR primers are useful in producing fingerprints and putative variety specific bands. While, specific markers could cluster the parental lines in different groups and show significant correlation with hybrid performance, the data also supports the proposition that epistasis is the basis of heterosis.

Molecular Green Revolution: A New Era in Crop Improvement

In this dissertation, I have attempted to explain about the approaches used by me for exploiting the potential of DNA markers in rice genome analysis. It is true that several workers have explored such approach over a decade and yet real success stories are awaited. However, the tremendous flow of information involving use of DNA markers clearly suggests their vast potential in producing plants with superior traits. Efforts in this direction are being made all over the world and molecular green revolution remains a distinct future for plant breeding in years to come.

During the course of this work, a number of observations were made which prompt us to look upon the results discussed here as the stepping stone and in future focus on certain aspects that come to our mind. For example,

1] How can we utilize the putative genome specific markers generated in our work? It would be definitely interesting to convert these specific bands into SCAR markers and use them for actual genotyping, cultivar identification and to detect micro levels of introgression in back cross breeding programs.

2] In sectionII of chapter2, we have generated and characterized ESTs that are homologous to cyclophilin genes. Owing to the importance of cyclophilins, an exercise can be carried out to detect if variable allele introgression can cause functional effect to alter response. Back cross breeding programs between wild and cultivated rice are being carried out for past several years. Such populations can possibly be used for initial analysis before undertaking a proper program with focus on these genes.

3] Lastly utility of molecular markers in prediction of hybrid performance will probably become more efficient if specific markers are identified for specific environmental conditions and also from related germplasm population. It is therefore necessary at least for QTLs to study the contribution of each QTL under individual environmental conditions and design programs for region specific variety development and development of variety for all the environmental conditions with base level phenotypic contribution. This type of study can also facilitate the basic understanding of how markers can be effectively utilized in hybrid programs, especially for traits governed by QTLs. Despite a decade of research in the area of molecular biology, there is still dearth of such information and this may be one reason why breeders are unable to carry out marker assisted selection of parents for commercial use.

References

Adams M D, Kelley J M, Gocayne J D, Dubrick M, Polymeropoulos M H, Xiao H, Merril C R, Wu A, Olde B, Moreno R F, Kerlavage A R, McCombie W R, Venter J C (1991) Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science.* 252: 1651-1656

Aggarwal R K, Brar D S, Khush G S (1997) Two new genomes in the *oryza* complex identified on the basis of molecular divergence analysis using total genomic DNA hydridization. *Mol Gen Genet.* 254: 1-12

Aggarwal R K, Brar D S, Nandi S, Huang N, Khush G S (1999) Phylogenetic relationships among *Oryza* species revealed by AFLP markers. *Theor Appl Genet.* 98: 1220-1228

Ainsworth P J, Surh L C, Coulter-Mackie M B (1991) Diagnostic single strand conformational polymorphism (SSCP): a simplified non-radio isotopic method as applied to a Tay-sachs B1 variant. *Nucl Acids Res*. 19: 405-406

Akagi H, Yokozeki Y, Inagaki A, Nakamura A, Fujimura T (1996) A co-dominant DNA marker closely linked to the rice nuclear restorer gene RF-1 identified with inter SSR fingerprinting. *Genome.* 39: 61205-1209

Alpert K B, Tanksley S D (1996) High resolution mapping and isolation of a yeast artificial chromosome contig containing *fw2*.2: A major fruit weight quantitative trait locus in tomato. *Proc Natl Acad Sci USA*. 93: 15503-15507

Arcade A, Faivre-Rampant P, Le Guerroue B, Paques L E, Prat D (1996) Heterozygosity and hybrid performance in Larch. *Theor Appl Genet.* 93: 1274-1281

Barbosa-Neto J F, Sorrells M E, Cisar G (1996) Prediction of heterosis in wheat using coefficient of parentage and RFLP based estimates of genetic relationship. *Genome.* 39: 1142-1149

Bernardo R (1992) Relationship between single- crosses performance and molecular marker heterozygosity. *Theor Appl Genet*. 83: 628-634.

Blair M W, Panaud O, McCouch S R (1999) Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (*Oryza sativa* L.). *Theor Appl Genet*. 98: 780-792.

Bligh H FJ, Blackhall N W, Edwards K J, McClung A M (1999) Using amplified fragment length polymorphisms and simple sequence length polymorphisms to identify cultivars of brown and white milled rice. *Crop Sci.* 39: 1715-1721

Boppenmaier J, Melchinger A E, Brunklaus-Jung E, Geiger H H, Herrmann R G. (1993) Genetic distance for RFLPs in European maize inbreds. III, Performance of crosses within versus between heterotic groups for grain traits. *Plant Breeding.* 111: 217-226.

Boppenmaier J, Melchinger A E, Seitz G, Geiger H H, Herrmann R G (1992) Genetic diversity for RFLPs in European maize inbreds: I. relation to Performance of Flint X Dent crosses for forage traits. *Crop Sci.* 32: 895-902.

Botstein B, White R L, Skolnick M and Davis R W (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet.* 32 314-331

Brar D S, Khush G S (1997) Alien introgression in rice. *Plant Mol Biol.* 35: 35-47

Broun P Tanksley S D (1993) Characterization of tomato DNA clones with sequence similarity to human minisatellites 33.6 and 33.15. *Plant Mol Biol*. 23: 231-242

Brown S M, Kresovich S (1996) in Genome Mapping in Plants. (ed Paterson AH), Clandes New York. PP 85-93

Buchholz W G, Harris- Haller L, Derose R T, Hall T C, (1994) Cyclophilins are encoded by a small gene family in rice. *Plant Mol biol.* 25: 837-843

Burnboim H C, Dolly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acid Research*. 7: 1513

Bustos A, Soler C, Jouve N (1999) Analysis of PCR-based markers using: between species of *Hordeum* (*Poaceae*). *Genome.* 42 (1): 129-138

Caetano-Anolle's G, Bassam B J, Gresshoff P M (1991) DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Bio/Technology.* 9: 553-557

Callow J A (1997) Molecular markers and the management of genetic resources in seed gene Banks: a case study of rice. Biotechnology and Plant Genetic Resources Conservation and Use. Callow J.A. Newbury H.J.(eds). PP 103-118

Cerna F J, Cianzio S R, Rafalski A, Tingey S, Dyer D (1997) Relationship between seed yield heterosis and molecular marker heterozygosity in Soybean. *Theor Appl Genet*. 95: 460-467

Chang C, Bowman A W, Lander E S, Meyerowitz E W (1988) Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana. Proc Natl Acad Sci USA.* 85: 6856-6860

Charcosset (1992) Proc XIII EUCARPIA cong. Springer, Berlin Heidelberg NewYork. 355-369

Chen X, Temnykh S, Xu Y, Cho Y G, McCouch S R (1997) Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). *Theor Appl Genet.* 95: 553-567

Cho Y G, Blair M W, Panaud O, McCouch S R (1996) Cloning and mapping of variety specific rice genomic DNA sequences: Amplified fragment length polymorphisms (AFLP) from silver stained poly acryl amide gels. *Genome.* 39: 373-378

Cho Y G, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch S R, Park W D, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and Gene Bank sequences in rice (*Oryza sativa L.*). *Theor Appl Genet*. 100: 713-722.

Chou I T, Gasser C S (1997) Characterization of the cyclophilin gene family of *Arabidopsis thaliana* and phylogenetic analysis of known cyclophilin proteins. *Mol Biol.* 35: 873-892

Chowdari K V, Venkatachalam S R, Davierwala A P, Gupta V S, Ranjekar P K, Govila O P (1998) Hybrid performance and genetic distance as revealed by $(GATA)_4$ microsatellite and RAPD markers in Pearlmillet. *Theor Appl Genet.* 97: 163-169

Cordesse F, Second G, Delsney M (1990) Ribosomal gene spacer length variability in cultivated and wild rice species. *Theor Appl Genet.* 79: 81-88

Cox T S, Murphy J P (1990) The effect of parental divergence on F2 heterosis in winter wheat crosses. *Theor Appl Genet*. 79: 241-250

Dallas J F, McIntyre C L, Gustafson J P (1993) An RFLP species-specific DNA sequence for the A genome of rice. *Genome.* 36: 445-448

Dally A M, Second G (1990) Chloroplast DNA diversity in wild and cultivated species of rice genus *Oryza*, section *Oryza*. Cladistic- mutation and genetic distance analysis. *Theor Appl Genet*. 80: 209-222

Davierwala A P, Chowdari K V, Shiv Kumar, Reddy A P K, Ranjekar P K and Gupta V S (2000) Use of three different marker systems to estimate genetic diversity of Indian elite rice varieties. Genetica. (in press)

Dellaporta S L Wood J Hick J B (1983) A plant DNA minipreparation: version II. *Plant Mol Biol Rep.* 1:19-21

Deshpande A D Ramakrishna W Mulay G P Gupta V S Ranjekar P K (1998) Evolutionary and polymorphic organization of *Knotted-1* homeobox in cereals. *Theor Appl Genet*. 97: 135-140

Deng H, Wang G (1984) A study on prediction of heterosis in crops (in chinese). *Hunan Agric Sci.* 3: 1-5

Diers B W, McKetty P B E, Osborn T C (1996) Relationship between heterosis and genetic distance based on restriction fragment length polymorphism markers in oilseed rape (*Brassica napus* L). *Crop Sci.* 36: 79-83

Dong N V, Subudhi P K, Luong P N, Quang V D, Quy T D, Zheng H G, Wang B, Nguyen H T (2000) Molecular mapping of a rice gene conditioning thermosensitive genic male sterility using AFLP, RFLP and SSR techniques. *Theor Appl Genet.* 100: 727-734

Dorweiler J, Stec A, Kermicle J, Doebley J (1993) Teosinte glume architecture- 1: A genetic locus controlling a key step in maize evolution. *Science.* 262: 233

Dudley J W, Saghai Maroof M A, Rufener G K (1991) Molecular markers and grouping of parents in maize breeding programs. *Crop Science.* 31: 718-723

Dudley J W (1993) Molecular markers in plant improvement: Manipulation of genes affecting quantitative traits. *Crop Science.* 33: 660-668

Eathington S R (1997) Usefulness of marker - QTL association in early generation selection. *Crop Sci.* 37: 1686-1693

Ehrlich H A, Gelfand D H, Sninsky J J (1991) Recent advances in the polymerase chain reaction. *Science.* 252 1643-1651

Epplen J T (1992) Advances in electrophoresis. Inchrambach A, Dunn N J and Radola B J (eds), VCH Cambridge. Vol 5. PP 59-112

Falconer D S (1981) Introduction to quantitative genetics. 2nd Ed. Longman. London.

Flavell R B (1995) Plant biotechnology R&D – The next ten years. *Tibtech.* 13: 313-319

Fukuchi A, Kikuchi F, Hirochika H (1993) DNA fingerprinting of cultivated rice with rice retrotransposon probes. *Jpn.J.Genet*. 68: 195-204.

Fukuoka S, Inoue T, Miyao A, Monna L, Zhong M S, Sasaki T, Minobe Y (1994) Mapping of sequence-tagged sites in rice by single strand conformation polymorphism. *DNA Res.* 1: 271-277

Gardiner J M, Coe E H, Melia-Hancock S, Hoisington D A, Chao S (1993) Development of a core RFLP map in maize using an immortalized F2 population. *Genetics.* 134: 917-930

Ge S, Sang T, Bao-Rong L, De-Yuan Hong (1999) Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc Natal Acad Sci USA*. 96: 25: 14400-14405

Gimelfarb A, Lande R (1995) Marker assisted selection and marker- QTL associations in hybrid populations. *Theor Appl Genet.* 91: 522-528

Glaszmann J C (1987) Isozyme and classification of Asian cultivated rice varieties. *Theor Appl Genet.* 74: 21-30

Godshalk E B, Lee M, Lankey K R (1990) Relationship of restriction fragment length polymorphism in single- cross hybrid performance of maize. *Theor Appl Genet.* 80: 273-289

Gortner G, Pfenninger M, Khal G, Weising K (1996) Northern blot analysis of simple repetitive sequence transcription in plants. *Electrophoresis.* 17: 1183-1189

Gottlieb L D, Ford V S (1996) Phylogenetic relationships among the sections of Clarkia (onagraceae) inferred from the nucleotide sequences of PgiC. *Syst Bot.* 21: 45-62

Grandillo S, Tanksley S D (1996) Genetic analysis of RFLPs GATA microsatellites and RAPDs in a cross between *L.esculentum* and *L. pimpinellifolium. Theor Appl Genet.* 92: 957-965

Griffings B (1956) Concept of general and specific combining ability in relation to diallel crossing system. *Aust J Biol Sci.* 9: 463-493

Gu W K, Waden N F, Yu J, Wallace D H (1995) Large scale, cost effective screening of PCR products in marker-assisted selection applications. *Theor Appl Genet.* 91: 465-470

Guimaraes E P, Borrero J, Ospina-Rey Y (1996) Genetic diversity of upland rice germplasm distributed in Latin America. *Pesq Agropec Bras Brasilia*. 31: 187-194.

Gupta M, Chyi Y-S, Romero-Severson J, Owen J L (1994a) Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple sequence repeats. *Theor Appl Genet*. 89: 998-1006

Gupta V S, Ramakrishna W, Rawat S R, Ranjekar P K (1994b) $(CAC)_5$ detects DNA fingerprints and sequences homologous to gene transcripts in rice. *Biochemical Genetics.* 32, No. ½ -8

Gustafson JP, Yano M (2000) Genetic mapping of hypervariable minisatellite sequences in rice (*Oryza sativa* L.). *Theor Appl Genet.* 100: 447-453

Gupta P K, Balyan H S, Sharma P C, Ramesh B (1996) Microsatellites in Plants: A new class of molecular markers. *Current Sci.* 70 (1): 45-54

Harlan J R (1987) Plant Genetic Resources : A conservation Imperative. Yeatman C W, Kafton D, Wilkes G (eds). West View, Boulder Co. PP 111-129

Hayes P M, Liu B H, Knapp S J, Chen F, Jones B, Blake T, Franckowiak J, Rasmusson D, Sorrells M, Ullrich S E, Wesenberg D, Kleinhofs A (1993) Quantitative trait locus effects and environmental interaction in a sample of North American barley germplasm. *Theor Appl Genet.* 87:392-401

Hearne C M, Ghosh S, Todd J A (1992) Microsatellites for linkage analysis of genetic traits. *Trends Genet.* 8: 288-294

Heath D D, Iwama G K, Devlin R H (1993) PCR primed with VNTR core sequences yields species specific patterns and hypervariable probes. *Nucl Acids Res.* 21: 5782-5785.

Hedges S B (1992) The number of replications needed for accurate estimation of the bootstrap value in phylogenetic studies. *Mol Biol Evolution* 9: 366-369

Helentjaris T, Slocum M, Wright S, Schaefer A, Nienhuis J (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. *Theor Appl Genet*. 72: 761-769.

Hillel J, Schaap T, Haberfield A, Jeffreys A J, Plotzky Y, Cahaner A, and lavi U (1990) DNA fingerprints applied to gene introgression breeding programs. *Genetics*. 124: 783

Hirabayashi H, Ogawa T (1995) RFLP mapping of *Bph-1* (brown planthopper resistance gene) in rice. *Breed Sci.* 45: 369-371

Hittalmani S, Foolad M R, Mew T, Rodriguez R L, Huang N (1995) Development of a PCR based marker to identify rice blast resistance gene *Pi* (t) in a segregating population. *Theor Appl Genet*. 91: 9-14

Hittalmani S, Parco A, Mew T V, Zeigler R S, Huang N (2000) Fine mapping and DNA marker assisted pyramiding of the three major genes for blast resistance in rice. *Theor Appl Genet*. 7: 1121-1128

Huang H, Kochert G (1994) Comparative RFLP mapping of an allotetraploid wild rice species (*Oryza latifolia*) and cultivated rice (*O.sativa*). *Plant Mol Biol.* 25: 633-648

Ishii T, Tsunewaki K (1991) Chloroplast genome differentiation in Asian cultivated rice. Genome. 34: 818-826

Ishii T, Brar D S, Multani D S, Khush G S (1994) Molecular tagging of genes for brown plant hopper resistance and earliness introgressed from *Oryza australiensis* in to cultivated rice *O.sativa*. *Genome.* 37: 217-221

Ishii T, Brar D S, Second G, Tsunewaki K, Khush G S (1995) Nuclear genome differentiation in Asian cultivated rice as revealed by RFLP analysis. *Jpn J Genet*. 70: 643-652

Ishii T, Nakano T, Maeda H, Kamijima O (1996) Phylogenetic relationships in A-genome species of rice as revealed by RAPD analysis. *Genes Genet Syst*. 71: 195-201

Izawa T, Foster I, Nakajima M, Shimamoto K, Chua N H (1994) The rice bZIP transcriptional activator RITA 1 is highly expressed during seed development. *Plant Cell.* 6: 1277-1287

Jackson M T (1997) Conservation of rice genetic resources: The role of international gene bank at IRRI. *Plant Mol Biol*. 35:61-67

Jarvis P, Lister C, Szabo V, Dean C (1994) Integration of CAP_5 markers in to the RFLP map generated using recombinant inbred lines of *Arabidopsis thaliana.* *Plant Mol Biol*. 24: 685-687

Jena K K, Khush G S, Kochert G (1994) Comparative RFLP mapping of a wild rice *Oryza officinalis* and cultivated rice *Oryza sativa*. *Genome.* 37: 382-389

Jeffery A J, Wilson V, Thein S L (1985) Hypervariable minisatellite regions in human DNA. *Nature.* 314: 67-73

Kim N S, Shim K M, Yoon K M, Shin Y B, Cho Y S, Eun M Y, Cho Y G (1996) Polymorphism and phylogenetic analyses in rice cultivated in Korea by RAPDs and RFLPs. *Korean J. Genetics*. 18: 211-224.

Kellog EA (1998) Who's related to whom? Recent results from molecular systematic studies. *Current Opinion in plant Biology.* 1 (2): 149-158

Kellog E A, Apples R, Mason-Gamer R J (1996) When genes tell different stories: the diploid genera of *Triticeae* (*Gramineae*). Syst Bot. 21: 321-347

Khush G S (1977) Disease and insect resistance in rice. *Adv Agron.* 29: 265-361

Khush G S, Toenniessen G H (1991) Rice Biotechnology: *Biotechnology in Agriculture No. 6*. CAB International .PP 55-81

Ko H L, Cowan D C, Henry R J, Graham G C, Blakeney A B, Lewin L G (1994) Random amplified polymoprhic DNA analysis of Australian rice (*Oryza sativa* L.) varieties. *Euphytica*. 80: 179-189.

Koniecyzn A, Ausubel F M (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype specific PCR based markers. *Plant J.* 4: 403-410

Kumar L S (1999) DNA markers in plant improvement: An overview. *Biotechnology Advances*. 17: 143-182

Lampe K L (1991) Foreword: Rice Biotechnology. *Biotechnology in Agriculture No.6*. IRRI, C A B International Publication. PP vii.

Landry B S, Michelmore R W (1987) Methods and applications of restriction fragment length polymorphism analysis to plants in tailoring genes for crop improvement. An *Agricultural Perspective*. G Bruening, J Harada and A Hollaender (eds) (New York: Plenum) PP: 134-157

Lang NT, Subudhi PK, Virmani SS, Brar DS, Khush GS, Li Z, Huang N (1999) Development of PCR-based markers for thermosensitive genetic male sterility gene *tms*3(t) in rice. *Hereditas.* 131: 121-127

Lee M (1995) DNA markers and plant breeding programs. *Advances in Agronomy*. 55: 265-344

Lee M, Godshalk E B, Lamkey K R, Woodman W W (1989) Association of restriction fragment length polymorphism among maize inbreds with agronomic performance of their crosses. *Crop Sci.* 29: 1067-1071

Li Z B, Xiao Y H, Zhu Y H, Li R Q, Liu C L, Wang J M (1982) The study and practice of hybrid rice (in Chinese). Shanghai academic and technical press, Shanghai, China

Litt M, Luty J A (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotid repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet*. 44: 397-401

Lorieux M, Nadjiondjop M N, Ghesquiere A (2000) A first interspecific *Oryza sativa X Oryza glaberrima* microsatellite based genetic linkage map. *Theor Appl Genet*. 100: :¾ 593-601

Mackill D J, Zhang Z, Redona E D, Colowit P M (1996) Level of polymorphism and genetic mapping of AFLP markers in rice. *Genome* 39: 969-977

Mahalanobis P C (1936) On the generalized distance in statistics. *Proc Nat Inst Sci India.* 2 (1): 49-55

Maheshwaran M, Subudhi P K, Nandi S, Xu J C, Parlo A, Yang D C, Huang N (1997) Polymorphism, distribution and segregatiopn of AFLP markers in a double haploid rice population. *Theor Appl Genet.* 94: 39-45

Malecot G (1948) Les mathematiques de l'heredite. Masson et Cie, Paris.

Marivet J, Frendo P, Burkard G (1992) Effects of abiotic stresses on cyclophilin gene expression in maize and bean and sequence analysis of bean cyclophilin cDNA. *Plant Sci*. 84: 171-178

Marivet J, Margis- Pinheiro M, Frendo P, Burkard G (1994) Bean cyclophilin gene expression during plant development and stress conditions. *Plant Mol Biol.* 26: 1181-1189

Martin G B (1998) Gene discovery for crop improvement. *Current Openion in Biotechnology.* 9:220-226

Martin G B, Williams J G K, Tanksley S D (1991) Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near isogenic lines. *Proc Natl Acad Sci USA.* 88: 2336-2340

Martin J M, Talbert L E, Lanning S P, Blake N K (1995) Hybrid performance in wheat as related to parental diversity. *Crop Science.* 35: 104-108

Mason-Gamer R J, Kellog E A (1996) Potential utility of the nuclear gene *waxy* for plant phylogenetic analysis. *Amer J Bot.* 83:178

Mason-Gamer R J, Weil C F, Kellogg E A (1998) Granule bound starch synthase: structure function and phylogenetic utility. *Mol Biol Evol.* 15 (12): 1658-1673

Mathew S, Sharrock R A, Lavin M (1995) Evolution of phytochrome gene family and its utility for phylogenetic analyses of angiosperms. *Ann Missouri Bot Gard.* 82: 296–321

Mathew S, Sharrock R A (1996) The phytochrome gene family in grasses (Poaceae): a phylogeny and evidence that grasses have a subset of the loci found in dicot angiosperms. *Mol Biol Evol.* 13: 1141- 1150

McCouch S R, Kochert G, Yu Z H, Wang Z Y, Khush G S, Coffman W R and Tanksley S D (1988) Molecular mapping of rice chromosomes. *Theor Appl Genet*. 76: 815-829

McCouch S R, Chen X, Panaud O, Temnykh S, Xu Y, Cho Y G, Huang N, Ishii T, Blair M (1997) Microsatellite marker development, mapping and applications in rice genetics and breeding. *Plant Mol Biol.* 35: 89-99

Melchinger A E, Lee M, Lamkey K R, Woodman W W (1990a) Genetic diversity for restriction fragment length polymorphisms, relation to estimated genetic effects in maize inbreds. *Crop Sci.* 30: 1033-1040

Melchinger A E, Lee M, Lamkey K R, Hallaue A R, Woodman W W (1990b) Genetic diversity for restriction fragment length polymorphisms and heterosis for two diallel sets of maize inbreds. *Theor Appl Genet.* 80: 488-496

Melchinger A E, Boppenmaier J, Dhillon B S, Pollmer W G, Hermann R G (1992) Genetic diversity for RFLPs in European maize inbreds: II. Relation to performance of hybrids within versus between heterotic groups for forage traits. *Theor Appl Genet.* 84: 672-681

Michelmore R W, Paran I, Kesseli R V (1991) Identification of markers linked to disease resistance genes by bulked segregant analysis a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci USA.* 88: 9828-9832

Miller J C, Tanksley S D (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. *Theor Appl Genet*. 80: 437-448

Miyao A, Zhong HS, Monna L, Yano M, Yamamoto K, Havukkala I, Minobe Y, Sasaki T (1996) Characterization and genetic mapping of simple sequence repeats in the rice genome. *DNA Res.* 3: 233-238

Mohan M, Nair S, Bhagwat A, Krishna T G, Yano M, Bhatia C R and Sasaki T (1997a) Genome mapping, molecular markers and marker assisted selection in crop plants. *Mol Breed*. 3 (2): 87-103

Mohan M, Sathyanarayanan P V, Kumar A, Srivastava M N, Nair S (1997b) Molecular mapping of a resistance specific PCR based marker linked to a gall midge resistance gene (*GM*4t) in rice. *Theor Appl Genet.* 95: 777-782

Mohan M, Nair S, Bentur J S, Prasad Rao U, Bennett J (1994) RFLP and RAPD mapping of rice *Gm2* gene that confers resistance to biotype 1 of gall midge (*Orseolia oryzae*). *Theor Appl Genet.* 87: 782-788

Moll R H, Lonnquist J H, Velez Fortuno J, Johnson E C (1965) The relationship of heterosis and genetic divergence in maize. *Genetics.* 52: 139-144

Morinaga T (1964) Cytogenetical investigation on *Oryza* species in: IRRI (ed.) Rice genetics and cytogenetics, Elsevier Amsterdam. PP. 91-103

Nair S, Bentur J S, Prasad Rao U., Mohan M. (1995) DNA markers tightly linked to a gall midge resistance gene (*Gm*2) are potentially useful for marker aided selection in rice breeding. *Theor. Appl. Genet.* 91: 68-73

Nakamura Y, Julier C, Wolff R, Holm T, O'Connell P, Leppert M, White R (1987) Characterization of a human 'midisatellite' sequence. *Nucl Acids Res*. 15: 2537-2547

Nakase M, Aoki N, Matsuda T, Adachi T (1997) Characterization of a novel bZIP protein which binds to the α-globulin promoter. *Plant Mol Biol.* 33: 513-522

Nam H G, Giraudat J, Den Boer B, Moonan F, Loos W D B, Hauge B M and Goodman H M (1989) Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana*. *Plant Cell.* 1: 699-705

Naqvi N I, Chattoo B B (1996) Development of a sequence characterized amplified region (SCAR) based indirect selection method for a dominant blast resistance gene in rice. *Genome*. 39: 26-30

Nei M (1987) Molecular evolutionary genetics. Columbia University Press. New York.

Nei M, Li W H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA.* 76: 5269-5273

Nelson R J, Yap V I (1996) Win Boot: A program for performing boot strap analysis of binary data to determine the confidence limits of UPGMA based dendrograms. *IRRI Discussion paper series* No. 14 International Rice Research Institute, PO Box 933, Manila, Philippines

Oba S, Hanzawa M, Anai T, Umeda M, Uchimiya H (1996) Novel DNA probes capable of discriminating indica and japonica rice cultivars. *DNA Seq* 6: 303-306

Oka H I, Morishima H (1982) Phylogenetic differentiation of cultivated rice, 23 Potentiality of wild progenitors to evolve the indica and joponica types of rice cultivars. *Euphytica*. 31: 41-50

Oka H I (1988) Origin of cultivated rice Japan Scientific Societies Press, Tokyo, Elsevier, Amsterdam Oxford, New York, Tokyo

Oka H I (1991) Genetic diversity of wild and cultivated rice Rice Biotechnology Eds. Khush G S Toenniessen G H *Biotechnology in Agriculture No. 6* PP: 55-81

Olson M, Hood L, Cantor C, Botstein D (1989) A common language for phycial mapping of the human genome. *Science.* 245: 1434-1435

Orita M, Suzuki Y, Sekiya T, Hayashi K (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymarase chain reaction. *Genomics.* 5: 874-877

Panaud O, Chen X, McCouch S R (1995) Frequench of microsatellite sequences in rice (*Oryza sativa* L). *Genome.* 38: 1170-1176

Panaud O, Chen X, McCouch S R (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). *Mol Gen Genet*. 252: 597 - 607

Parsons B J, Newbary H J, Jackson M T, Ford-Loyd B V (1997) Contrasting genetic diversity relationships are revealed in rice (Oryza sativa L.) using different marker types. *Mol Breeding*. 3: 115-125

Paterson A H, Tanksley S D, Sorrells M E (1991) DNA markers in plant improvement. *Advances in Agronomy*. 46: 39-90

Paterson A H, Deverna J W, Lanini B, Tanksley S D (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in interspecies cross of tomato. *Genetics.* 124: 735-742

Paterson A H (1995) Molecular dissection of Quantitative traits: Progress and prospects. *Genome Research.* 5: 321-333

Peng J Y, Glaszmann J C, Virmani S S (1988) Heterosis and isozyme diversions in Indica rice. *Crop Sci*. 28: 561-563

Peng J Y, Virmani S S, Julfiquar A W (1991) Relationship between heterosis and genetic divergence in rice. *Oryza*. 28: 129-133

Powell W, Machray G C, Provan J (1996) Polymorphism revealed by simple sequence repeats. *Trends Plant Sci*. 1: 215-222

Qian H R, Zhuang J Y, Lin H X, Lu J, Zheng K L (1995) Identification of a set of RFLP probes for subspecies differentiation in *Oryza sativa* L. *Theor Appl Genet*. 90: 878-884

Rafalski J A, Tingey S V (1993) Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. *Trends Genet*. 9: 275-280

Ramakrishna W, Lagu M D, Gupta V S, Ranjekar P K (1994) DNA fingerprinting in rice using oligonucleotide probes specific for simple repetitive DNA sequences. *Theor Appl Genet*. 88: 402-406

Ramakrishna W, Chowdari K V, Lagu M D, Gupta V S, Ranjekar P K (1995) DNA fingerprinting to detect genetic variation in rice using hypervariable DNA sequences. *Theor Appl Genet*. 90: 1000-1006

Ramakrishna W Davierwala A P Gupta V S and Ranjekar P K (1998) Expansion of a (GA) dinucleotide at a microsatellite locus associated with domestication in rice. *Biochem Genet.* 36: 9/10 323-327

Reddy O U K, Siddiq E A, Sarma N P, Ali J, Hussain A J, Nimmakayala P, Ramasamy P, Pammi S, Reddy A S (2000) Genetic analysis of temperature-sensitive male sterility in rice. *Theor Appl Genet.* 100: 794-801

Rhoads D D, Roofa D J (1989) SEQUID II Kansas State University Molecular Genetics Laboratory, Manhatten

Richardson T, Cato S, Ramser J, Kahl G, Weising K (1995) Hybridization of microsatellite to RAPD: a new source of polymorphic marker. *Nucleic Acids Research* 23 (18): 3798-3799

Ribaut J-M, Hoisington D (1998) Marker assisted selection: new tools and strategies. *Trends in Plant Sciences.* 3 (6): 236-239

Rogers S O, Bendich A J (1988) Extraction of DNA from plant tissues. Plant Molecular Biology Manual A6: 1 Ed. Gelvin S B and Schilperoort R A (eds), Kulwer Academic Publishers, Dordrecht.

Rongwen J, Akkaya M S, Bhagwat A A, Lavi U, Cregan P B (1995) The use of microsatellite DNA markers for soybean genotype identification. *Theor Appl Genet.* 90: 43-48

Saghai Maroof M A, Yang G P, Zhang Q and Gravois K A (1997) Correlation between molecular marker distance and hybrid performance in U S southern long grain rice. *Crop Science.* 37: 145-150

Saiki R K, Gelfand D H, Stoffel S, Scharf S J, Higuchi R, Horn G T, Mullis K B, Erlich H A (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Nature.* 239: 487-497

Saiki R K, Scharf S, Faloona F, Mullis K B, Horn G T, Erlich H A, Arnheim N (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science.* 230: 1350-1354

Saito T, Niwa Y, Ashida H, Tanaka K, Kawanukai M, Matsuda H, Nakagawa T (1999) Expression of a gene for cyclophilin which contains an amino terminal endoplasmic

reticulum targeting signal. *Plant Cell Physiol.* 40 (1): 77-87

Saito H, Saito K, Nakamura S, Kawasaki S, Iwasaki M (2000) Fine physical mapping of the rice stripe resistance gene locus *Stvb-iy*. *Theor Appl Genet.* 101 (½): 59 -63

Sant V J, Patankar A G, Gupta V S, Sarode ND, Mhase L B, Sainani M N, Deshmukh R B, Ranjekar P K (1999) Potential of DNA markers in detecting divergence and in analyzing heterosis in Indian chickpea cultivars. *Theor Appl Genet.* 98: 1217- 1225

Sano Y, Morishima H, Oka H I (1980) Intermidiate perennial- annual populations of *Oryza.perennis* found in Thailand and their evolutionary significance. *Botany Magazine Tokyo*. 93: 291-305

Sano Y, Sano R (1990) Variation of the intergenic spacer region of ribosomal DNA in cultivated and wild rice species. *Genome.* 33: 209-218

Schlotterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. *Nucl Acids Res.* 20: 211-215

Second G (1982) Origin of the genetic diversity of cultivated rice (*Oryza* sp.) Study of the polymorphism scored at 40 isozyme loci. *Jpn. J. Genet*. 57: 25-57.

Second G, Wang Z Y (1992) Mitochondrial DNA RFLP in genus *Oryza* and cultivated rice. *Theor Appl Genet.* 39: 125-140

Sertedaki A, Lindsay S (1996) CAC- the neglected repeat. *Bioessay.* 18 (3): 237-242

Shastry S V S, Sharma S D (1965) Taxonomic studies in genus *Oryza L.*III. *O.rufipogon* Griff Sensor strictr *O nivara* sharma et shastry nom. *Indian J Genet Plant Breed*. 25: 157-161

Shull G H (1914) Duplicate genes for capsule form in *Bursa bursa- pastoris*. Zeitschr. Induki. *Abstamn. U. Verebungsl*. 12: 97-149

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science.* 236: 787-792.

Smith O S, Smith J S C, Bowen S L, Tenborg R A, Wall S J (1990) Similarities among a group of elite maize inbreds as measured by pedigree, F1 grain yield, grain yield heterosis and RFLPs. *Theor Appl Genet.* 80: 833-840

Sokal R R, Michener C D (1958) A statistic method for evaluating systematic relationships. *Univ Kans Sci Bull.* 28: 1409-1438

Song W, Wang G, Chen L (1995) A receptor kinase like protein encoded by the rice disease resistance gene *Xa*21. *Science.* 270: 1804-1806.

Somers D J, Zhou Z, Bebeli P, Gustafson J P (1996) Repetitive genome specific probes in wheat *Triticum aestivum* I emthell amplified with minisatellite core sequences. *Theor Appl Genet*. 93: 983-989

Southern E M (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol*. 98: 503-517

Stuber C W, Lincoln S E, Wolff D W, Helentjaris T, Lander E S (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics.* 132: 823-839

Swartz D (1989) Users manual, Information resource group, Maryland Biotechnology Institute, University of Maryland, College park, MD 20742 USA.

Tanksley S D, Young N D, Paterson A H, Bonierbale M W (1989) RFLP mapping in plant breeding : New tools for an old science. *Bio/Technology*. 7: 257-264

Tanksley S D, Ganal M W, Martin G B (1995) Chromosome Landing: A paradigm for map based gene cloning in plants with large genomes. *Trends Genets.* 11: 63-68

Tanksley S D (1993) Mapping polygenes. *Annu Review Genet*. 27: 205-233

Thanh N D, Zheng, Dong N V, Trinh L N, Ali M L, Nguyen H T (1999) Genetic variation in root morphology and microsatellite DNA loci in upland rice (*Oryza sativa* L.) from Vietnam. *Euphytica.* 105: 43-51.

Tateoka (1962) Taxonomic studies of Oryza II several species complex. *Bot Mag Tokyo* 75: 165-173

Temnykh S, Park W D, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho Y G, Ishii T, McCouch S R (2000) Mapping and genomic organization of microsatellite sequences in rice (*Oryza sativa L.*). *Theor Appl Genet*. 100: 697-712.

Ting Y (1957) The origin and evolution of cultivated rice in China (in Chienese). *Acta Agron Sinica.* 8 (3): 243-260

Tingey S V, Rafalski J A, Williams J G K (1993) Genetic analysis with RAPD markers, in *Application of RAPD Technology to Plant Breeding* ed M Neff (St. Paul, MN: ASHS Publishers) PP 3-8

Toenniessen G H, Khush G S (1991) Prospects for the future. Rice Biotechnology: *Biotechnology in Agriculture No 6: CAB International and IRRI*: PP.309-313

Udupa SM, Robertson LD, Weigand F, Baum M, Kahl G (1999) Allelic variation at $(TAA)_n$ microsatellite loci in a world collection of chickpea (*Cicer arietinum* L.) germplasm. *Mol Gen Genet.* 261: 354-363

Ulstrop A J (1978) Agric. Hand Book, 199: 1

Vaughan D (1989) The genus *Oryza L*. current status of taxonomy. *IRRI Research paper Series.* Number 138: 21.

Virk P S, Ford-Lloyd B V, Jackson M T, Newbury H J (1995) Use of RAPD for the study of diversity within plant germplasm collections. *Heredity*. 74: 170-179.

Virmani S S (1994) Prospects of hybrid rice in the tropics and subtropics. In " Hybrid Rice Technology: New developments and future prospects" (Virmani Ed. IRRI, Manila Philippines). PP. 7-19

Wang X K, Cheng K S, Lu Y X, Luo J, Huang N W, Liu G R (1984) Studies on indigenous rice in Yunnan and their utilization, III. Glabrous rice in Yunnan. Acta Univ Agric Pekinensis 10: 333-344

Wang ZY, Tanksley SD (1989) Restriction fragment length polymorphism in *Oryza sativa* L. *Genome*. 32: 1113-1118

Wang Z M, Second G, Tanksley S D (1992) Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. *Theor Appl Genet*. 83: 565-581.

Wang G, Mackill D J, Bonman J M, McCouch S R, Champoux M C, Nelson R J (1994) RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. *Genetics*. 136: 1421-1434

Wang B, Zheng J, Xu W, Wu W, Zheng H, Nguyen H T (1995) Tagging and mapping of rice thermosensitive genic male sterile (TGMS) gene. Plant Genome III, January

Wang B, Fu J, Zheng H, Qiu F, Li C, Xie W, Wang J, Jin D, Yang R (1998) Construction of rice BAC libraries and screening with AFLP markers linked to the TGMS gene. Plant and Animal Genome VI Conference, San Diego, California, USA, January 18-22

Waters E R, Lee G J, Vierling E (1996) Evolution structure and function of small heat shock proteins in plants. *J Exp Bot.* 47: 325- 338

Weber J L, May P E (1989) Abundant Class of human DNA polymorphism which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet*. 44: 388-396

Weising K, Nybom H, Wolff K and Meyer W (1995) *DNA fingerprinting in plants and fungi*. A Arbor (eds) CRC Press Inc. Boca Raton PP: 1-3

Weising K, Winter P, Huttel B, Kahl G (1998) Microsatellite markers for molecular breeding. *J. Crop prod*. 1: 113-143

Welsh J and McCelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res*. 18: 7213-7218

Welsh J, Hoercutt R J, McClelland M and Sobral B W S. (1991) Parentage determination in maize hybrids using the arbitrary primed polymerase chain reaction (AP- PCR). *Theor Appl Genet*. 82: 473-476

Williams J G K, Kubelik A R, Livak K J, Rafalski J A, Tingey S V (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl Acids Res.* 18: 6531-6535

Winter P, Kahl G (1995) Molecular marker technologies for plant improvement. *World Journal of Microbiology and Biotechnology.* 11: 438-448

Wu K S, Tanksley S D (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. *Mol Gen Genet*. 241: 225-235

Xiao J, Li J, Yuan L, Tanksley S D (1995) Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. *Genetics* 140: 745-754

Xiao O J, Grandillo S, Ahn S N, McCouch S R, Tanksley S D (1996a) Genes from wild rice improve yield. *Nature*. 384: 223-224

Xiao J, Li J, Yuan L, McCouch S R, Tanksley S D (1996b) Genetic diversity and its relationships to hybrid performance and heterosis in rice as revealed by PCR- based markers. *Theor Appl Genet*. 92: 637-643

Yanagisawa T, Hayashi M, Hirai A, Harada K (1994) DNA fingerprinting in soybean (Glycine Max (L.) Meurill) with oligonucleotide probe for simple repetitive sequences. *Euphytica.* 80: 129-136

Yang G P, Saghai Maroof M A, Xu C G, Zhang Q, Biyashev R M (1994) Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. *Mol Gen Genet*. 245: 187-194

Ye X, Al-Babili S, Kloti A, Zhang J, Lucca P, Beyer P and Potrykus I (2000) Engineering the provitamin A (b-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science.* 287: 303-305.

Yu S B, Li J X, Xu C G, Tan Y F, Gao Y J, Li X H, Zhang Q, Saghai Maroof M A (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid*. Proc Natl Acad Sci USA.* 94: 9226-9231

Yu L X, Nguyen H T, (1994) Genetic variation detected with RAPD markers among upland and lowland rice cultivars (*Oryza sativa* L.). *Theor Appl Genet.* 87: 668-672.

Yu Y G, Saghai Maroof M A, Buss G R, Maughan P J, Tolin S A (1994) RFLP and microsatellite mapping of a gene for soybean mosaic virus resistance. *Phytopathology.* 84: 60-4

Yu Z H, Mackill D J, Bonman J M, Tanksley S D (1991) Tagging genes for blast resistance in rice via linkage to RFLP markers. *Theor Appl Genet.* 81: 471-476

Zabeau M (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting. *European patent Application Publication* No 0534858A1

Zhang Q F, Zhou Z Q, Yang G P, Xu C G, Liu K D, Saghai Maroof M A (1996) Molecular marker heterozygosity and hybrid performance in *Indica* and *Japonica* rice*. Theor Appl Genet.* 93: 1218-1224

Zhang Q F, Gao Y J, Yang S H, Saghai Maroof M A, Li J N (1995) Molecular divergence and hybrid performance in rice. *Mol Breed*. 1: 133-142

Zhang Q, Shen B Z, Dai X K, Mei M H, Saghai Maroof M A, Li Z B (1994a) Using bulked extremes and recessive class to map genes for photoperiod- sensitive genic male sterility in rice. *Proc Natal Acad Sci USA*. 91:8675-8679

Zhang Q F, Gao Y J, Yang S H, Ragab R A, Saghai Maroof M A, Li Z B (1994b) A diliallel analysis of heterosis in elite hybrid rice based on RFLPs and microsatellites. *Theor Appl Genet.* 89: 185-192

Zhang Q F, Saghai Maroof M A, Lu T Y, Shen B Z (1992) Genetic diversity and differentiation of indica and japonica rice detected by RFLP analysis. *Theor Appl Genet.* 83: 495-499.

Zhao X and Kochert G 1992 Characterisation and genetic mapping of short, highly repeated interspersed DNA sequence from rice (*Oryzae sativa* L). *Mol Gen Genet*. 2311: 353-359

Zhao X, Kochert G (1993) Phylogenetic distribution and genetic mapping of a $(GGC)_n$ microsatellite from rice (*Oryza sativa* L.). *Plant Mol Biol*. 21: 607-614

Zhou Z, Gustafson J P (1995) Genetic variation detected by DNA fingerprinting with a rice minisatellite probe in *Oryza sativa* L. *Theor Appl Genet*. 91: 481-488

Zhu J H, Stephenson P, Laurie D A, Li W, Tang D, Gale M D, Jackson M T (1999) Towards rice genome scanning by map-based AFLP fingerprinting. *Mol Gen Genet.* 184-195

Zhuang C X, Zhang G Q, Mei M T, Lu Y G (1999) Molecular mapping of the *S-a* locus for F1 pollen sterility in cultivated rice. *I Chuan Hseuh Pao* 26: 213-218

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics*. 20: 176-183

CURRICULUM VITAE

Swati P Joshi

Plant Molecular Biology Unit
Division of Biochemical Sciences
National Chemical Laboratory
Pune 411008
Maharashtra
India.
Email: spj30@usa.net

EDUCATION

Doctoral research on "Rice Genomics: Use of DNA markers for phylogenetic and hybrid performance analysis." Submitted to the University of Poona, Pune, India, for the degree of Ph.D. in Biochemistry (Molecular Biology). Research was carried out at The National Chemical Laboratory, Pune, India. Research adviser: Dr. Vidya Gupta. From July 1995 to present.

M.Sc. Biochemistry (University of Poona, Pune, India) May 1994 with First Class.

B.Sc. Chemistry (University of Poona, Pune, India) April 1992 First Class with Distinction.

HONORS

- Junior and Senior Research Fellowship Award (Council of Scientific and Industrial Research, India (1995-2000).
- Manolikar Foundation Award for best Science Project and Scientific Essay writing (1991-1992).

PUBLICATIONS

1] S. P. Joshi, P. K. Ranjekar and V. S. Gupta, "Molecular markers in plant genome analysis " (Review article) Current Science, Vol. 77. No. 2; 1999

2] S. P. Joshi, V. S. Gupta, R. K. Aggarwal, P. K. Ranjekar and D. S. Brar, " Genetic diversity and phylogenetic relationship in the genus *Oryza*", Theoretical and Applied Genetics, 2000, Vol. 100, 1311-1320.

3] S. P. Joshi, S. G. Bhave, K. V. Chowdari, G. S. Apte, B. L. Dhonukshe, Lalitha K., P. K. Ranjekar and V. S. Gupta, " Use of DNA markers in prediction of hybrid performance and heterosis for a three line hybrid system in rice (communicated to Biochemical Genetics)

4] S. P. Joshi, M. Katti, K. V. Chowdari, R. Sami Subbu, D. S. Brar, P. K. Ranjekar and V. S. Gupta, " Detection of variable alleles of transcriptionally active microsatellite (CAC) n containing genes in the genus *Oryza*", (submitted to Genome).

Posters presented at International conferences/ symposia

1] Gupta V. S., Ranjekar P. K., Ramakrishna W., Chowdari K. V., Joshi S. P., Wadia A. M., " Characterization of (CAC)n and (GATA)n microsatellites in rice". 4th International DNA fingerprinting conference, 2-7 December 1996, Melbourne, Australia.

2] V. S. Gupta, S. P. Joshi, A. P. Davierwala, and P. K. Ranjekar, "Rice Genomics: Exploitation of DNA markers in hybrid rice program, diagnosis of BLB resistant lines and genotype profiling of wild rice." Meeting of the Rockefeller Foundations International Program on Rice Biotechnology at Phuket, Thailand, September 20 - 24 1999.

Training Courses Attended

- "Emerging trends in DNA based diagnosis" 24th Feb to 7th March 1999 a training course in techniques like, Comparative genomic hybridization (CGH) analysis on DNA arrays, Fluorescence in-situ hybridization on human metaphase chromosomes for detection of malignant diseases, application of chromosome painting,
- Training in Plant tissue culture under the guidance of Dr. N P. Malpathak, Dept. of Botany, University of Poona, Pune, India. (August 94 to May 95)

Technical Skills

- Extensive experience in molecular biology techniques, like DNA isolation, Southern and Northern Blotting, Radioactive and non-radioactive hybridizations, PCR and RT PCR techniques, cloning, in plasmid vectors, genomic library preparation in λ zap, library screening and DNA sequencing.
- Hands on experience with basic techniques in analytical biochemistry and Microbiology.
- Familiarity with plant tissue culture and transformation using *Agrobacterium rhizogenase*.