EXPLOITING THE UTILITY OF MICROSATELLITES IN GENOTYPE PROFILING AND BACTERIAL BLIGHT RESISTANCE DIAGNOSIS IN RICE

A Thesis Submitted to the University of Pune

for the Degree of

DOCTOR OF PHILOSOPHY

IN

CHEMISTRY (BIOCHEMISTRY)

BY

ARMAITY PIROZE DAVIERWALA

PLANT MOLECULAR BIOLOGY UNIT

DIVISION OF BIOCHEMICAL SCIENCES,

NATIONAL CHEMICAL LABORATORY,

PUNE 411008 (INDIA)

JANUARY 2001

DECLARATION

Certified that the work incorporated in this thesis entitled **"Exploiting the Utility of Microsatellites in Genotype Profiling and Bacterial Blight Resistance Diagnosis in Rice"**, submitted by Ms Armaity Piroze Davierwala was carried out by the candidate under my supervision. The material obtained from other sources has been duly acknowledged in this thesis.

(P.K. Ranjekar)

Research Guide

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I consider it a privilege to work under the guidance of Dr P.K. Ranjekar. His continued encouragement and faith in my abilities has helped me to set higher goals for myself and achieve all that I have been able to. He has not only been a guide but a mentor too.

I am extremely grateful to Dr Vidya S. Gupta for her untiring help, guidance, encouragement and keen interest in my work. She has rendered a patient ear to all my problems and has always strived to help me solve them.

I would like to express my sincere thanks to Ramakrishna, Venkat and Anjali for their invaluable help and advice throughout the tenure of my work. I have learnt a lot from my association with them for which I am deeply indebted.

I especially wish to acknowledge the help rendered by Bhushan and Rahul at all times. I am exceedingly lucky to have friends like Aditi, Aparna, Swati and Vrinda who have always boosted my morale especially in times of difficulty.

I would like to thank all my friends and lab-mates – Aarohi, Abhay, Ajay, Ajit, Aparna Deshpande, Archana Damle, Archana Deshpande, Arundhati, Ashok Aspatwar, Ashok Giri, Ashwini, Mr B.G. Patil, Bimba, Dipak, Gauri Apte, Gauri Bhat, Jakir, Juri, Kanchan, Dr Lalitha, Manasi, Maneesha, Manisha, Medhavinee, Meena Arvindakshan, Mikhil, Milind, Monali, Mukund, Dr Nirmala, Pavan Prakash, Prachi, Rajesh, Rajsehkar, Raju, Rakhee, Rangarajan, Rashmi, Renu, Renuka, Ritika, Sami, Sanjay, Sastry, Shashi, Dr Shubhada, Smita, Suresh, Trupti, Varsha and Vijay for the lively atmosphere maintained in the lab.

I am grateful to Dr Meena Lagu, Dr Mohini, Suvarna and Sadhana for their help at all times. They have always tried to make available all the chemicals enzymes, etc required for my work. I am also thankful to Indira, Usha and Satyali for the help rendered by them.

I would also like to mention Mr Modak, Mr Kamthe, Mr Trehan and Mr Karanjkar for their technical help with instrumentation. Thanks are also due to Karunakaran, Jagtap, Mari, Kamble and Shikhre for their assistance.

I would like to thank Dr Paul Ratnasamy, Director, National Chemical Laboratory for providing the laboratory facilities during the tenure of my work. I acknowledge the financial support from Council of Scientific and Industrial Research, New Delhi in the form of Junior and Senior Research Fellowships. My work has mainly been supported by the grant from Rockefeller Foundation, USA, to whom I express my gratitude.

I would like to acknowledge Dr N.P. Sarma and Dr A.P.K Reddy, Directorate of Rice Research, Hyderabad, for making available the elite cultivars and bacterial blight resistant and susceptible genotypes of rice, respectively. I am grateful to Dr D.S. Brar for providing DNA of wild rice genera from the germplasm collection at International Rice Research Institute, Manila, Philippines. I also wish to thank Central Rice Research Institute, Cuttack, India, Agricultural Research Station, Vadgaon, India, USDA ARS, National Small Grains Collection, Aberdeen, USA, Indian Agricultural Research Institute, New Delhi, India and Agharkar Research Institute, Pune, India for providing genotypes of rice and other cereals. I am indebted to Dr Shiv Kumar Aggarwal for providing the coefficient of parentage data for the elite rice cultivars used in my work. The invaluable discussions with Dr Bapat and Prof E.A. Siddiq have been a tremendous help in planning my work. I especially wish to mention Prof M.V. Hegde, who is instrumental in initiating me into the field of molecular biology.

I would like to acknowledge all my teachers, friends, relatives and well-wishers who have helped me a different times but whose names I have been unable to mention here. I wish like to express my heartfelt gratitude to my parents, parents-in-law, brother Pettarusp and sister Piloo for their unstinting support and encouragement which has egged me on throughout my work. I am at a loss for words to thank my father whose invaluable advice I have always sought in times of need. Mere words are insufficient for the understanding and patience shown by my husband Piroze during the last four years, without whose continuous support and encouragement my thesis would not have seen the light of day.

Armaity Davierwala

LIST OF ABBREVIATIONS

AFLP	amplified fragment length polymorphism
ALP	amplicon length polymorphism
AP-PCR	arbitrarily primed polymerase chain reaction
BAC	bacterial artificial chromosome
BB	bacterial blight
BC	back-cross
bp	base pairs
CAPs	cleaved amplicon polymorphisms
cM	centimorgan
CP	coefficient of parentage
CTAB	hexadecyl-trimethyl-ammonium bromide
^{o}C	degree centigrade
dATP	deoxyadenosine 5' triphosphate
dCTP	deoxycytidine 5' triphosphate
dGTP	deoxyguanosine 5' triphosphate
dNTP	deoxynucleotide 5' triphosphate
dTTP	deoxythymidine 5' triphosphate
DNA	deoxyribonucleic acid
DHL	double haploid line
DTT	dithiothreitol
EST	expressed sequence tag
h	hour
ISSR	inter-simple sequence repeat
kb	kilo base pairs
MAS	marker-assisted selection
µCi	microcurie
µg	microgram
µl	microlitre
µM	micromolar
ml	millilitre
mM	millimolar
M	molar

min	minute
mya	million years ago
Myr	million years
ng	nanogram
PCR	polymerase chain reaction
pmoles	pico-moles
QTL	quantitative trait loci
RAPD	random amplified polymorphic DNA
RDA	representational difference analysis
RFLP	restriction fragment length polymorphism
RIL	recombinant inbred line
RNA	ribonucleic acid
rDNA	ribosomal DNA
rpm	revolutions per minute
SAP	specific amplicon polymorphism
SCAR	sequence characterized amplified region
SDS	sodium dodecyl sulphate
sec	second
SLV	spacer length variation
SSLP	simple sequence length polymorphism
SSR	simple sequence repeat
STMS	sequence tagged microsatellite site
STS	sequence tagged site
TAE	Tris-acetate-EDTA buffer
TBE	Tris-borate-EDTA buffer
TE	Tris-EDTA buffer
Tm	melting temperature
Tris	Tris-hydroxymethyl amino methane
U	units of enzyme
V	volt
YAC	yeast artificial chromosome

LIST OF FIGURES

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF TABLES

THESIS ABSTRACT

Rice is agronomically and nutritionally, one of the world's most important staple food crops, with approximately half of the world's population dependant on it for a significant proportion of their caloric intake. Different genotypes of rice are adapted to a wide range of environmental conditions from tropical flooding to temperate dryland, making it an excellent example for real-life, adaptive responses. Rice, a member of the Gramineae family, has a genome size of $0.45X10^{-9}$ bp (Arumunagathan and Earle, 1991), which is one-tenth the size of the human genome and is a model system for cereal genome analysis. A wide variety of DNA markers have been extensively used in rice for genetic diversity analysis, phylogenetic and evolutionary studies, mapping and tagging genes and quantitative traits of agronomic importance and marker-assisted selection. Moreover, these marker studies have been invaluable for positional cloning and synteny mapping of cereals. Of the several DNA markers available, microsatellites or SSRs (simple sequence repeats) are considered to be ideal molecular genetic markers. Microsatellite sequences are abundantly present in eukaryotic genomes and have, therefore, been used for genome analysis of many crop plants including rice (Gupta *et al*, 1994; Ramakrishna *et al*, 1994; 1995). Many microsatellite loci have been isolated and sequences flanking these loci have been used in polymerase chain reaction (PCR) to amplify different genotypes. Presence of a high level of allelic diversity at these loci makes them informative and valuable markers, with many applications in agriculture.

In my research I have used PCR amplifiable microsatellites to study a few specific aspects in rice and the important findings are as follows:

(1) Genetic diversity evaluation of Indian elite rice varieties using molecular markers

In India, rice is the most important crop, grown all over the country, in different ecological niches with a wide range of selected cultivars. In an attempt to quantify the genetic diversity in a large sample of elite rice cultivars developed in India from 1968 to 1994 and to correlate the diversity data with the coefficient of parentage, I used two different microsatellite-based DNA marker systems, namely, simple sequence length polymorphism SSLP () and inter simple sequence repeat (ISSR). The results obtained with these two

marker systems were correlated with those obtained earlier using random amplified polymorphic DNA (RAPD) markers. Although 153 bands were polymorphic out of a total of 168 bands amplified, the informativeness of the polymorphisms obtained using individual marker bands was low. A comparison of the dendrogram based on the polymorphism data using all the three marker systems with that using coefficient of parentage data revealed a very poor correlation, which could be due to several reasons as discussed in the thesis. The information that I have obtained from the dendrograms will aid rice breeders to select suitable genetically diverse parents in their crossing program to increase heterosis.

(2) Molecular characterization of $(GATA)_n$-containing microsatellites from rice and their potential in studying inter- and intra-specific variability in rice

Microsatellites or simple sequence repeats (SSRs) are 2-5bp in length and are widely dispersed throughout the nuclear genome of eukaryotes (Tautz and Renz, 1984). They are considered to be the most informative molecular genetic markers (Tautz *et al*, 1986; Tautz, 1989). Three polymorphic $(GATA)_n$-containing microsatellite loci viz. OS1A6, OS1H10 and OS2E7, containing 7-13 repeat motifs were isolated from the cultivated rice, *Oryza sativa* var. Basmati-370. Primers flanking these loci were used to screen 26 wilds (representing all the genomes of rice), 16 cultivars, 47 elite rice varieties of India and 37 lines resistant / susceptible to bacterial leaf blight to determine the degree of polymorphism. When the $(GATA)_n$-flanking primers were used to amplify DNA from different bacterial blight resistant and susceptible lines, it was observed that the lines clustered into different groups depending on the resistance genes present in them. Amplification of maize, wheat, barley and oat DNA with *O.sativa* flanking primers, indicated that these $(GATA)_n$-containing loci were conserved across different cereal genera.

(3) Expansion of $(GA)_n$ dinucleotide at a microsatellite locus associated with domestication in rice

Microsatellites undergo rapid changes over short evolutionary time periods, which can be phylogenetically informative in related species.

Dinucleotide repeats are abundantly distributed in the genome and are highly polymorphic in nature. I was therefore, interested in studying the dynamic changes occurring in microsatellite sequences during the process of evolution and to evaluate how these changes could provide useful information for phylogenetic studies. A microsatellite locus harboring $(GA)_n$ repeats was amplified from several wild and cultivated rices. Sequencing of these loci revealed an increase in repeat number from 14 in distantly related wilds to 24 in the widely grown present-day indica rice cultivars. Thus, I have observed repeat unit expansion of a $(GA)_n$ microsatellite in the process of cultivation of rice from its wild ancestors.

(4) Sequence variations at a complex microsatellite locus in rice and its conservation in cereals.

Microsatellites exhibit a high degree of polymorphism mainly due to variation in copy number of the basic motif. In an attempt to study changes associated with microsatellites in rice, DNAs of cultivated rice including indica and japonica varieties and wild rice genotypes were amplified by polymerase chain reaction with primers flanking the $(GATA)_n$ and $(AC)_n$ repeats at a microsatellite containing locus, OS1E6 (GenBank accession number AFO16647), previously reported from *Pst*I rice (var. Malkolam) genomic library in pUC18. Eight alleles of varying sizes were obtained which were cloned and sequenced. Sequencing data indicated that the size variations of different alleles were due to differences in the repeat number as well as sequence variations in the region flanking the microsatellite motifs.

Among cereals, rice has the smallest genome size and is considered to be an ideal crop to study cereal genome evolution and transpecies crop improvement. I was interested in studying if non-coding microsatellite-containing loci are also conserved / variable in different cereals. Different cereal DNAs were, therefore, amplified using primers flanking the OS1E6 locus and it was observed that this locus was present in various cereal genotypes analyzed indicating its conservation across different members of cereals.

Organization of Thesis

This thesis is organized into six chapters and the contents of each chapter are as follows:

Chapter1: Review of Literature - Molecular markers for unraveling the rice genome and their potential in rice improvement

In this chapter, I have discussed the wide and varied applications of molecular markers in rice genome research.

Chapter 2: Estimation of genetic diversity in Indian elite rice varieties using molecular markers

Here, I have described the ability of different molecular marker systems viz. ISSR and STMS to estimate genetic diversity of 42 Indian elite rice cultivars. The results obtained using these two marker systems have been compared with those obtained in an earlier study using RAPD markers.

Chapter3: Potential of $(GATA)_n$ microsatellites from rice for studying inter- and intra-specific variability

The intra- and inter-specific variability studies using the $(GATA)_n$-containing microsatellites from rice are included in this chapter.

Chapter 4: Expansion of $(GA)_n$ dinucleotide at a microsatellite locus associated with domestication in rice

The repeat expansions at a microsatellite locus amplified from different wild and cultivated genotypes of rice have been described in this chapter and the results are correlated with the process of domestication of this cereal.

Chapter 5: Sequence variations at a complex microsatellite locus in rice and its conservation in cereals

Here, I have included the results involving analysis of a complex microsatellite locus containing various types of repeat motifs in different wild and cultivated rice genotypes. This locus from rice has also been amplified in different cereal genera.

Chapter 6: General Discussion - Microsatellites: universal genetic mapping reagents

Here, I have very briefly outlined the summary of all the results obtained in the earlier four chapters and explained the utility of microsatellites in rice genome research.

The list of references and my curriculum vitae are given at the end.

CHAPTER 1

REVIEW OF LITERATURE

Molecular Markers for Unraveling the Rice Genome and Their Potential in Rice Improvement

1.1 Rice as a crop

Rice, which has been documented as a source of food as far back as 2500 BC in history books, has fed more people over a longer period of time than any other crop in the world (Rost, 1997). It provides the main source of food for approximately half of the world's population and, hence, may be the most important plant on this earth (Shimamoto, 1995; Goff, 1999) (Fig. 1.1). Rice is one of the three cereals produced annually at worldwide levels of approximately half a billion tons (Goff, 1999). However, unlike other major cereals, more than 90% of the rice is milled almost exclusively as food for human consumption (Goff, 1999) and forms three-fourths of the total diet for millions of people. This member of the grass family (Gramineae) is abundant in carbohydrates and is a major source of protein for the masses of Asia (Chang, 1984). Rice is the only cereal that can withstand flooding and produces more calories to sustain a larger number of persons per unit of land than any other cereal in monsoonal areas (Chang, 1984).

Considering the growth in human population every year all over the world, the demand of rice is ever increasing. However, the total area under rice cultivation has remained stable since 1980 (Khush, 1995), and hence, increasing yield is the only possible alternative to meet the anticipated higher demand of rice. Since conventional breeding tools which were responsible for the first 'Green Revolution', were by themselves unable to meet the growing demand for rice, the Rockefeller Foundation launched its rice initiative in the early 1980s, focussing on integrating biotechnology, chiefly genetic engineering, into rice research in order to boost its productivity. As a result, extensive work has been carried out on engineering rice varieties with resistance to stemborer, sheath blight, bacterial blight and viruses, enhanced tolerance to water submersion and improved nutritional qualities such as high iron and provitamin A content ('Golden Rice').

The development of various types of DNA markers in the late 1980s has provided a new platform for rice genome research, and there has been a tremendous progress in this field during the last decade (Fig. 1.2). The main purpose of this review is to give a glimpse into the various facets of rice genomics, where DNA markers have played a crucial role.

Figure 1.1: Rice plant – the major food source for 50% of the world's population

1.2 Molecular markers: a tool to study genetic diversity of rice

Modern agricultural practices have resulted in reduction in the genetic diversity of most major crops. Since the late 1960s and the 'Green Revolution' in rice, traditional rice varieties have been replaced by semi-dwarf cultivars, which now occupy more than 50% of the irrigated fields of Asia (Ikeda and Wasaka, 1997). However, this genetic uniformity has resulted in the loss of several valuable genes making rice more vulnerable to the vagaries of climate and associated biotic and abiotic stresses (Lee, 1995). An awareness of genetic diversity and management of crop genetic resources are, therefore, essential components of plant improvement programs.

Several approaches such as morphological and phenotypic characteristics (Tateoka, 1962; Morinaga, 1964), biochemical and isozyme markers (Second, 1982) and DNA markers (Tanksley *et al*, 1989; Paterson *et al*, 1991) have been used for detecting genetic diversity. However, DNA markers being innumerable and independent of growth stage, physiology of the plant and environment, provide a more powerful source of genetic polymorphism (Beckmann and Soller, 1986).

I am giving below a brief account of how DNA markers have been used in assessment of genetic diversity of rice germplasm and their exploitation to estimate genetic relationships, select suitable parents for crossing programs and for management and protection of germplasm.

1.2.1 DNA fingerprinting and genotype identification for management of germplasm collections and parentage analysis

For genotyping cultivars, multi-locus probes such as repetitive and arbitrary DNA markers are usually the markers of choice. In rice, DNA fingerprints were first obtained in rice by Dallas (1988), when he used a human minisatellite probe, 33.6, on genomic DNA from rice cultivars of *O.sativa* and *O.glaberrima* to obtain unique fingerprint profiles. Subsequently, several microsatellites such as $(GATA)_4$, $(GACA)_4$, $(GGAT)_4$, $(GAA)_6$ and $(CAC)_5$ (Gupta *et al*, 1994; Ramakrishna *et al*, 1994), interspersed repeated DNA sequences (Zhao and Kochert, 1993a) and minisatellites (Dallas *et al*, 1993; Winberg *et al*, 1993; Ramakrishna *et al*, 1995; Zhou and Gustafson,

1995) have been used to detect genetic variation and for cultivar identification in rice.

Restriction fragment length polymorphism (RFLP) probes have also been used to differentiate between indica and japonica rice cultivars at the nuclear (Wang and Tanksley, 1989; Zhang *et al*, 1992a; Ishii *et al*, 1995; Qian *et al*, 1995; Oba *et al*, 1996) and chloroplast (Ishii and Tsunewaki, 1991) DNA level and as species specific markers for the A genome of rice (Dallas *et al*, 1993). Sano and Sano (1990) and Cordesse *et al* (1990) have used ribosomal DNA (rDNA) spacer length variants to investigate genetic diversity in rice while Fukuchi *et al* (1993) have shown retrotransposon-mediated fingerprinting to be an efficient method to distinguish or identify cultivars of rice. Kajiya *et al* (1996) have employed a representational difference analysis (RDA) or genomic subtractive hybridization technique to develop RFLP markers between two closely related japonica varieties. This technique involves digestion of DNA with restriction endonucleases and addition of oligonucleotide adapters to the ends of the restriction fragments followed by PCR amplification of the DNA fragments. RDA is specifically designed to clone restriction fragments from one of the samples referred to as tester and cannot be used to clone the fragment from the other sample called as driver. Using rDNA intergenic spacer length polymorphism, Liu *et al* (1996) have assessed the genetic diversity in Chinese and Asian rice germplasm and observed that indica rice, grown in tropical and subtropical areas, has longer spacer length variations (SLVs), while japonica rice, grown in temperate regions has shorter SLVs. Accessions from Dongxiang County of Jiangxi Province, which is the highest altitude of a natural habitat for wild rice ever observed, have shorter SLVs than those from lower latitudes. Such differential occurrence may have significant implications regarding the origin and evolution of cultivated rice.

In addition to hybridization-based markers, which are time consuming and tedious, polymerase chain reaction (PCR)-based markers, which obviate these limitations of hybridization-based markers have been employed to estimate genetic relatedness in rice. Random amplified polymorphic DNAs (RAPDs) have been used for evaluation of genetic diversity in *O.sativa* cultivars (Fukuoka *et al*, 1992; Ko *et al*, 1994; Virk *et al*, 1995a; Kim *et al*,

1996) and upland and lowland varieties (Yu and Nguyen, 1994; Guimaraes *et al*, 1996), identification and parentage determination (Wang *et al*, 1994a). Several researchers have exploited the utility of inter-simple sequence repeats (ISSRs) and amplified fragment length polymorphisms (AFLPs) to study diversity and differentiate between indica and japonica subspecies (Parsons *et al*, 1997; Zhu *et al*, 1998; Blair *et al*, 1999; Virk *et al*, 2000). PCR-based sequence tagged microsatellite site (STMS) markers have played an important role for estimation of genetic variation between indica and japonica cultivars (Wu and Tanksley, 1993; Zhao and Kochert, 1993b; Yang *et al*, 1994; Panaud *et al*, 1996), upland rice accessions (Thanh *et al*, 1999) and identification of white milled cultivars from brown milled cultivars (Bligh *et al*, 1999). Kawase (1994) has demonstrated how restriction landmark genome scanning (RLGS), which provides unbiased information on the genetic polymorphism throughout the whole genome and, therefore, an accurate estimation of genetic similarity, can be a tool to fingerprint rice cultivars.

The accurate, rapid and unambiguous identification of cultivars has gained importance for plant variety protection with the agreement on Trade Related Intellectual Property Rights (TRIPS), which is a part of the General Agreement on Tariffs and Trade (GATT), to protect plant breeder's rights. The choice of appropriate DNA profiling technique depends on the aims of the testing. In rice, ISSRs have revealed slightly higher levels of polymorphism as compared to RAPDs (Parsons *et al*, 1997), while STMS markers are observed to be more robust and efficient, resulting in a greater number of alleles than either RFLPs or AFLPs (Olufowote *et al*, 1997). Thus, STMS markers which generate simple sequence length polymorphisms (SSLPs) have been found to discriminate between closely related accessions and genotypes with a narrow genetic base (Yang *et al*, 1994; Olufowote *et al*, 1997).

Genotyping of cultivars is essential for effective management and utilization of germplasm collections. Germplasm management is a multi-faceted endeavor, involving acquisition, maintenance and characterization such that the biodiversity of plant genetic resources is conserved and utilized for crop improvement. The rice germplasm collection at the International Rice Genbank at the International Rice Research Institute (IRRI), Los Banos,

Philippines, comprises more than 80,000 registered samples, including landraces, modern and obsolete varieties, breeding lines and special genetic stocks, twenty-one wild species in the genus *Oryza* and related genera in the tribe *Oryzae* (Jackson, 1997). Large germplasm collections are also maintained in India, China, Thailand, Indonesia, Japan and several other countries. Totally 120,000 varieties of rice are estimated to exist in the world (Khush, 1997). Reliable identification of duplicate accessions will lead to efficient and effective management of germplasm collections. RAPDs have shown the ability to identify duplicate accessions within a germplasm collection (Virk *et al*, 1995b). Martin *et al* (1997) have further proposed that five of the forty accessions designated as *O.glumaepatula* have been mis-identified based on RAPD data.

Microsatellites and minisatellites represent single loci and can, therefore, be used for parentage and pedigree analysis. These markers being multiallelic can reliably detect allelic variability across a wide range of germplasm, allowing individuals to be uniquely genotyped, so that gene flow and paternity can be established (Joshi *et al*, 1999). Dallas (1988) has been the first researcher to carry out parentage analysis in plants using a human minisatellite probe, 33.6. Similarly, microsatellites are stable enough to reliably trace the flow of monogenic or quantitative trait loci of interest in rice pedigrees as demonstrated by Panaud *et al* (1996) and Ayres *et al* (1997) in analyses spanning over 60 years of rice breeding and seed amplification history.

1.2.2 Estimation of genetic relatedness for selection of suitable parental genotypes

Information on genetic diversity in a crop species is important for selection of suitable parental genotypes. Predicting hybrid performance and heterosis has always been the primary objective in all hybrid crop-breeding programs (Hallauer and Miranda, 1988). In a crop such as rice, where hybrids are commercially important, it is imperative to exploit the heterosis among genetically divergent germplasm because the level of genetic diversity has been proposed as a possible predictor of F1 performance and heterosis. It has been observed that rice hybrids showing strong heterosis were usually

developed from parental lines diverse in relatedness, ecotype, geographic origin, etc (Lin and Yuan, 1980; Yuan, 1985; Yuan and Cheng, 1986).

Two measures of heterozygosity used to calculate the correlations of heterozygosity with hybrid performance and heterosis, are general heterozygosity, which is the heterozygosity at all the marker loci and specific heterozygosity, which is the heterozygosity calculated on the basis of positive markers for each trait. Using RFLP and STMS markers, low correlation was obtained between heterozygosity and F1 performance, while high correlations were obtained between specific heterozygosity and mid-parent heterosis (Zhang *et al*, 1994a; 1995), indicating that specific heterozygosity may find practical applications for prediction of heterosis. Thus, in reality, it may be much more practical to predict heterosis on the basis of a small number of informative markers rather than to use a large number of markers covering the entire genome corroborating the suggestion from theoretical calculations (Bernardo, 1992, Charcosset, *et al*, 1991). If such correlations can be confirmed using a larger sample size rather than just eight parental lines used in these analyses, it can help in planning the most productive crosses in hybrid breeding programs. A heterotic group is a collection of germplasm, that when crossed to germplasm external to its group (usually another heterotic group), tends to exhibit a higher degree of heterosis on an average, than when crossed to a member of its own group (Lee, 1995). Analysis of the molecular data resolved the eight lines, representing a major portion of the elite rice germplasm used in hybrid rice programs in China, into three well separated groups which essentially agreed with the available pedigree information (Zhang *et al*, 1995).

Zhang *et al* (1996c) have observed varying degrees of correlation between molecular marker heterozygosity and heterosis in indica and japonica rice. When RFLP and STMS markers have been used to analyze U.S. Southern long grain rice, F1 heterozygosity has been found to be highly correlated with rough rice yield and head rice yield, and significantly correlated with heterosis of these two traits, suggesting that molecular markers may be partially useful for predicting the performance of inter-subspecific hybrids (Saghai Maroof *et al*, 1997). Xiao *et al* (1996b) have observed a significant positive correlation between heterosis in F1 hybrid and

genetic distance based on RAPD and STMS markers in ten elite inbred lines consisting of four japonicas and six indicas, widely used in the hybrid rice research program. They have concluded that genetic distance measures based on these two molecular marker types, may be useful for predicting yield potential and heterosis of intra-subspecific hybrids, but not of inter-subspecies hybrids. These results suggest that parental genotyping based on RAPD and STMS markers could be useful in reducing the field work associated with making crosses and hybrid field testing when attempting to identify intra-subspecies hybrids possessing high yield potential as targets for transfer of key genes in rice.

1.2.3 Wild and exotic germplasm: a means to widen the genetic base of rice

Intense selection pressure for specific traits such as yield, due to domestication and modern breeding practices, has resulted in the narrowing of the genetic base of rice thus rendering it vulnerable to epidemics and disasters. Although wild and unadapted germplasm is less desirable than modern varieties due to its overall appearance and performance, it is a unique reservoir of genetic variation, containing genes which confer resistance to biotic stresses such as pests and diseases and tolerance to abiotic stresses such as drought, cold and salinity. Such important traits from landraces and wild species when incorporated into economically important varieties, will have an enormous impact on the productivity of rice. Using biotechnological tools several hybrids have been produced between *O.sativa* and many wild species (Khush *et al*, 1993).

There are very few reports on the assessment of genetic variability in landraces of rice as compared to modern cultivars. Studies using different molecular markers such as RFLP, RAPD, STMS, AFLP, etc to analyze genetic diversity among landraces from different ecogeographic locations (Second, 1991; Chen *et al*, 1994; Yang *et al*, 1994; Virk *et al*, 1995a; Ghareyazie *et al*, 1995; Kim *et al*, 1996) have revealed that STMS markers detected the largest number of alleles per locus as compared to any other marker system (Yang *et al*, 1994). Using rDNA intergenic spacer length

polymorphism, Liu *et al* (1996) assessed 272 Chinese landraces, which represented most of the rice growing areas of China.

Genetic diversity of wild rice genotypes has been analyzed using RFLP (Zhao *et al*, 1989; Dally and Second, 1990), rDNA spacer length polymorphisms (Sano and Sano, 1990), direct amplification of minisatellite-region DNA (DAMD) with minisatellite core sequences as primers (Zhou *et al*, 1997), simple sequence repeats from chloroplast genomes (Provan *et al*, 1997) and sequence variations at specific genes (Barbier and Ishiyama, 1990). Several genome-specific sequences have been identified (Zhao *et al*, 1989; Reddy *et al*, 1993; Kiefer-Meyer *et al*, 1995; Zhou *et al*, 1997), which might be useful in classifying unknown species of wild or domestic rice, reclassifying misclassified accessions of wild rice genera, tracing the genomic origin of different species and studying genome evolution at the molecular level. Moreover, one could also follow the transfer of a genome-specific DNA sequence when monitoring the success of crosses between different rice genome types during plant breeding experiments as observed by Jena *et al* (1992) and Brar and Khush (1997) during introgression of RFLP alleles from *O.officinalis*, *O.brachyantha* and *O.granulata* into cultivated rice, *O.sativa*.

1.3 Taxonomic, phylogenetic and evolutionary studies

Information on the genetic variation and phylogenetic relationships in the germplasm of a crop is the basic requirement in plant breeding. For improving crop cultivars, it is very important to find plant material, which has the appropriate genotype for a given purpose. The genus *Oryza*, which comprises 23 species and 10 recognized genome types, represents an enormous gene pool for genetic improvement of rice cultivars (Table 1.1). Clarification of phylogenetic relationships of rice genomes is, therefore, crucial for effective utilization of this rice germplasm.

1.3.1 Classification of germplasm

The taxonomic classification of germplasm is the first step in evolutionary and phylogenetic studies as it helps to determine whether the germplasm is a part of the primary, secondary or tertiary gene pool of the

Table 1.1: Chromosome number, genomic composition and potential useful traits of different species in the genus *Oryza* (Khush, 1997; Brar and Khush, 1997; Aggarwal *et al*, 1997, Ge *et al*, 1999; USDA, ARS, National Genetic Resources Program, 2001)

Species	**Chromosome no. (2n)**	Genome	Distribution	**Useful or potentially useful traits**[#]
***O.sativa* complex**				
O.sativa L.	24	AA	Worldwide	Cultigen
O.nivara Sharma et Shastry	24	AA	Tropical and subtropical Asia	Resistance to grassy stunt virus*, blast, drought avoidance
O.rufipogon Griff.	24	AA	Tropical and subtropical Asia, tropical Australia	Elongation ability*, resistance to BB, tungro*, source of CMS, tolerance to acid sulfate soils*
O.breviligulata A. Chev. et Roehr.	24	AA	Africa	Resistance to GLH, BB, drought avoidance
O.glaberrima Steud.	24	AA	West Africa	Cultigen
O.longistaminata A. Chev. et Roehr.	24	AA	Africa	Resistance to BB*, drought avoidance
O.meridionalis Ng.	24	AA	Tropical Australia	Elongation ability, drought avoidance
O.glumaepatula Steud.	24	AA	South and Central America	Elongation ability, source of CMS*
***O.officinalis* complex**				
O.punctata Kotschy	24	BB	Africa	Resistance to BPH, zigzag leafhopper
O.punctata Steud	48	BBCC	Africa	Resistance to BPH, zigzag leafhopper
O.minuta J.S. Pesl. ex C.B. Presl	48	BBCC	Philippines and Papua New Guinea	Resistance to blast*, sheath blight*, BB*, BPH*, GLH
O.officinalis Wall ex Watt	24	CC	Tropical and subtropical Asia, tropical Australia	Resistance to thrips, tungro*, BB*, BPH*, GLH, WBPH*
O.rhizomatis Vaughan	24	CC	Sri Lanka	Drought avoidance, rhizomatous
O.eichengeri A. Peter	24	CC	South Asia and East Africa	Resistance to yellow mottle virus, BPH, WBPH, GLH
O.latifolia Desv.	48	CCDD	South and Central America	Resistance to BB*, BPH*, high biomass production
O. alta Swallen	48	CCDD	South and Central America	Resistance to striped stemborer, high biomass production
O.grandiglumis (Doell) Prod.	48	CCDD	South and Central America	High biomass production
O.australiensis Domin.	24	EE	Tropical Australia	Drought avoidance, resistance to BB*, BPH*
O.brachyantha A. Chev. et Roehr.	24	FF	Africa	Resistance to BB*, YSB*, leaf - folder, whorl maggot, tolerance to laterite soil
***O.meyeriana* complex**				
O.granulata Nees et Arn. ex G. Watt	24	GG	South and Southeast Asia	Shade tolerance, adaptation to aerobic soil, resistance to BPH*
O.meyeriana (Zoll. et Moritzi ex Steud.) Baill.	24	GG	Southeast Asia	Shade tolerance, adaptation to aerobic soil
***O.ridleyi* complex**				
O.longiglumis Jansen	48	HHJJ	Irian Jaya, Indonesia and Papua New Guinea	Resistance to blast, BB
O.ridleyi Hook. f.	48	HHJJ	South Asia	Resistance to stemborer, whorl maggot, blast, BB, YSB*
New genome				
O.schlechteri Pilger	48	HHKK	Papua New Guinea	Stoloniferous
Porteresia coarctata (Roxb.) Tateoka	48	HHKK	Bangladesh	-

[#]BPH: Brown planthopper, GLH: green leafhopper, WBPH: white backed planthopper, BB: bacterial blight, YSB: Yellow stemborer, CMS; cytoplasmic male sterility

*Material from which genes have been or are being introgressed into cultivated rice

crop. Glaszmann (1987) has classified 1688 traditional rices from Asia into six varietal groups based on isozyme analysis with eight enzymes. Later on Mochizuki *et al* (1993) have classified rice genotypes with AA genome into ten groups using short interspersed elements (SINEs) while Mackill (1995) and Akagi *et al* (1997) have classified japonica rice cultivars using RAPD and STMS markers, respectively. Mackill *et al* (1996) have observed that rice cultivars could be classified into the same subspecies (indica / japonica) and into the same subgroup (temperate japonica / tropical japonica) using RAPDs and AFLPs. Amplicon length polymorphisms (ALPs) and PCR-based RFLP have also been used to classify rice germplasm (Ghareyazie *et al*, 1995; Xu *et al*, 1998) and the classification using ALPs corresponded well with that of isozyme analysis (Xu *et al*, 1998). Thus, the ALP mediated rice germplasm classification system is observed to be at least as accurate and robust as that using morphological, physiological, biochemical or DNA (RFLP, RAPD and SSR) markers.

1.3.2 Study of phylogenetic relationships in the genus *Oryza*

Prior to the last decade, understanding of the phylogenetic relationships in the genus *Oryza* (Oka, 1988; Vaughan, 1989) has been mainly due to cytogenic (Morinaga, 1964; Hu, 1970; Katayama, 1982), morphometric (Tateoka, 1962), isozyme (Second, 1982; 1985a; 1985b) and chloroplast DNA restriction (Dally and Second, 1990) analysis. However, with the introduction of DNA markers, additional inputs have been available for reconstructing phylogenies in the genus *Oryza*.

Barbier *et al*, (1991) and Barbier and Ishiyama (1990) have proposed a model for phylogenetic relationships of annual and perennial wild rice species based on DNA sequence variation data at the phytochrome and prolamin genes, respectively. Use of nuclear (Wang *et al*, 1992) and mitochondrial (Second and Wang, 1992) DNA RFLP represents one of the initial efforts to determine phylogenetic relationships among species in the genus *Oryza*. Later, Kim *et al* (1996) have used RFLPs and RAPDs to construct phylogenies of Korean varieties and landraces. When RAPDs were used to classify phylogenetic relationships in the A genome species, most of the accessions clustered together corresponding to their taxonomic grouping (Ishii

et al, 1996). AFLPs have also been used to determine phylogenetic relationships among *Oryza* species (Aggarwal *et al*, 1999). Joshi *et al* (2000) have used ISSRs to determine phylogenetic relationships between 42 genotypes of the genus *Oryza*, including 17 wild species representing nine different genomes (AA, BB, CC, EE, FF, GG, BBCC, CCDD, HHJJ), two cultivated species, *O.sativa* (AA) and *O.glaberrima* (AA) and three related genera *Porteresia coarctata*, *Leersia* and *Rhynchoryza subulata*. However, Ge *et al* (1999) have recognized the presence of an additional type HHKK for *Oryza schlechteri* and *Porteresia coarctata*, suggesting that *Porteresia coarctata is an Oryza* species based on the phylogenetic relationships obtained by comparing two nuclear gene (*Adh*-1 and *Adh*-2) trees and a chloroplast gene (*mat*k) tree of all 23 rice species and nine recognized genomes.

1.3.3 Evolution of the genus *Oryza*

The origin, evolution, cultivation and diversification of two cultivated species of rice viz. *O.sativa* and *O.glaberrima*, interest not only biologists but also geographers, archeologists, anthropologists, philologists, historians and other social scientists. Based on restriction endonuclease analysis of chloroplast DNA, RFLP and RAPD analysis of nuclear DNA from diploid A genome species of rice it has been proposed that these two cultivated rices originated from the Asian form of *O.perennis* and from *O.breviligulata* (also called *O.barthii*), respectively (Ishii *et al*, 1988; Wang *et al*, 1992; Ishii *et al*, 1996). This supports the earlier hypothesis put forth by several workers that *O.barthii* is the progenitor of the African cultivated rice *O.glaberrima* (Morishima *et al*, 1963; Second, 1982). However, Wang *et al* (1992) and Yi *et al* (1995) have proposed that the origin of cultivated rice *O.sativa* is Asian *O.rufipogon* using RFLP and arbitrarily primed PCR (AP-PCR) analysis. Joshi *et al* (2000) have further shown that *O.nivara*, the annual form of *O.rufipogon*, is the progenitor of cultivated rice using ISSR markers. They have observed that all the cultivars first grouped with *O.nivara* and then clustered with the perennial forms of *O.rufipogon* and *O.perennis*. Further differentiation of the cultivated species *O.sativa* into its two subspecies viz. indica and japonica is an interesting feature in the evolution of rice. The two schools of thought

regarding the evolution of these two ecotypes are monophyletic and diphyletic evolution. The monophyletic path has been supported by several workers (Ting, 1957; Oka and Chang, 1962; Oka and Morishima, 1982; Wang *et al*, 1984; Glaszmann, 1987; Oka, 1988; 1991) and more recently by Joshi *et al* (2000), who used ISSR markers for the same. The diphyletic origin of indica and japonica rice has been proposed by Chinese workers (Oka, 1988) and later supported by restriction endonuclease analysis of chloroplast DNA from A genome species (Ishii *et al*, 1988), RFLP analysis (Wang *et al*, 1992) and AP-PCR analysis (Yi *et al*, 1995).

Wang *et al* (1992) have proposed that BBCC tetraploids (*O.malampuzhaensis*, *O.punctata* and *O.minuta*) are likely to be of independent origin or, if they originated from a single ancestral tetraploid species, introgression has been occurring between these tetraploids and sympatric C genome relatives. They have also shown that CCDD tetraploid species, *O.latifolia*, *O.alta* and *O.grandiglumis* have a closer affinity to each other than to any known diploid species which has also been observed by Aggarwal *et al* (1997; 1999) and Joshi *et al* (2000) using whole genome hybridization, AFLP and ISSR analyses, respectively. The uniqueness of the CCDD genome may be due to the unique DD genome component, whose origin still remains a mystery suggesting the possibility of an ancient origin for these American rice species. The closest living diploid relatives of these CCDD genome species belongs to the C genome (*O.eichengeri*) and E genome (*O.australiensis*) species (Wang *et al*, 1992; Ge *et al*, 1999). Although the different marker types have provided valuable information regarding evolution and phylogeny of various species, the trend is now increasingly shifting towards the use of expressed sequence tags (ESTs) for such analyses, since one actually looks at the evolution of functional genes here (Deshpande *et al*, 1998; Ge *et al*, 1999). Successful amplification of fragments of chloroplast DNA from charred grains of rice which are 2700 to 6000 years old (Sato *et al*, 1995) will further aid in evolutionary studies of rice. Thus, molecular marker studies have been extremely useful in classifying rice germplasm, reconstructing phylogenies and studying the origin and evolution of the genus *Oryza* and are expected to contribute more in the near future.

1.4 Mapping and tagging genes and QTLs of agronomic importance

The improvement of even the simplest of characteristics, often requires the manipulation of a large number of genes (Flavell, 1995). Recent advances in the construction of saturated linkage maps in rice has made it possible to map and tag genes of economic importance with molecular markers (Beckmann and Soller, 1986; Barnes, 1991; Mohan *et al*, 1997a; Kumar, 1999).

1.4.1 Construction of genetic linkage maps

Over the past 80 years, a classical genetic map of rice, comprising 200 phenotypic trait markers, including loci for morphological mutants, disease resistance and isozymes has been described by many geneticists and breeders (Kinoshita, 1995). The first rice molecular genetic map has been developed at Cornell University by McCouch *et al* (1988) followed by the second RFLP map by Saito *et al* (1991) (Table 1.2). The third molecular map constructed from an inter-specific cross could readily assemble 12 linkage groups without any unlinked segments, which were observed in the intraspecific cross (McCouch *et al*, 1988) using the same number of markers.

The group at Rice Genome Project (RGP) in Japan, have constructed a high resolution linkage map consisting of 1383 DNA markers at an average interval of 300kb, which is the densest genetic map available in plants (Kurata *et al*, 1994b). The total number of markers on this map has now increased to approximately 2300 (Nagamura *et al*, 1997) including 883 expressed sequence tags (EST) markers, 260 newly identified rice genes, one-fifth of which are ribosomal protein genes (Wu *et al*, 1995) and four telomere-associated sequences (Ashikawa *et al*, 1994). The mapping of expressed gene sequences is an important tool for constructing a detailed genetic map since it resolves genome organization in terms of their functional genes and their coordinated expression. Xiong *et al* (1997) have observed that markers in their map (Table 1.2) have bridged some of the major gaps in the individual Cornell and RGP maps due to integration of the markers from both the maps and a majority of the marker loci are well conserved between the two maps. Using the centromere- mapping technique, a linear correspondence has been observed between these two maps (Cornell and RGP) facilitating their

integration (Singh *et al*, 1996). Table 1.2 lists the different genetic linkage maps available in rice using molecular markers.

In order to further saturate the existing maps, Williams *et al* (1991) have used PCR-based RFLP for the first time in rice, followed by minisatellite, STMS, AFLP and centromere-associated sequences (Wu and Tanksley, 1993; Akagi *et al*, 1996a; Cho *et al*, 1996; Mackill *et al*, 1996; Miyao *et al*, 1996; Panaud *et al*, 1996; Chen *et al*, 1997a; Gustafson and Yano, 2000; Temnykh *et al*, 2000; Wang *et al*, 2000). Totally 323 STMS markers have been mapped onto the rice genetic map (Cho *et al*, 2000), a high percentage of which are ESTs. Maheshwaran *et al* (1997) have mapped 208 AFLP markers derived from only 20 primer combinations onto the molecular map of rice and observed very little intermingling of AFLP and RFLP loci. Recently, Zhu *et al* (1999) have mapped 300 AFLP markers in two rice populations and have observed that the genetic maps are aligned with almost full coverage of the rice genome. The AFLP markers are distributed throughout the 12 chromosomes and have helped to fill several gaps left by RFLPs and SSLPs (Cho *et al*, 1998).

Molecular maps constructed using 300 DNA markers in five different indica / japonica crosses in cultivated rice, *O.sativa*, show the same linkage order (Antonio *et al*, 1996). Comparative mapping using RFLP markers has revealed that gene order is highly conserved between the CC genome of a wild rice, *O.officinalis* and the AA genome of a cultivated rice, *O.sativa* (Jena *et al*, 1994) and between the diploid AA genome of *O.sativa* and the allotetraploid CCDD genome of *O.latifolia* (Huang and Kochert, 1994) with the exception of some rearrangements, inversions and deletions. Maps developed from inter-specific crosses show comparable overall map length and good colinearity with intraspecific maps (Lorieux *et al*, 2000). Comparative mapping will help to predict the positions of the genes, which are mapped in one species but not in the other.

Table 1.2: Different genetic linkage maps of rice using molecular markers

Sr. No.	Cross combination	Type of population used for mapping	Number of markers mapped	Types of markers mapped	Reference
1	Intra-specific indica X japonica (IR 34583) (Bulu dalam)	F2	135	RFLP	McCouch *et al*, 1988
2	Intra-specific indica X japonica (Kasalath) (Fl 134)	F2	347	RFLP	Saito *et al*, 1991
3	Inter-specific *O.sativa* X *O.longistaminata*	Backcross	726	RFLP (from rice, barley, maize, oat,), micro-satellite, telomere markers, cloned genes, isozymes, morphological mutants	Causse *et al*, 1994
4	Intra-specific indica X japonica (Kasalath) (Nipponbare)	F2	1383	RFLP, RAPD, ESTs	Kurata *et al*, 1994b
5	Intra-specific indica X japonica	DHL	181	RFLP	Zhu *et al*, 1993
6	Intra-specific indica X japonica (IR 64) (Azucena)	DHL	135	RFLP	Huang *et al*, 1994
7	Intra-specific indica X japonica (IR 24) (Asominori)	RIL	375	RFLP	Tsunematsu *et al*, 1996
8	Intra-specific indica X japonica (Kasalath) (Fl 134)	DHL – re-generated plants from anther culture of F1 hybrid	-	RFLP	Yamagishi *et al*, 1996
9	Intra-specific japonica X tongil (Gihobyeo) (Milayang)	RIL	545	RFLP, AFLP, SSLP, isozymes, morphological mutants	Cho *et al*, 1998
10	Inter-specific *O.glaberrima* X *O.sativa*	BC1	-	STMS, STS	Lorieux *et al*, 2000
11	Inter-specific *O.sativa* X *O.rufipogon*	F2	771	RFLP	Xiong *et al*, 1997
12	First genetic map of wild rice *Zizania palustris* L.	-	121	RFLP	Kennard *et al*, 2000

1.4.2 Mapping and tagging genes of economic importance

Mapped DNA markers provide a framework for locating and tagging a number of agronomically important genes. Rice crop experiences severe losses due to damage by fungal and bacterial diseases and pests. Developing resistant cultivars is often the most economical and effective method of controlling diseases. Tagging resistance genes with molecular markers will help in marker-assisted selection (MAS) in breeding programs.

Blast is the most devastating disease of rice, followed by bacterial blight (BB) caused by *Xanthomonas oryzae* pv *oryzae* (*Xoo*). Several genes conferring resistance to diseases caused by different fungi, bacteria and insects have been tagged in rice using molecular markers (Table 1.3). By classical linkage analysis, the brown planthopper gene, *Bph*-1 has been located on chromosome 4. However, RFLP analysis reveals that the *Bph*-1 gene is located on chromosome 12 and not chromosome 4 (Hirabayashi and Ogawa, 1995).

Several genes responsible for quality improvement have also been mapped in rice (Table 1.4). The nuclear restorer gene for fertility, *Rf*-1, has been tagged using an ISSR marker (AG)8YC (UBC835) (Akagi *et al*, 1996b). Codominant markers developed from the flanking ISSR polymorphisms will not only be useful for breeding restorer lines but will also accelerate the breeding of maintainer lines by eliminating the *Rf*-1 gene. It can also be applied to seed purity management of hybrid rice seeds by eliminating contamination of CMS lines, maintainer lines and restorer lines from F1 hybrid rice at seedling stage.

1.4.3 Analysis of quantitative trait loci (QTLs)

Most of the agronomically and economically important traits in rice, like yield and yield components (grain number, grain weight, plant height, days to flowering, maturity date, etc.) are controlled by a relatively large number of loci, each of which makes a small positive or negative contribution to the final phenotypic value of the trait. Such loci are called as quantitative trait loci (QTLs), and the traits showing continuous variation in phenotype are termed as polygenic traits, because the final expression of the phenotype is governed by the genetic variation at a large number of loci, modified by

Table 1.3: List of disease resistance genes mapped in rice

Sr. No.	Resistance gene	Marker system	Chromosomal location	Reference
A	**Bacterial blight resistance genes**			
1	*Xa*-1	RFLP	4	Yoshimura *et al*, 1992
		RAPD		Yoshimura *et al*, 1995b
2	*Xa*-3	RFLP	11	Yoshimura *et al*, 1995a
3	*Xa*-4	RFLP	11	Yoshimura *et al*, 1995a
4	*xa*-5	RFLP	5	McCouch *et al*, 1992; Blair and McCouch, 1997
		CAPs, STMS		Blair and McCouch, 1997
5	*Xa*-10	RAPD	11	Yoshimura *et al*, 1995a
6	*xa*-13	RAPD	8	Zhang *et al*, 1996a
		CAPs		Zhang *et al*, 1996a; Huang *et al*, 1997
7	*Xa*-21	RFLP, RAPD	11	Ronald *et al*, 1992
		SCAR		Chungwongse *et al*, 1993; Wang *et al*, 1996; Williams *et al*, 1996
		STS		Lu *et al*, 1996; Williams *et al*, 1996
8	*Xa*-22(t)	RFLP	11	Lin *et al*, 1996
B	**Blast resistance genes**			
1	*Pi*-1	RFLP	11	Hittalmani *et al*, 2000
2	*Pi*-2(t)	RFLP	6	Yu *et al*, 1991
		CAPs		Hittalmani *et al*, 1994
3	*Pi*-3(t)	RFLP	6	Inukai *et al*, 1992
4	*Pi*-4(t)	RFLP	12	Yu *et al*, 1991
5	*Pi*-5(t)	RFLP	4	Wang *et al*, 1994b
6	*Pi*-7(t)	RFLP	11	Wang *et al*, 1994b
7	*Pi*-9(t)	RAPD	-	Naqvi *et al*, 1995
8	Pi-10	RAPD, SCAR	5	Naqvi *et al*, 1995; Naqvi and Chattoo, 1996; Hittalmani *et al*, 2000
9	*Pi*-19(t)	-	12	Babujee & Gnanamanickam, 2000
10	*Pi*-b	RFLP, RAPD	2	Miyamoto *et al*, 1996
11	*Pi*-z5	RFLP, CAPs	6	Hittalmani *et al*, 2000
12	*Pi*-t	STS	-	Hittalmani *et al*, 1995
13	*Pi*-ta	**RFLP**	12	Hittalmani *et al*, 2000
C		Gall midge resistance genes		
1	*Gm*-2	RFLP	4	Mohan *et al*, 1994
		RAPD		Mohan *et al*, 1994; Nair at al 1995a
2	*Gm*-4	**RAPD**	8	Nair at al 1996
		RFLP, STS		Mohan *et al*, 1997b
3	*Gm*-2, *gm*-3, *Gm*-4, *Gm*-5, *Gm*-6t	RAPD, STS	-	Katiyar *et al*, 1999
D	**Brown planthopper resistance gene**			
1	*bph*	RFLP	12	Ishii *et al*, 1994; Hirabayashi and Ogawa, 1995
E		Stripe disease resistance gene		
1	*stvb*-i	**RFLP**	11	Saito *et al*, 2000
F	**White backed planthopper resistance gene**			
1	*wbph*1	RFLP	-	McCouch *et al*, 1991

Table 1.4: **Genes mapped for rice quality improvement using molecular markers**

Sr. No.	Gene of interest	Marker system	Chromosomal location	Reference
A	**Genes for rice grain quality**			
1	Aroma gene, fgr	RFLP STS	8	Ahn *et al*, 1992 Garland *et al*, 2000
2	*Waxy* gene for amylose biosynthesis, *wx*	STMS	6	Bligh *et al*, 1995
3	Gene for super-giant embryo character, ge^s	RFLP, STMS	7	Koh *et al*, 1996
B	**Genes for rice improvement**			
1	Photoperiod sensitivity gene, *se*-1	Isozyme, RFLP RAPD, SCAR	6	Mackill *et al*, 1993 Monna *et al*, 1995
2	Semi-dwarfism genes			
	(a) *sd*-1	RFLP	1	Cho *et al*, 1994
	(b) *sd*-g	RFLP	5	Liang *et al*, 1994
3	Gene for heading date	RFLP	7	Li *et al*, 2000
4	Gene for earliness	RFLP	10	Ishii *et al*, 1994
C	**Genes for heterosis**			
1	Wide compatibility gene	Isozyme,	6	Liu *et al*, 1992
	(a) *S-5*	RFLP		Liu *et al*, 1992; Zhang *et al*, 1992b; Yanagihara *et al*, 1995; Liu *et al*, 1997
	(b)-	RFLP	2	Liu *et al*, 1997
	(c)-	RFLP	12	Liu *et al*, 1997
2	Pollen sterility gene, *S-a*	RFLP, RAPD	1	Zhuang *et al*, 1999
3	Nuclear restorer gene for fertility, *Rf-1*	ISSR	10	Akagi *et al*, 1996b
4	TGMS genes			
	(a) -	RAPD	8	Wang *et al*, 1995a
	(b) *tms*-3(t)	RAPD, SCAR	6	Lang *et al*, 1999
	(c) *tms*-4(t)	AFLP, STMS	2	Dong *et al*, 2000
	(d) -	AFLP, RAPD, STMS, STS	9	Reddy *et al*, 2000
5	PGMS genes			
	(1) *pms*-1	-	7	Zhang *et al*, 1994b
	(2) *pms*-2	-	12	Zhang *et al*, 1994b

environmental effects. While the theory and techniques of quantitative genetics (Falconer, 1960) have proven useful in the study of quantitative traits, these characters continue to be more difficult to manipulate than single gene traits in breeding programs (Tanksley *et al*, 1989). However, if these complex traits could be resolved into their individual genetic components, it might be possible to deal with these characters with the efficacy of single gene traits (Tanksley *et al*, 1989). The progress in development of newer and newer DNA markers followed by the subsequent 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). By using DNA markers for QTL analysis, it is possible to distinguish individual genetic components that are sometimes masked by the interaction of major genes and by the environment (Ikeda and Wasaka, 1997). With molecular markers, it is possible to assign chromosomal positions to individual QTLs, to determine the types and magnitude of gene effects of individual QTLs, and also to determine which parent possesses the positive allele at each QTL (Edwards *et al*, 1987; Tanksley *et al*, 1989; Lander and Botstein, 1989; Paterson, 1995; Stuber, 1995). Also, DNA markers linked to QTLs are exceedingly valuable for genetic diagnostics, as the breeding processes in quantitative traits require an enormous time frame and logistics.

Several morphological (Tables 1.5 and 1.6) and physiological (Table 1.5) traits and QTLs responsible for disease resistance (Table 1.7) have been tagged in rice using molecular markers. Li *et al* (1999) have observed that the allele from the indica rice, Teqing, called *Xa*-4T, was found to act as a dominant resistance gene against two strains of *Xoo*, while it acted as a recessive QTL against one strain of *Xoo* indicating that in addition to single genes, QTLs are also responsible for governing resistance to diseases.

Although wild and unadapted germplasm is a unique source of genetic variation, very often the superior trait of interest cannot be identified phenotypically making it virtually impossible to utilize wild accessions for improvement of quantitatively inherited traits. The reason for this is that in wild germplasm these desirable alleles are present in low frequency and are often masked by the effects of the deleterious alleles while in elite cultivars these positive trait-enhancing alleles are present in high frequency and can,

Table 1.5: List of mapped morphological and physiological traits in rice

Sr. No.	Quantitative trait loci (QTLs)	No. of QTLs	Marker system	Chromo-somal location	Reference
A	**Seedling vigor**				
1	Shoot length	4	RFLP	1, 3, 5, 9	Redona & Mackill, 1996a
2	Root length	7	RFLP, RAPD	1, 2, 5, 7, 9, 10, 12	Redona & Mackill, 1996a; 1996b
3	Coleoptile length	4	RFLP, RAPD	3, 6, 10, 11	Redona & Mackill, 1996a; 1996b
4	Mesocotyl length	10	RFLP, RAPD	1, 3, 5, 6, 7, 11	Redona & Mackill, 1996a; 1996b
B	**Plant height**				
1	(QPh2a, QPh3a, QPh8a, QPh9a	4	RFLP	3, 12	Li *et al*, 1995a
2	ph2, ph3, ph5, ph6, ph8	5	RFLP	2, 3, 5, 6, 8	Xiao *et al*, 1995
3	qph-3, qph-4, qph-11	3	-	-	Zou *et al*, 2000
C	**Heading date**				
1	QHd3a, QHd8a, QHD9a	3	RFLP	3, 6, 8	Li *et al*, 1995a
2	dth-3-1, dth-3-2, dth-4, dth-7, dth-8	5	RFLP	3, 4, 7, 8	Xiao *et al*, 1995
3	qHD-2, qHD-3, qHD-5, qHD-7	4	RFLP	-	Zou *et al*, 2000
4	Hd1, Hd2, Hd3	3	-	-	Lin et al, 2000
D	**Root traits**				
1	Total root number, penetrated root number, root penetration index, penetrated root thickness, penetrated root length	28	RFLP, AFLP	1, 2, 3, 4, 6, 7, 10, 11	Ali *et al*, 2000
2	Root penetration ability	7	RFLP, AFLP	-	Price *et al*, 2000
E	**Cell-membrane stability**	9	RFLP, AFLP, STMS	1, 3, 7, 8, 9, 11, 12	Tripathy *et al*, 2000
F	**Shoot regeneration**	2	RFLP	2, 4	Takeuchi *et al*, 2000

Table 1.6: Quantitative trait loci for grain characteristics in rice

Sr. No.	Quantitative trait loci (QTLs)	No. of QTLs	Marker system	Chromo-somal location	Reference
A	**Grain yield**				
1	Kernel weight (QKw1, QKw2a, QKw2b, QKw3a, QKw3b, QKw3c, QKw5, QKw6, QKw10)	9	RFLP	1, 2, 3, 5, 6, 10	Li *et al*, 1997
2	Grain number(QGn1, QGn3a, QGn3b, QGn4, QGn8, QGn10)	6	RFLP	-	Li *et al*, 1997
3	Grain weight per panicle (QGwp3, QGwp4, QGwp5, QGwp8)	4	RFLP	-	Li *et al*, 1997
4	Days to maturity (dtm4, dtm7, dtm8)	3	RFLP	4, 7, 8	Xiao *et al*, 1995
5	Panicle length (pl5, pl9)	2	RFLP	4, 8	Xiao *et al*, 1995
6	Panicles per plant (ppp4)	1	RFLP	4	Xiao *et al*, 1995
7	Spikelets per panicle (spp3, spp5)	2	RFLP	3, 5	Xiao *et al*, 1995
8	Grains per panicle (gpp3, gpp4, gpp5)	3	RFLP	3, 4, 5	Xiao *et al*, 1995
9	Percent seed set (pssr6, pssr7)	2	RFLP	6, 7	Xiao *et al*, 1995
10	1000-grain weight (gw3, gw4, gw5, gw7, gw8)	5	RFLP	3, 4, 5, 7, 8	Xiao *et al*, 1995
11	Spikelets per plant (sppl3, sppl5, sppl11)	3	RFLP	3, 5, 11	Xiao *et al*, 1995
12	Grains per plant (gppl3, gppl4, gppl5, gppl11)	4	RFLP	3, 4, 5, 11	Xiao *et al*, 1995
13	Grain yield (gy8, gy11)	2	RFLP	8, 11	Xiao *et al*, 1995
B	**Cooking and eating quality**				
	Gelatinization temperature, gel consistency, amylose content	1 (*waxy* locus)	RFLP	6	Tan *et al*, 1999
C	**Grain quality**				
1	Grain length	1 major & minor QTLs	RFLP	3	Tan *et al*, 2000
2	Grain width	1 major & minor QTLs	RFLP	5	Tan *et al*, 1999
3	White belly responsible for opacity of the endosperm	1	RFLP	5	Tan *et al*, 1999
C	**Paste viscosity characteristics**	20		6 (*waxy* locus)	Bao *et al*, 2000

Table 1.7: List of quantitative trait loci for biotic and abiotic stress in rice

Sr. No.	Quantitative trait loci (QTLs)	No. of QTLs	Marker system	Chromo-somal location	Reference
A	**Biotic stress**				
1	Bacterial blight	1	RFLP	11	Li *et al*, 1999
2	Bacterial leaf streak	11	-	-	Tang *et al*, 2000
3	Sheath blight				Zou *et al*, 2000
	(a) qSB-2, qSB-3, qSB-7, qSB-9-1, qSB-9-2, qSB-11	6	-	2, 3, 7, 9	
	(b) -	6	-	-	Li *et al*, 1995b
4	Blast				
	(a) Lesion number	10	RFLP	1, 3, 5, 6, 7, 8, 11, 12	Wang *et al*, 1994b
	(b) Diseased leaf area	7	RFLP	1, 3, 6, 8, 11	Wang *et al*, 1994b
	(c) Lesion size	2	RFLP	1, 11	Wang *et al*, 1994b
B	**Abiotic stress**				
1	Submergence tolerance sub-1	1	RFLP	9	Xu & Mackill, 1996
2	Aluminium tolerance	2	RFLP. AFLP	1, 12	Wu *et al*, 2000

therefore, be easily detected. When populations derived from crosses between one high-yielding parent and one low-yielding parent have been examined with molecular markers and the loci controlling yield identified, it has been observed that although the high yielding parent does contain a great number of positive alleles at loci associated with yield, there are almost always some loci for which the inferior parent contributes a superior allele (Xiao *et al*, 1996a). In order to discover and transfer such valuable QTLs from unadapted germplasm into elite breeding lines, Tanksley and Nelson (1996) have devised a strategy called advanced backcross QTL analysis. They have predicted that the modified elite lines thus obtained will perform even better than the original elite cultivar. Xiao *et al* (1998) have observed that 35 out of the 68 QTLs identified totally using RFLPs and microsatellites have a beneficial allele derived from the phenotypically inferior, wild parent, *O.rufipogon*, when the transgressive segregants of a cross between the wild parent and an elite hybrid rice were analyzed.

1.5 Physical mapping and map-based cloning or positional cloning of genes

Various factors such as availability of saturated molecular maps, small DNA content (Arumunagathan and Earle, 1991) and large percentage of low copy DNA sequences (Deshpande and Ranjekar, 1980; McCouch *et al*, 1988) make rice a model monocot for molecular genetic studies and map-based cloning of agronomically important genes. Map-based cloning or positional cloning, also known as reverse genetics, is a strategy which requires knowledge about the chromosomal location of the gene but does not require prior knowledge about the gene or its products (Young, 1990; Wicking and Williamson, 1991; Collins, 1992). Map based cloning and transposon tagging are now being employed to isolate genes corresponding to desirable traits of agronomic importance.

The first step in map-based cloning is the identification of markers tightly linked to the gene and on either side of the gene. The second prerequisite for map-based cloning is the construction of a physical map of the region encompassing the gene of interest, in which the distance between the markers is defined in terms of nucleotides rather than recombination

frequencies (i.e. cM distances) as in a genetic linkage map, followed by chromosome-walking. In rice, physical mapping has been done in regions spanning important genes such as BB resistance genes, *Xa*-21 (Ronald *et al*, 1992; Wang *et al*, 1995b), *Xa*-1 (Yoshimura *et al*, 1996) and *xa*-13 (Sanchez *et al*, 1999), stripe disease resistance locus, *stvb*-I (Saito *et al*, 2000) and *waxy* locus (Nagano *et al*, 2000). The availability of libraries containing large inserts of genomic DNA has made it possible to identify positive clones and order the overlapping clones to facilitate 'walking' on the gene of interest. Yeast artificial chromosome (YAC) libraries, the cloning capacity of which is 1000kb, have been constructed in rice (Umehara *et al*, 1995; Kurata *et al*, 1997) and used to produce physical maps of large regions. The rice YAC library contains an average DNA insert size of 350kb, which may be extremely valuable for constructing large DNA contigs over a particular genomic region or whole chromosomes. However, the large percentage (40%) of chimeric YAC clones (Umehara *et al*, 1995), difficulties in isolation of YAC insert DNA and low transformation efficiency have affected the construction of YAC libraries and hindered their utility. In order to obviate the difficulties associated with YAC libraries, bacterial artificial chromosome (BAC) libraries have been constructed, which can accept inserts upto 400kb (Wang *et al*, 1995b; Zhang *et al*, 1996b; Zhang and Wing, 1997). In BACs DNA can be cloned with high efficiency, manipulated easily and maintained stably (Shizuya *et al*, 1992), making BAC cloning less time consuming and less expensive for construction of complete libraries (Wang *et al*, 1995b). Fluorescent *in situ* hybridization (FISH) technique has also been used for physical mapping in case of the *Gm*-2 gene locus for gall midge resistance (Rajyashri *et al*, 1998).

The final step in map-based cloning is isolation of the gene corresponding to the target trait using a combination of genetic analysis and molecular tools. In rice, Song *et al* (1995), Wang *et al* (1995c), Yoshimura *et al* (1998) and Wang *et al* (1999) have isolated the *Xa*-21, *waxy*, *Xa*-1 and *Pi*-b genes, respectively, by positional cloning. These studies have revealed that *Xa*-21 is probably a transmembrane protein containing extracellular leucine-rich repeats (LRR) and a cytoplasmic kinase domain (Song *et al*, 1995), while *Xa*-1 and *Pi*-b genes belong to the nucleotide binding site (NBS)-LRR class of

plant disease resistance genes (Yoshimura *et al*, 1998; Wang *et al*, 1999) and *Xa*-1 is believed to encode a cytoplasmic protein (Yoshimura *et al*, 1998). The characterization of these resistance genes has facilitated a better understanding of plant disease resistance and has led to genetic engineering of lines containing *Xa*-21 gene in rice (Wang *et al*, 1996).

Sequence comparison of the cloned genes with other linked disease resistance genes on the same chromosome may lead to clues regarding evolution of plant disease resistance. A better knowledge of signal transduction pathway in plants will help to design new approaches for disease control. The intra and inter-genic transfer of disease resistance genes may provide an additional tool for breeders in combating plant disease. For example, the phytobacterial genus *Xanthomonas* infects virtually every crop species worldwide and, hence, engineering and transfer of BB resistance genes may help in reducing the losses resulting from *Xanthomonas* infection in recipient species.

1.6 Marker assisted selection (MAS) and gene pyramiding

Molecular markers offer great scope for improving the efficiency of conventional plant breeding by selecting markers linked to the trait rather than the trait itself. In case of disease resistance, availability of markers tightly linked to the resistance genes will help in identifying plants carrying these genes simultaneously, without subjecting them to pathogen or insect attack and with no limitation on the number of rounds of selection in a year.

MAS in plant breeding requires not only suitable markers closely linked to the known genes but also a simple, precise, rapid, efficient and economic approach to handle a large number of samples. Since RFLPs are very cumbersome and RAPDs are extremely sensitive to various reaction parameters, both these markers cannot directly be used for MAS. However, once these markers are converted into STS markers or SCARs as seen in the case of *Xa*-21 gene (Abenes et al, 1993; Huang et al, 1997), they can be readily used for MAS. Specific amplicon polymorphisms (SAP) or cleaved amplicon polymorphisms (CAPs), generated between resistant and susceptible genotypes when the STS markers obtained by sequencing the RFLP clones were found to be monomorphic, have been used for MAS of *Pi*-

2(t) gene conferring resistance to blast and *xa*-5 and *xa*-13 genes against BB (Hittalmani *et al*, 1994; Blair and McCouch, 1997; Huang *et al*, 1997). Hittalmani *et al* (1999; 2000) have used these SAP markers and RFLPs closely linked to the resistance genes in MAS to select agronomically superior lines into which the desired gene has been incorporated. In addition to SAP, STMS markers have also been found to be tightly linked to the xa-5 gene (Blair and McCouch, 1997) and can be used for MAS. SCAR markers tightly linked to the gall midge resistance genes, *Gm*-2 and *Gm*-4(t) (Nair *et al*, 1995a; 1995b; 1996) and RFLPs cosegregating with the semidwarfing gene, *sd*-1 (Cho *et al*, 1994) have facilitated their use in MAS. Nair *et al* (1995a) have also developed two allele-specific associated primers (ASAPs), one for each allele (resistant and susceptible), which could be used either individually or multiplexed in a single PCR reaction to discriminate between plants that are resistant and susceptible to gall midge biotype 1.

Another scope for MAS is in gene introgression studies where Chen *et al* (2000) for example, have used it to introgress *Xa*-21 into Minghui 63, a restorer line widely used for hybrid rice production in China, which has become increasingly susceptible to BB over a period of time. Using a PCR-based MAS system comprising a marker that is a part of *Xa*-21 and two markers located 0.8cM and 0.3cM on either side of *Xa*-21, an improved version of Minghui 63, which is exactly the same as the original except for a fragment of less than 3.8cM in length surrounding the *Xa*-21 locus, has been developed. Field examination has shown that the improved version of Minghui 63 shows significantly higher grain weight and spikelet fertility than the original genotype under heavily diseased conditions while they are identical when there is no disease stress. Rapid, nondestructive DNA isolation methods from leaf discs, leaf blades and half seeds have made it easier to implement MAS (Deragon and Landry, 1992; Wang *et al*, 1993; Zheng *et al*, 1995; Zhai *et al*, 1996).

Gene pyramiding refers to the combining of two or more major genes for resistance into a single plant genotype (Mundt, 1990). While the use of a single major gene limits the useful lifespan of resistant cultivars to a few years, gene pyramiding could delay resistance breakdown by conferring horizontal resistance effective against all prevalent pathotypes of the

pathogen. Yoshimura *et al* (1995a) and Huang *et al* (1997) have combined two to four BB resistance genes into a single line by MAS using RFLP, RAPD, SCAR and SAP markers. They have observed that pyramid lines show a wider spectrum and higher level of resistance than their single gene parental lines, through both ordinary gene action and quantitative complementation. Hittalmani *et al* (2000) have also observed a similar phenomenon when they carried out marker-assisted pyramiding of three major blast resistance genes. The gene-pyramided lines can be used to conduct quantitative analysis on the effect of interaction with other genes.

1.7 Comparative mapping or synteny mapping in cereals

Rice, a monocotyledonous plant with a genome size of $0.45X10^9$ bp (Deshpande and Ranjekar, 1980; Arumunagathan and Earle, 1991), is the smallest genome among cereals. The smaller genome size of rice results in a higher gene density, making it an attractive target to study cereal genome discovery and cereal genome analysis (Goff, 1999). Thus, rice is a model system for the study of fundamental aspects of plant biology such as yield, hybrid vigor; and single and multigenic disease resistance. The presence of several genotypes of rice, which are adapted to a wide variety of environmental situations from tropical flooding to temperate drylands (Fig. 1.3), makes it a model to study real life adaptive responses. On the basis of the conserved units present between cereal genomes and the advantages offered by the rice genome, it has been proposed that rice would become a pivotal genome for the analysis of cereal genomes (Moore, 1995).

Comparative mapping basically involves the use of a common set of molecular markers to map the genomes of different species. Based on RFLPs, Ahn and Tanksley (1993) first observed that there was a conservation of linkage groups between rice and maize, which together accounted for more than two-thirds of both the genomes. Later on, several workers observed a colinearity between the genomes of rice, wheat, maize, barley and oat (Ahn *et al*, 1993; Kurata *et al*, 1994a; Van Deynze *et al*, 1995; Saghai Maroof *et al*, 1996; Devos and Gale, 1997) despite gross differences in the genome size of these species. Dunford *et al* (1995) and Kilian *et al* (1995) have

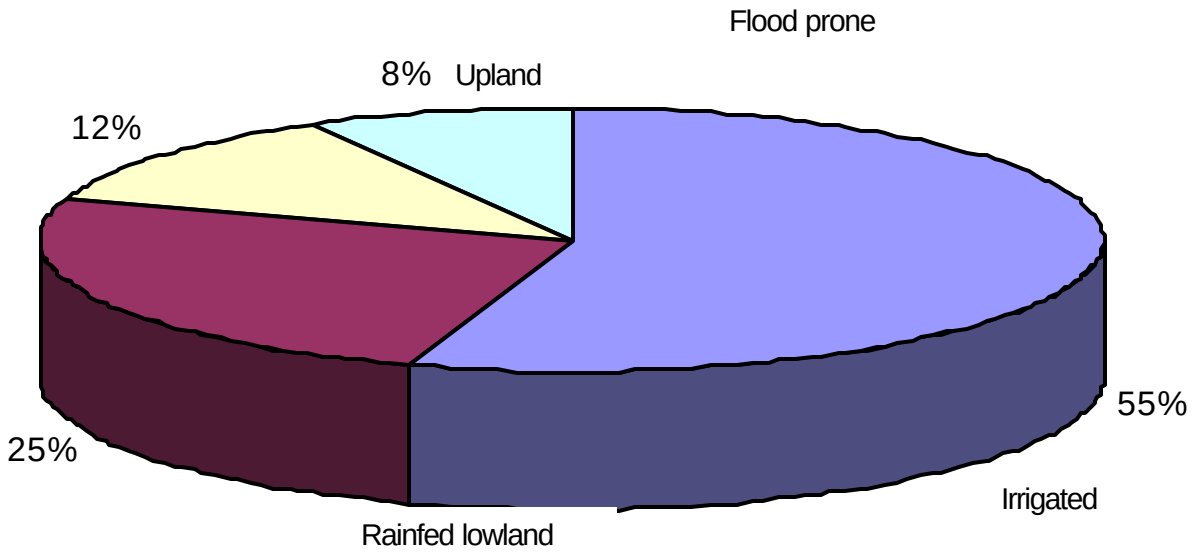


Figure 1.3: Distribution of world rice area in different ecologies (Khush, 1997)

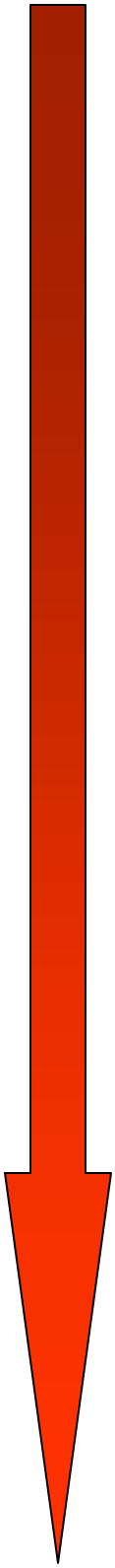
demonstrated that colinearity exists even at the physical level, at least for the regions studied between rice and Triticeae genomes, based on their analysis of YAC clones. Several genes and QTLs have also been found to correspond closely in cereals (Paterson *et al*, 1995; Chen *et al*, 1997b; 1998). The loci controlling heading date in hexaploid oat (chromosome 5) and rice (chromosome 3) are homologous (Causse *et al*, 1994) while those controlling heading date in hexaploid oat (chromosome 17), vernalization gene *Vrn*5 in wheat (chromosome 7A) and the photoperiod response gene, *Se*1 in rice (chromosome 6) are homologous. Even the positions of the centromeres correspond closely in rice and wheat (Singh *et al*, 1996). Recently, Ishii and McCouch (2000) have observed a high level of microsynteny at microsatellite loci in the chloroplast genomes of rice and maize. Conservation of synteny relationships facilitates bridging of genetic information between taxa. The molecular markers produced in one species can be used to accelerate the development of genetic linkage maps in species with less characterized genetic maps. Also, DNA markers spanning conserved regions can serve as heterologous probes to saturate regions encompassing genes of interest, which will ultimately lead to tagging the particular gene of interest. The microsynteny between rice and barley has been effectively used to saturate the region encompassing the *Rpg*-1 locus in barley, using rice clones, and one of the clones has been observed to map 0.3cM from the *Rpg*-1 locus (Kilian *et al*, 1995). The *Rpg*-1 gene confers stable resistance in barley to the wheat stem rust pathogen, *Puccinia graminis* f. sp. *tritici*. When rice YAC, BAC and cosmid clones have been used to isolate probes mapping to the *Rpg*-1 region in barley, it has been observed that three probes at the distal end of one of the BAC clones are out of synteny with rice (Kilian *et al*, 1997). Also, the barley *Rpg*-1 homologue has not been identified in the rice BAC, although a gene encoding a hypothetical polypeptide with similarity to a membrane protein is present (Han *et al*, 1999). These results indicate that even in regions of high microsynteny, there may be small DNA fragments that have transposed and are no longer in syntenous position. Since disease resistance genes evolve quickly, it may not always be possible to use colinearity among disease resistance genes for further studies.

Synteny relationships can help to predict the positions of the orthologous genes of agronomic importance in related species (Jena, *et al*, 1994; Harrington *et al*, 1997). If related species show conserved linkages in regions of shared orthologous loci, such genes can be characterized more efficiently and cloned in those species that have well-developed genetic linkage maps and small and less complex genomes. Thus, the greatest advantage of the conservation of gene order and composition lies in using grasses with smaller genomes such as rice and *Sorghum* for chromosome walking and cross-genome map-based cloning of genes. However, in spite of the colinearity at the genetic map level, few studies have shown micro-colinearity at the level of a few hundred kilobases. Studies on micro-colinearity would tell us which genomes would be suitable for map-based cloning of genes from large genomes. This will allow a gene to be located in the smallest possible number of steps, even if the locus was originally identified in a species such as wheat or maize, in which large genome size and uneven distribution of recombination along chromosomes hampers the isolation of genes. Examination of the comparative maps of barley and rice have revealed the conservation of several genomic intervals containing barley disease resistance genes (Saghai Maroof *et al*, 1996). Markers flanking these intervals could provide a starting point for the characterization and positional cloning of these genes in the smaller and less complex rice genome. Comparison of locus order and distribution of recombination events in these studies may be useful to considerably improve the gene pool of the crop (Devos *et al*, 1993).

The completion of rice genome sequencing (Sasaki and Burr, 2000) will aid in using the compact rice genome to isolate agronomically important genes from large cereal genomes.

Thus, in summary, rice is one of the most important crops in the world with about 50% of the world's population depending on it for food. It is also one of the best studied cereals, where molecular markers have played an important role in its genome analysis and structure-function relationships (Fig. 1.4). The use of these markers will help in reducing the time span required for developing new and improved varieties and will be of tremendous help to rice breeders in crop improvement programs. Also, the small size of the rice

genome and its colinearity with other cereal genomes make it a model plant for study of cereal genomics. Several reviews are available summarizing the information about molecular markers for specific individual applications involving mapping, marker-assisted selection, gene pyramiding, microsatellites, synteny relationships and so on. Here, we have made a concerted effort to compile all the available information related to various aspects of genome research in rice for the first time. The review gives an up to date account of the application of molecular markers in six different areas viz. genetic diversity analysis, phylogenetic and evolutionary studies, mapping and tagging genes and QTLs, marker-assisted selection and gene pyramiding, physical mapping and map-based cloning and comparative mapping. Fig. 1.4 lists the milestones in rice genome research using molecular markers. Thus, our review is possibly a unique effort summarizing DNA marker research in rice and will be of great help not only to researchers but also to rice breeders in exploiting the potential of DNA marker strategy in their future work.



1988	Dallas	DNA fingerprinting for the first time in rice. Parentage analysis performed for the first time in plants
1988	McCouch *et al*	First molecular map of rice based on RFLPs
1991	Yu *et al*	Gene tagging for the first time in rice of blast resistance gene
1991	Williams *et al*	**PCR-based RFLP performed for the first time**
1992	Jena *et al*	Use of RFLP to study gene introgression from wilds
1993b	Zhao and Kochert	First microsatellite mapped - $(GGC)_n$
1993	Abenes *et al*	Marker-assisted selection carried out for bacterial blight resistance gene, *Xa*-21
1993	Ahn and Tanksley	Synteny between cereal genomes observed for the first time
1994	Kawase	A new technique called Restriction landmark genome scanning (RLGS) developed
1994b	Kurata *et al*	Densest molecular map ever developed in plants, containing 1383 DNA markers. First significant gene expression map in plants
1994	Gupta *et al*	Transcriptional activity of microsatellites reported in plants for the first time using oligonucleotide $(CAC)_5$
1995	Song *et al*	Map-based cloning of the first gene from cereals viz. *Xa*-21
1995a	Yoshimura *et al*	Gene pyramiding carried out for the first time. Multiple genes conferring resistance to bacterial blight introduced into a single cultivar
1996	Tanksley and Nelson	Advanced backcross QTL analysis
2000	Sasaki and Burr	Rice genome sequencing nearly completed

FIGURE 1.4: Milestones in rice genome research using molecular markers

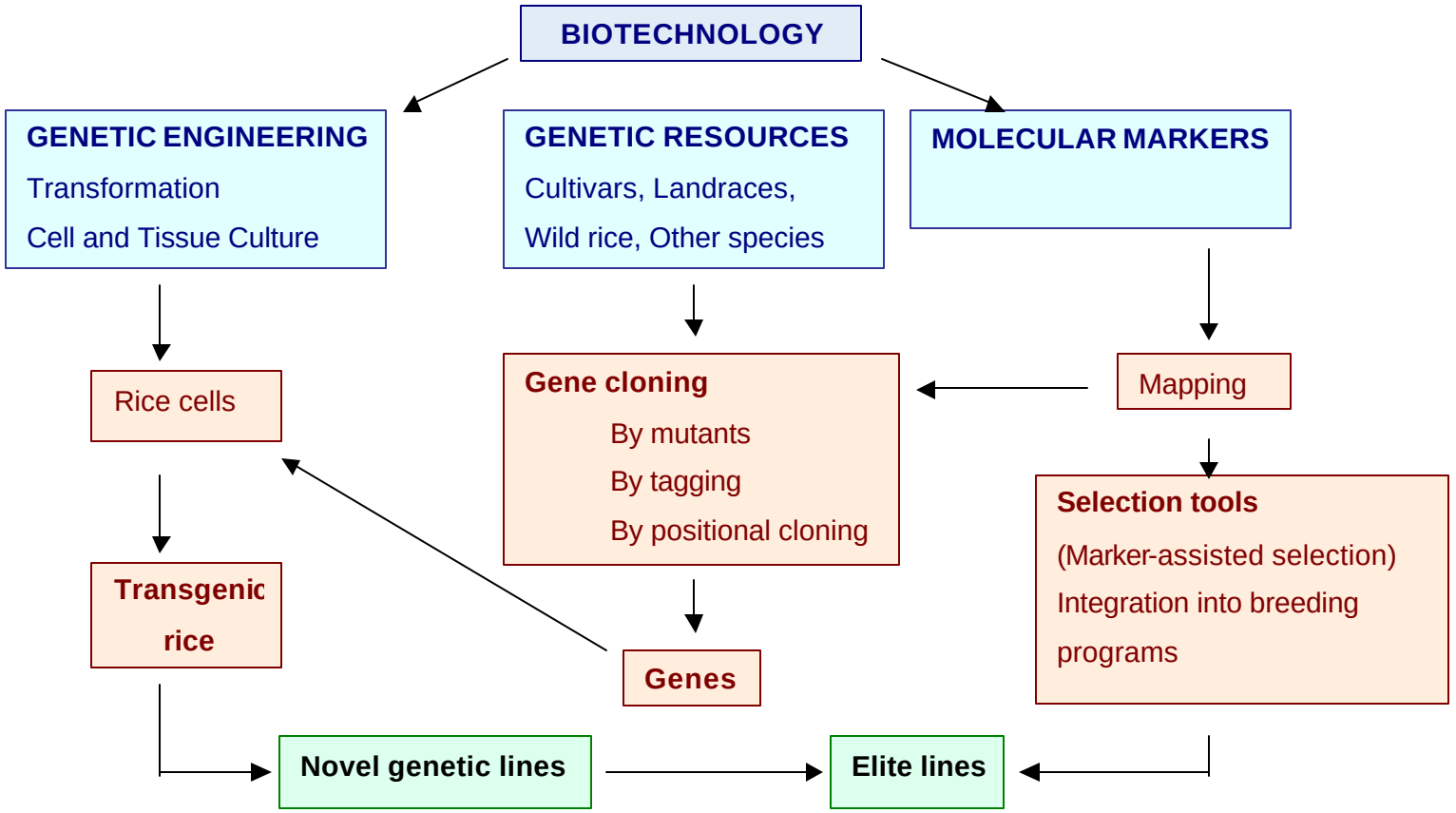


Figure 1.2: Perspective process of new breeding: applying molecular biological and genetic engineering tools for developing new and novel lines (Ikeda and Wasaka, 1997)

CHAPTER 2

Estimation of Genetic Diversity in Indian Elite Rice Varieties Using Molecular Markers

The contents of this chapter have been accepted as a full-length paper in Genetica (2001) (In Press)

ABSTRACT

Genetic diversity, an important criterion for selection of parents for conventional breeding and hybrid program was evaluated among 42 Indian elite rice varieties using inter-simple sequence repeat (ISSR) and sequence tagged microsatellite site (STMS) markers. The data thus obtained was compared with that generated in an earlier study using random amplified polymorphic DNA (RAPD) markers. RAPD, ISSR and STMS markers resulted in mean heterozygosity values of 0.429, 0.675 and 0.882 over all loci, respectively with marker index values of 2.21, 4.05 and 5.49, respectively. Totally 153 bands (91%) were polymorphic out of 168 bands amplified, and the average genetic similarity coefficient across all the 861 cultivar pairs was 0.70 while the average coefficient of parentage was 0.10. Cluster analysis revealed that there was very poor correlation (correlation coefficient < 0.1) between dendrograms generated using coefficients of parentage and molecular marker generated genetic similarities. The three molecular marker systems together provided a wider genome coverage and, therefore, would be a better indicator of the genetic relationships among these elite rice cultivars than those revealed using individual molecular marker systems.

2.1 INTRODUCTION

In India, rice is the most important crop, grown widely in diverse ecological niches with a wide range of selected cultivars. In India, most rice varieties developed by rice research stations are regionally adapted and are high yielding leading to self sufficiency. However, rice yields have plateaued during the last two and a half decades (Viramani, 1994), which may have resulted from the narrow genetic base of released varieties (Carmona, 1990). Similar situations have also been reported from Japan (Kaneda, 1985), United States (Dilday, 1990), Taiwan (Lin, 1991), Latin America (Cuevas-Perez *et al*, 1992) and Australia (Ko *et al*, 1994), documenting a widespread reduction in genetic diversity of modern rice cultivars due to intensive breeding efforts. Crosses between genetically diverse parents are, therefore, important in hybridization programs to increase heterosis, maximize heterozygosity and maintain high levels of genetic variability in the progeny (Messmer *et al*, 1993), leading to increase in grain yield.

Several approaches have been used to estimate genetic diversity to select appropriate parental genotypes in crossing programs. A commonly used measure of genetic similarity is coefficient of parentage (CP) or coancestry, which is defined as the probability that a random allele of one individual is identical by descent to a random allele of another at the same locus (Malecot, 1948). In many autogamous species where pedigrees of cultivars can be traced, the coefficient of parentage has been determined (Murphy *et al*, 1986; Knauft and Gorbet, 1989; Cuevas-Perez *et al*, 1992;) and used to study genetic diversity within species (Carter, *et al*, 1993; Melchinger *et al*, 1994). Measurement of morphological and biochemical characteristics is another commonly used method to arrive at an estimate of genetic diversity in parental stock material (Tateoka, 1962; Morinaga, 1964; Second, 1982). During the last decade, PCR based markers, such as random amplified polymorphic DNA (RAPD) (Welsh and McClelland, 1990; Williams *et al*, 1990; Hu and Quiros, 1991), inter-simple sequence repeats (ISSR) (Zietkiewicz *et al*, 1994; Goodwin *et al*, 1997), simple sequence length polymorphism (SSLP) (Wu and Tanksley, 1993; Morgante *et al*, 1994; Plaschke *et al*, 1995), and amplified fragment length polymorphism (AFLP) (Vos *et al*, 1995), have been extensively used in genetic diversity analysis of crops since they are

technically simple to use, time saving, highly informative, need no information of template DNA sequence or synthesis of specific primers except for simple sequence repeats (SSRs) and they require small amounts of DNA (Parsons *et al*, 1997; Kojima *et al*, 1998).

SSR or SSLP or STMS (sequence-tagged microsatellite site) markers have been developed based on DNA sequence variation, and the STMS primers are designed flanking the repetitive regions which are found ubiquitously in eukaryotic genomes. The variation in PCR product length, referred to as alleles is thus a function of the number of SSR motifs. STMS markers are easily transferable among different laboratories by exchange of primer sequences and are co-dominant, highly reproducible, easily scorable (Powell *et al*, 1996a) and readily available in rice (McCouch *et al* 1997; Chen *et al*, 1997a; Cho *et al*, 2000; Temnykh *et al*, 2000). ISSR analysis also relies on the ubiquity of SSRs in eukaryotic genomes (Langercrantz *et al*, 1993) and enjoys the specificity of STS markers but requires no prior sequence information for primer synthesis. Here, the primer, which is terminally anchored at 5' or 3' end, amplifies the unique sequence between two inversely oriented, closely spaced SSRs of the same type, at several loci simultaneously, yielding a multilocus marker system useful for diversity analysis, DNA fingerprinting and genome mapping (Zietkiewicz *et al*, 1994).

Rice is the most well studied crop where PCR based markers have given valuable information about its genetic diversity as discussed in section 1.2.1 of chapter 1. Our laboratory has earlier shown the utility of micro- and minisatellite DNA sequences in fingerprinting of indica, japonica and wild *Oryza* species (Gupta *et al*, 1994; Ramakrishna *et al*, 1994; 1995). In addition, the genetic variability in rice germplasm at a specific locus, *knotted*-1 homeobox locus (Deshpande *et al*, 1998) has been studied. Recently, ISSRs have also been used in our laboratory to study genetic diversity and phylogenetic relationship in the genus *Oryza* (Joshi *et al*, 2000). However, except for these studies, there are no reports on the evaluation of genetic variability in Indian rice genotypes using molecular markers. Since the knowledge of genetic relationships among Indian elite rice cultivars is important for selection of parents both for conventional breeding and hybrid production, I undertook the present work to quantify the genetic diversity in a

large sample of elite rice cultivars developed in India, from 1968 to 1994, by using SSLP and ISSR markers, and to correlate the diversity data obtained with these two types of markers with that obtained earlier using RAPD markers and also with the coefficient of parentage.

2.2 MATERIALS AND METHODS

2.2.1 Plant material and DNA extraction

The seeds of 42 elite rice cultivars used in the study (Table 2.1) were obtained from the Directorate of Rice Research, Hyderabad and the Rice Research Station, Karjat, India. Seedlings were grown in a glass-house for 3 weeks after which leaf tissue was harvested. The leaf tissue of ten seedlings was bulked together, frozen in liquid nitrogen, stored at -70^{o}C and used later for DNA extraction. Five grams of this frozen bulked leaf tissue was ground to a fine powder in liquid nitrogen in a mortar and pestle and used for DNA isolation by hexadecyl-trimethyl-ammonium bromide (CTAB) method (Rogers and Bendich, 1988). 15ml of 2X CTAB extraction buffer [2% CTAB, 100mM Tris-HCl (pH 8.0), 20mM Na-EDTA (pH 8.0), 1.4% NaCl and 1% polyvinyl pyrrolidone] were added per 5g of frozen leaf tissue and incubated at 60^{o}C for one hour. Equal volume of chloroform : isoamyl alcohol (24:1) mixture was added and mixed thoroughly to form an emulsion, which was centrifuged at 10,000 rpm for 10min in SS34 rotor. To the supernatant, equal volume of CTAB precipitation buffer [1% CTAB, 50mM Tris-HCl (pH 8.0) and 10mM EDTA (pH 8.0)] was added, mixed gently and centrifuged as above. The DNA pellet was dissolved in high salt TE buffer [1mM NaCl, 10mM Tris-HCl (pH 8.0) and 1mM EDTA (pH 8.0)] and was precipitated with two and a half volumes of absolute ethanol. The DNA precipitate was washed with 70% ethanol, centrifuged and redissolved in TE buffer [10mM Tris-HCl (pH 8.0) and 1mM EDTA (pH 8.0)]. For removal of RNA, the DNA was incubated with RNase A (DNase free) at 37^{o}C for 2h. The DNA concentration was estimated using agarose gel electrophoresis by comparing with known concentrations of λ DNA.

Table 2.1: Indian elite rice cultivars used for analysis

Sr. No.	Name	Cross combination	Year of release	Days to 50% flowering	Grain type	Production system*
1	Karjat-2	RPW-6-17/RP-4-14	1994	105	LS	IRM
2	Sugandha	Prabhavati/IET8573	1994	85	LS	--
3	Jallahari	Pankaj/Mahsuri/TKM6	1993	150	MS	SDW
4	Jalnidhi	Selection of Goanth	1993	150	LB	DW
5	Jalpriya	Selection IET4060/Jalmagna	1993	130	LS	DW
6	Varsha	IR50/Mahsuri	1993	90	MS	IRE
7	Narmada	T(N)1/Basmati370	1991	110	LS	IRM
8	Vibhava	CR44/W12708	1989	105	LS	IRM
9	Chandana	Sona/Manoharsali	1989	100	LS	IRM
10	Salivahana	RP5-32/Pankaj	1988	128	SB	RSL
11	Pothana	IR579/W12708	1988	95	LS	IRE
12	Lalat	OBS677/IR2071/Vikram/W1263	1988	95	LS	IRM
13	Annada	Kumar/CR57-49	1988	85	MS	IRE
14	Indrayani	Ambemohr-157/IR8	1987	102	LS	IRM
15	Prasanna	IRAT8/N22	1986	65	LS	RUR
16	Rambha	Pankaj/W-1263	1985	120	MB	SDW
17	Gauri	T-90/IR-8/Vikram	1984	105	MS	IRM
18	Daya	Kumar/CR57-49	1984	105	MS	IRM
19	Shankar	Parijat/IET3225	1983	55	LB	RUR
20	Mandyavani	Cr1014/IR8(Sec.selection)	1982	100	MS	IRM
21	Subhadra	T(N)1/SR26B	1980	60	LB	RUR
22	Phalguna	IR-8/Siam29	1977	115	LS	RSL
23	Prakash	T90/IR8	1977	105	LS	IRM
24	Rasi	T(N)1/Co-29	1977	85	MS	IRE
25	Akashi	IR8/N-22	1977	80	SB	RUR
26	Bhavani	Peta/BPI76	1976	105	LS	IRM
27	Surekha	IR8/Siam29	1976	104	LS	IRM
28	Parijat	TKM6/T(N)1	1976	70	MS	RUR
29	Intan	Introd. From Indonesia	1975	135	LS	RSL
30	Mangala	Jaya/S317	1975	80	MS	IRE
31	Hema	T-141/IR-8-246	1974	105	LB	IRM
32	Rajeshwari	T-19/IR-8	1974	100	SB	IRM
33	Vani	IR8/CR1014	1973	105	LS	IRM
34	Tellahamsa	HR12/T(N)1	1971	90	MS	IRE
35	Jalmagna	Selection from Badhan	1969	130	SB	SDW
36	Madhukar	Selection from Gonda	1969	120	MS	SDW
37	Jaya	T(N)1/T141	1968	100	LB	IRM
38	Karjat-1	Halmaldiga/IR36	--	110	--	--
39	Karjat-35-3	HR-19/IR8	--	90	--	--
40	Karjat-184	T(N)1/Kolamba540	--	100	--	--
41	Ratnagiri-1	IR8/RTN24	--	110	--	--
42	Ratnagiri-2	RTN-60-8-1/Warangal	--	140	–	--

*Footnote: IRE [Irrigated Rice Early duration (120 to 135 days to maturity)], IRM [Irrigated Rice Medium duration (135 to 150 days to maturity)], IRAK (Irrigated Rice Alkaline tolerant variety), RUR [Rainfed Upland Rice (grown 900 metres above mean sea level (Arunachalam, 2000))], RSL [Rainfed Shallow Water or Low land (grown in flat plains (Arunachalam, 2000))], SDW (Semi Deep Water) and DW (Deep Water).

2.2.2 ISSR analysis

All the 3'-anchored microsatellite primers used for ISSR analysis were obtained from University of British Columbia, Canada. Each reaction was carried out in a 25µl volume containing 20ng template DNA, 100µM of each dNTP (U.S. Biochemical Corp., U.K.), 0.2µM of 3' anchored microsatellite primer (University of British Columbia, Canada), 0.5mM spermidine, 2% formamide, 0.8U Taq DNA polymerase (Perkin Elmer, U.S.A.) in 10mM Tris-HCl (pH 8.3), 1.5mM $MgCl_2$ and 50mM KCl according to *Joshi et al*, (2000). PCR was performed in a Perkin Elmer GeneAmp^{TM} PCR System 9600 (Perkin Elmer Inc., U.S.A.) with an initial denaturation at 94°C for 5 min, followed by 45 cycles of denaturation at 94°C for 1 min, primer annealing at 50°C for 45 sec and primer extension at 72°C for 2 min with a final extension at 72°C for 5 min. The PCR amplified products were resolved on 1.2% agarose gels in TAE buffer [40mM Tris-acetate, 1mM EDTA (pH 8.0)], stained with ethidium bromide and the fluorescence was viewed and photographed by using uv light.

2.2.3 STMS analysis

The STMS primers used for analysis were only those that amplified dinucleotide repeats, $(GA)_n$ and $(CT)_n$ from chromosomes 3, 4, 5, 6, 7, and 8 (Table 2.2) (Wu and Tanksley, 1993; Chen *et al*, 1997a). STMS reactions were performed according to the method described by Wu and Tanksley (1993). Each 10µl reaction consisted of 1µCi of α^{32}P dCTP, 25µM dCTP, 200µM each of dATP, dGTP, and dTTP (U.S. Biochemical Corp., U.K.), 5-10pmoles of each primer, 500µM of spermidine, 20ng of DNA, 0.24U of Taq DNA polymerase (Perkin Elmer Inc., U.S.A.), in 10mM Tris-HCl (pH 8.3), 1.5mm $MgCl_2$ and 50mM KCl. Reactions were amplified in a thermal cycler (DNA Engine; M.J. Research Inc., U.S.A.) programmed with an initial denaturation at 94°C for 5 min, followed by 35 cycles of denaturation at 94°C for 1 min, primer annealing at 55°C for 2 min and primer extension at 72°C for 1.5 min, with a final extension at 72°C for 5 min. After amplification, half volume of stop solution [98% deionized formamide, 10mM EDTA, 0.025% Bromophenol blue and 0.025% Xylene Cyanol] was added to individual PCR

reactions. Samples were denatured at 70-80^{o}C for 2-3 min and then electrophoresed on 6% denaturing polyacrylamide gels containing 7M urea in 0.5X TBE [22.5mM Tris-borate, 0.5mM EDTA (pH 8.0)] buffer at 1500 V for 2 hours and visualized by autoradiography. The gels were exposed to X-ray films at –70^{o}C for 12-24 hours (depending on the signal) with intensifying screens. Sequencing reactions of bacteriophage cloning vector, M13 mp18 single stranded DNA (Sequenase version 2.0 DNA sequencing kit, U.S. Biochemical Corp., U.K.) were used as molecular weight standards.

2.2.4 Statistical analysis

The amplified DNA fragments of each cultivar were scored as present (1) or absent (0) and the data was entered into a binary matrix for subsequent analysis using NTSYS-pc version 1.8 (Rohlf, 1993). Coefficients of similarity were calculated by using Dice coefficient (Sokal and Sneath, 1963) by SIMQUAL function and cluster analysis was performed by agglomerative technique using the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) method (Sneath and Sokal, 1973) by the SAHN clustering function of NTSYS-pc. Relationships between the elite lines were portrayed graphically as dendrograms.

Pedigrees of each of the 42 elite lines and their known ancestors were obtained from cultivar descriptions and breeding records. The coefficient of parentage was calculated as outlined by Kempthorne (1969) making the following assumptions as described by Cox *et al*. (1985): (i) each parent contributes equally in the cross, (ii) all cultivars, parental lines and known ancestors were homogeneous and self-pollinated to homozygosity without selection, (iii) ancestors without known pedigrees were unrelated and, therefore, coefficient of parentage between these ancestral lines was presumed to be zero and (iv) the value of coefficient of parentage between a cultivar or ancestor and a direct selection from that cultivar or ancestor was 0.75. The relatedness of every pair of lines was calculated as the sum of the products of all the parentage coefficients of the shared parents of the lines. A dendrogram was constructed based on a matrix of the coefficients of parentage by using the SAHN clustering function of NTSYS-pc as described above.

RAPD data generated earlier (Chowdari, 1998) was used for comparison with the data obtained using ISSR and STMS markers. Pearson's product moment correlation coefficients (Smouse *et al*, 1986) were calculated for the matrix generated using pedigree data and genetic similarity matrices based on RAPD, ISSR and STMS data individually, as well as the similarity matrix obtained using a combination of all the three molecular marker systems.

The definitions of various terms, which were used for calculations and statistical analysis are given below:

The expected heterozygosity, H_n for a genetic marker:

$H_n = 1 - \Sigma p_i^2$, where p_i is the allele frequency of the i^{th} allele (Nei, 1973).

The arithmetic mean heterozygosity, H_{av} for each marker class:

$H_{av} = \Sigma H_n/n$, where n is the number of markers or loci analyzed (Powell *et al*., 1996b).

The average heterozygosity for polymorphic markers $(H_{av})_p$:

$(H_{av})_p = \Sigma H_n/n_p$, where n_p is the number of polymorphic markers or loci (Powell *et al*., 1996b).

Marker Index (MI):

$MI = E(H_{av})_p$, where E is the effective multiplex ratio. ($E = n\beta$ where β is the fraction of polymorphic markers or loci) (Powell *et al*., 1996b).

The probability (PI) that two elite varieties would exhibit identical DNA fragment profiles:

$PI = (X_D)^n$, where X_D represents the average genetic similarity index for all pairwise comparisons and n is the mean number of fragments obtained per genotype (Wetton *et al*., 1987).

2.3 RESULTS

2.3.1 ISSR analysis

Only those ISSR primers which were earlier reported to be useful in rice and produced good amplification and polymorphic patterns (Joshi *et al*, 2000) were used in the present analysis. Twelve 3'-anchored dinucleotide primers, containing $(AG)_n$, $(GA)_n$, $(CT)_n$, $(CA)_n$, and $(GT)_n$ repeats (Table 2.2), when used to amplify 42 Indian elite rice cultivars, resulted in 72 amplicons,

Table 2.2: Comparison of DNA marker systems

Sr. No.	Parameters for comparison	RAPD	ISSR	STMS
1	Total number of primers tested	40	14	14
2	Total number of polymorphic primers used for final analysis	8	12	9
3	Polymorphic primers used	OPA 3, 7, 12, 13, 17 & 18, OPB 7 & 18	UBC 807, 808, 809, 810, 811, 812, 814, 834, 835, 836, 847 & 850	RM 25, 30, 38, 122, 214, 227, 241, 249 & 252
4	Total number of bands amplified	40	72	56
5	Average number of bands per primer	5	6	6.2
6	Maximum number of bands amplified by a single primer	13	9	9
7	Minimum number of bands amplified by a single primer	2	2	3
8	Number of polymorphic bands identified	32	66	55
9	Percentage of total bands that were polymorphic	80	91.6	98.2
10	Average number of polymorphic bands per primer	4	5.5	6.1
11	Maximum number of polymorphic bands amplified by a single primer	11	9	9
12	Minimum number of polymorphic bands amplified by a single primer	1	1	3
13	Genetic similarity coefficient for all possible pairs of genotypes			
	(A) Maximum	0.95	0.94	0.96
	(B) Minimum	0.59	0.58	0.08
	(C) Average	0.80	0.78	0.37
14	Probability of obta ining an identical match by chance	$2.20X10^{-3}$	$6.40X10^{-4}$	$3.46X10^{-5}$
15	Arithmetic mean heterozygosity (H_{av}) at			
	(A) All loci	0.43	0.68	0.88
	(B) Polymorphic loci	0.54	0.74	0.90
16	Marker Index	2.21	4.05	5.49

66 (91.6%) of which were polymorphic, yielding an average of 5.5 polymorphic bands per primer (Table 2.2). The primers UBC807 and UBC808 produced nine bands each, all of which were polymorphic. However, five and three bands produced by these two primers, respectively, were either unique to one cultivar or absent only in one cultivar, and were, therefore, considered to be poorly informative by the definition of Prevost and Wilkinson (1999). UBC811, on the other hand, resulted in amplification of only seven bands, all of which were polymorphic, with one band present in 57% of the cultivars, three bands present or absent in 20-30% of the cultivars and one band present in 14% of the cultivars, thus suggesting its informative potential.

During the genetic similarity calculations the maximum similarity index obtained was 0.94 between Ratnagiri-1 and Ratnagiri-2, both of which are from Ratnagiri district in Maharashtra state. The minimum similarity index value of 0.58 was between Tellahamsa, an irrigated rice cultivar of Andhra Pradesh and Jalpriya, deep-water rice of Bihar and Uttar Pradesh. The average genetic similarity coefficient was 0.78 (Table 2.2), which was very high, indicating high level of genetic similarity detected by ISSRs between the elite cultivars. The arithmetic mean heterozygosity using ISSRs was 0.68 across all loci while the marker index was 4.05, indicating that these markers are more polymorphic than RAPD markers (Table 2.2) on this set of Indian elite rice cultivars.

The dendrogram derived from the similarity coefficients consisted of seven groups with one major cluster containing 16 cultivars and six minor clusters containing two to five cultivars each (Fig. 2.1). The deepwater rice, Jalpriya, and its parent Jalmagna, formed a monophyletic group which was distinct from the main clusters. Salivahana, Rambha and Jallahari, having common parent, Pankaj grouped together in cluster 1. Some of the irrigated rices of medium duration (with 135-150 days to maturity) such as, Hema, Ratnagiri-1, Ratnagiri-2, Rajeshwari, Vibhava, Mandyavani, Jaya, Indrayani, and Surekha, also grouped together in cluster 1 in the dendrogram. Clusters 2 and 3, on the other hand, contained rainfed upland rice cultivars [Upland cultivars are 900m above mean sea level (Arunachalam, 2000)], such as Subhadra, Shankar, Akashi and Parijat, which are all rice cultivars of early duration maturity (120-135 days to maturity). With the exception of Karjat-

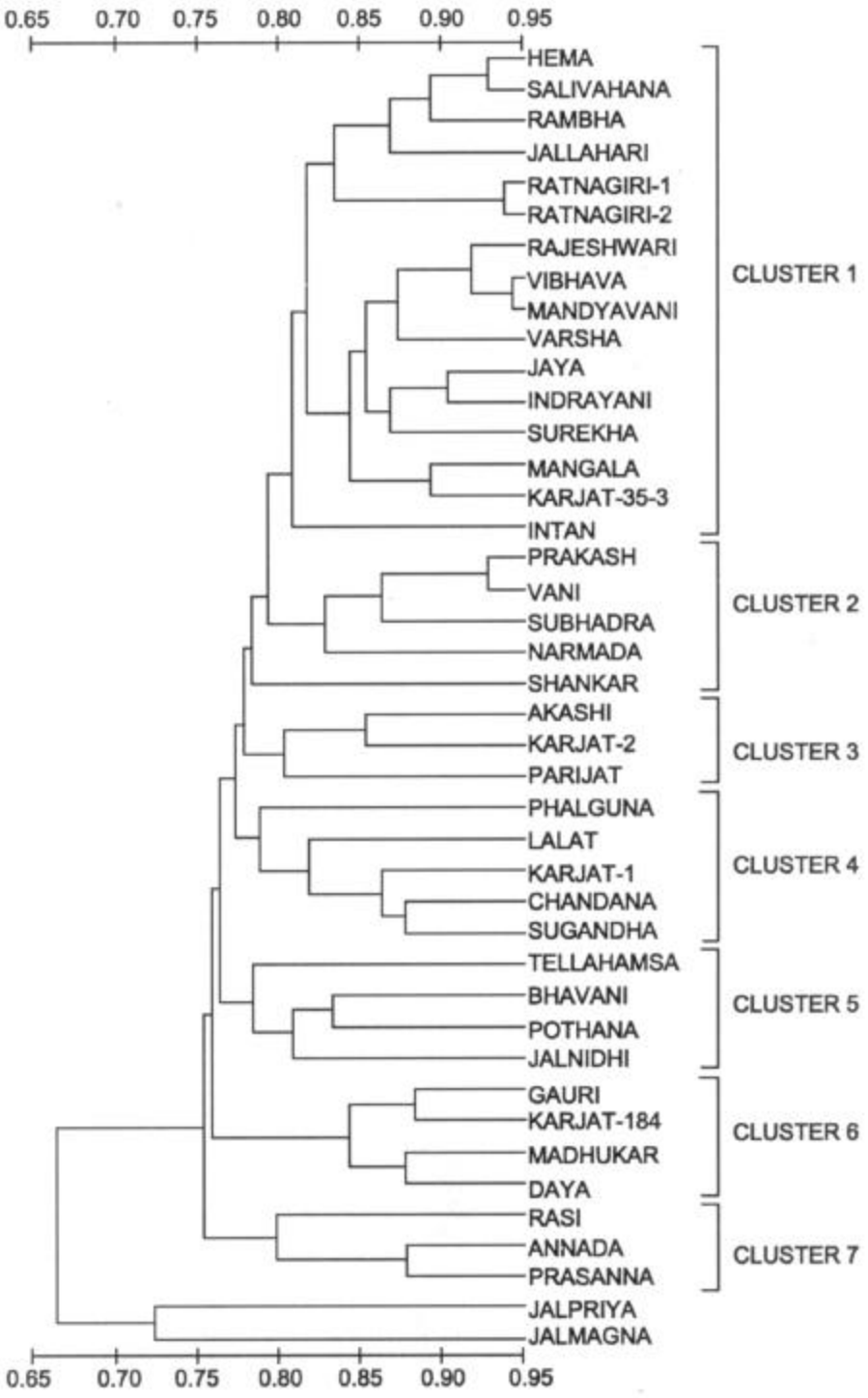


Figure 2.1: Dendrogram showing the genetic relationship among 42 Indian elite rice cultivars using ISSR markers

184, the remaining three cultivars in cluster 6, such as Gauri, Madhukar and Daya are all rices of medium duration maturity. All the three rice cultivars in cluster 7, such as Rasi, Annada and Prasanna have early duration maturity. The probability of obtaining an identical match by chance using ISSR data was found to be $6.40X10^{-4}$.

2.3.2 STMS analysis

Initially 12 STMS markers representing different chromosomes of rice were used to survey four elite rice cultivars, Rasi, Madhukar, Karjat-1 and Ratnagiri-2. However, three of these markers (RM39, RM168 and RM255) resulted in faint and blurred amplification patterns and were, therefore, not used for further analysis. By using the remaining nine STMS markers, 56 amplified products were obtained, of which only one was monomorphic. At 6.1 polymorphic bands per STMS marker, in the elite rice cultivars analysed, STMS markers generated the highest amount of polymorphism (98.2%) of the three marker types tested (Table 2.2). Two STMS markers RM30 and RM241 produced the highest number of bands at nine each, among all the STMS markers, while RM38 yielded the least number (three) of bands. Only one of the bands produced by RM30 flanking primers was monomorphic, while the remaining STMS primers yielded varying numbers of bands, all of which were polymorphic and many were cultivar-specific.

Based on the STMS data, genetic similarity coefficients were calculated and a dendrogram constructed (Fig. 2.2). The maximum similarity index of 0.96 was between Subhadra and Chandana, while the least similarity index of 0.08 was between Mangala and Hema and Karjat-35-3 and Hema. Relatively low values of genetic similarity were observed in STMS marker analysis of 42 Indian elite rice cultivars with an average genetic similarly coefficient of 0.37, which was the lowest among all the three molecular markers used (Table 2.2). The arithmetic mean heterozygosity was 0.88 across all the loci and the marker index was 5.49 (Table 2.2), indicating that of the three marker systems analyzed, STMS markers are extremely polymorphic and can detect the maximum genetic variation, and can, therefore, be used to distinguish among different Indian elite rice cultivars.

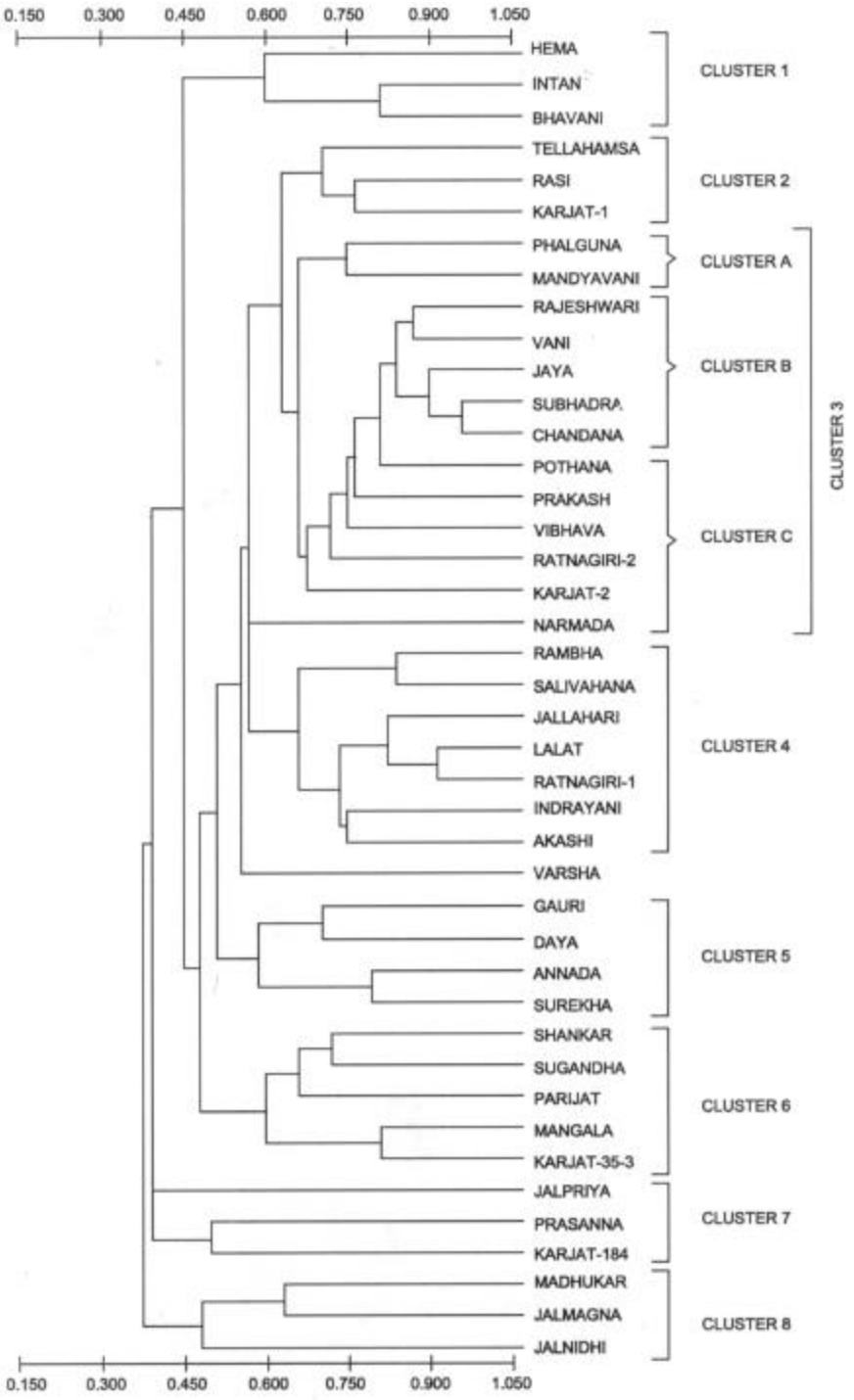


Figure 2.2: Dendrogram showing the genetic relationship among 42 Indian elite rice cultivars using STMS markers

The dendrogram (Fig. 2.2) revealed the presence of eight clusters, many of which were similar to those obtained using coefficient of parentage data. For example, three groups viz. (1) Tellahamsa, Rasi and Karjat-1, (2) Pothana, Vibhava and Ratnagiri-2 and (3) Jaya and Chandana clustered together in clusters 2, 3C and 3B, respectively, using STMS markers. Rambha, Salivahana and Jallahari, which have Pankaj as a common parent also grouped together in cluster 4. With the exception of Intan and Annada in clusters 1 and 5, respectively, and Phalguna, Subhadra and Pothana in cluster 3, all the remaining lines in these clusters were irrigated rice varieties of medium duration maturity. The three elite cultivars in cluster 2, Tellahamsa, Rasi and Karjat-1, are all irrigated rice genotypes of early duration maturity. In addition, the semi-deepwater and deepwater rices, Madhukar, Jalmagna and Jalnidhi clustered together in cluster 8, and formed a separate group. Some irrigated rice varieties of medium duration were also observed to be grouping together. The probability of obtaining an identical match by chance using STMS markers was calculated to be $3.46X10^{-5}$ (Table 2.2).

2.3.3 Combined analysis of RAPD, ISSR and STMS data

The earlier RAPD data generated by Chowdari (1998) in my laboratory, was next compared with the data generated using ISSR and STMS systems in this thesis. Eight RAPD, twelve ISSR, and nine STMS markers yielded a total of 168 bands, out of which 91% were polymorphic with an average of 5.3 polymorphic bands per primer. The average similarity index value was very high at 0.70 with a maximum value of 0.88 between Salivahana and Rambha, having a common parent, Pankaj, and a minimum value of 0.51 between Ratnagiri-2 and Jalmagna. Fig. 2.3 depicts the dendrogram constructed based on the similarity coefficients, which yielded eight clusters. In clusters 3, 4 and 5, many of the rice varieties such as Rajeshwari, Jaya, Vibhava, Prakash, Vani, Mandyavani, Chandana, Karjat-2, Surekha, Ratnagiri-1, Ratnagiri-2, Indrayani and Lalat (but not Subhadra, Karjat-1, Rasi, Annada, Rambha, Salivahana, Jallahari and Akashi), were irrigated rice cultivars of medium duration. Thus, when all the three marker systems were used in combination, 13 out of the 18 irrigated rice cultivars of medium duration were

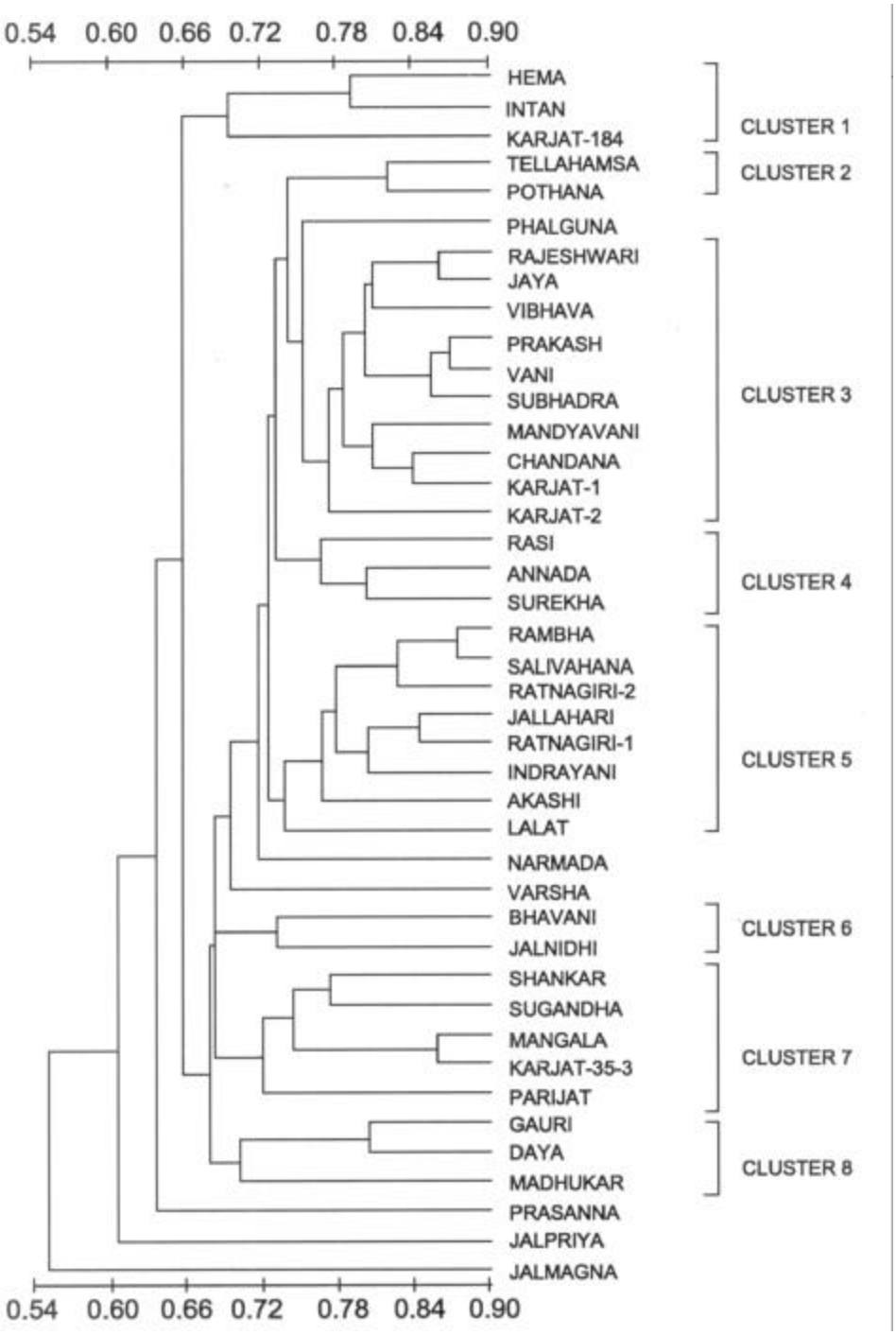


Figure 2.3: Dendrogram showing the genetic relationship among 42 Indian elite rice cultivars using a combination of three different molecular marker systems, RAPD, ISSR and STMS.

found to cluster together as against 9, 10 and 11 cultivars with ISSR, STMS and RAPD markers, respectively. All the five elite rice cultivars, Shankar, Sugandha, Mangala, Karjat-35-3 and Parijat, present in cluster 7, are genotypes of early duration maturity, while all the three cultivars in cluster 8, Gauri, Daya and Madhukar are varieties with medium duration maturity. Two of the deepwater rices, Jalpriya and Jalmagna were totally separate from all the remaining clusters. The probability of obtaining an identical match by chance using all the three molecular markers on the 42 elite rice cultivars was $3.04X10^{-11}$, which was much lower than that obtained by using any individual marker system.

2.3.4 Coefficient of parentage analysis

From the pedigree data of the elite rice cultivars, the coefficients of parentage were calculated and thesis values ranged from zero, between unrelated cultivars to 0.75 between closely related cultivars, with an average of 0.10 for all the 42 elite cultivars. A dendrogram that was constructed based on the coefficients of parentage, produced six clusters (Fig. 2.4). Many of the cultivars in cluster 1, such as Hema, Ratnagiri-1, Phalguna, Gauri, Surekha, Rajeshwari, Prakash, Vani, and Mandyavani (but not Karjat-2, Daya and Annada) as well as the three lines Indrayani, Karjat-35-3 and Akashi, which all joined cluster 1, shared IR8 as a common parent. Annada and Daya, having Kumar and CR-57-49 as parents clustered together with a similarity coefficient of 0.75, while Vani and Mandyavani, having common parents CR104 and IR8, were close together with a similarity coefficient of 0.75. All the cultivars present in cluster 5, with the exception of Chandana, contained T(N)1 as one of the parents. Parijat, which is the parent of Shankar, grouped together with Shankar, while Jaya, the parent of Mangala, was in the same group as Mangala. The cultivars Rambha, Jallahari and Salivahana, which have Pankaj as one of their parents, were grouped together in cluster 3, while Jalmagna, which is the parent of Jalpriya, and Jalpriya grouped together in the last cluster.

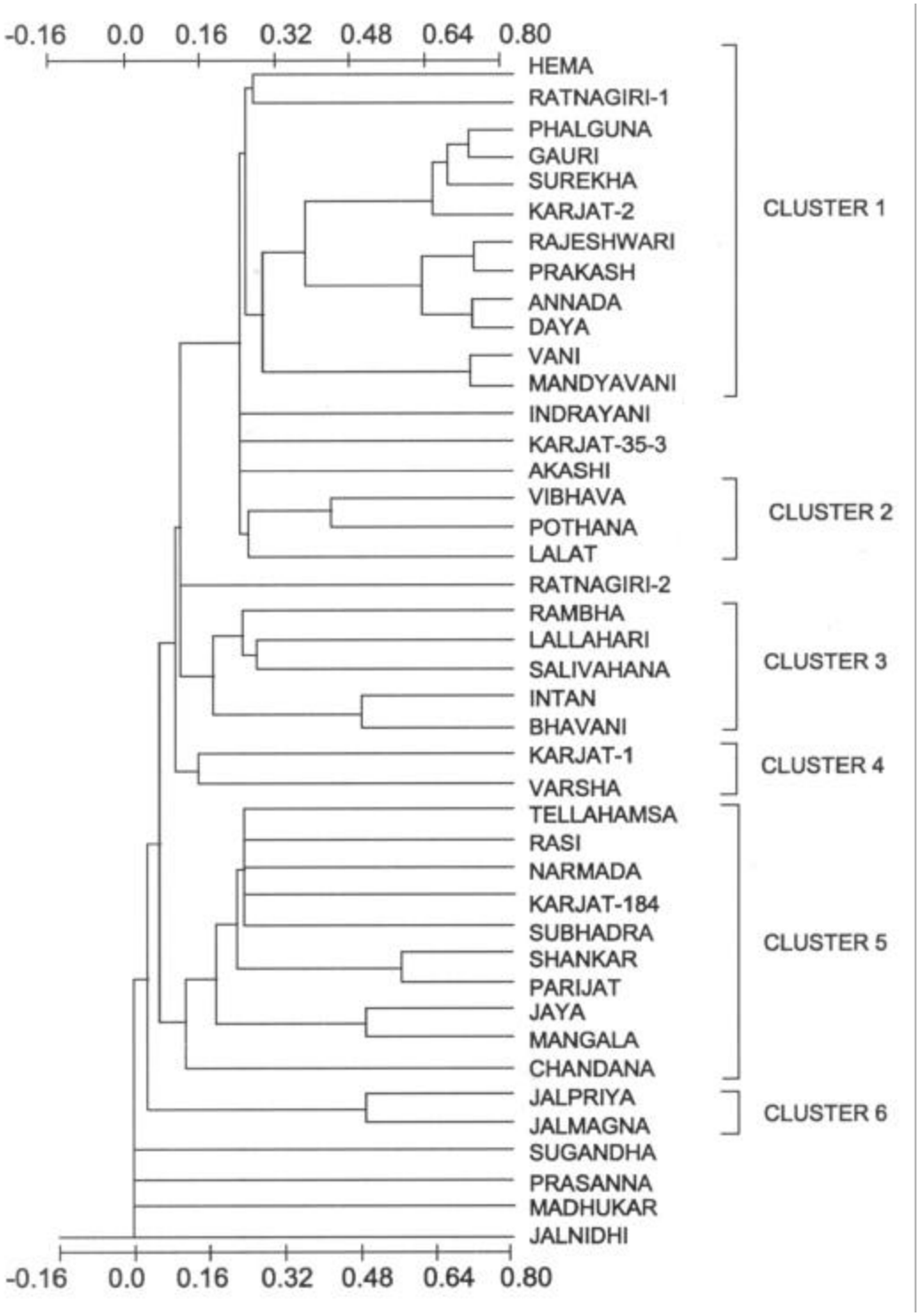


Figure 2.4: Dendrogram showing the genetic relationship among 42 Indian elite rice cultivars based on pedigree data (coefficient of parentage).

2.3.5 Correlations between different measures of genetic similarity

When the genetic similarity matrices generated using individual marker systems were compared with the matrix generated using coefficient of parentage data, RAPD markers resulted in the highest value of 0.085 for Pearson's product moment correlation, while STMS markers showed the least correlation of 0.002 (Fig. 2.5, Table 2.3). However, when related elite rice cultivars (with coefficient of parentage, $r \geq 0.1$) were considered, the value of correlation coefficient was found to increase to 0.294 between the similarity matrix generated using coefficient of parentage and that of all the three molecular marker systems together. Thus, when $r \geq 0.1$, correlation coefficients of 0.281, 0.183 and 0.139 were obtained on comparing the similarity matrix generated with pedigree data with that generated using RAPD, STMS and ISSR markers, respectively (Fig. 2.5, Table 2.3). When similarity matrices generated using different molecular marker systems were compared, the highest correlation of 0.4 was between RAPD and STMS markers followed by 0.310 between ISSR and STMS markers and the least correlation of 0.098 was between RAPD and ISSR markers (Fig. 2.5, Table 2.3). Finally, when the dendrograms based on coefficient of parentage and combined marker data were compared, a very low positive correlation of 0.064 (Table 2.3) was observed between these two dendrograms (Fig. 2.3 and 2.4).

2.4 DISCUSSION

The work embodied in this chapter is the first attempt to characterize elite varieties of rice commonly employed in Indian breeding programs using ISSR, STMS and RAPD markers (Chowdari, 1998)

2.4.1 Rationale behind the choice of marker systems employed

Molecular markers such as RAPD are more or less randomly distributed throughout the genome (Kurata *et al*, 1994b), while ISSR markers span short chromosomal regions between SSRs and hence sample a somewhat different subset of the genome than that assayed with RAPD markers (Parsons *et al*, 1997). ISSR and RAPD markers together span

Table 2.3: Correlations among estimates of coefficient of parentage and genetic similarity based on RAPD, ISSR and STMS data as well as combined data using all the three types of molecular marker systems calculated across (a) all 861 pairwise comparisons of elite rice cultivars (above the diagonal) and (b) across 342 comparisons of related elite cultivars (coefficients of parentage ≥ 0.10) (below the diagonal).

Matrices generated using	**Matrices generated using**				
	Coefficient of parentage	**RAPD**	**ISSR**	**STMS**	**All three molecular marker systems together**
Coefficient of parentage	-	0.085	0.062	0.002	0.064
RAPD	0.281	-	0.098	0.400	0.819
ISSR	0.139	0.082	-	0.310	0.425
STMS	0.183	0.369	0.278	-	0.823
All three molecular marker systems together	0.294	0.810	0.402	0.813	-

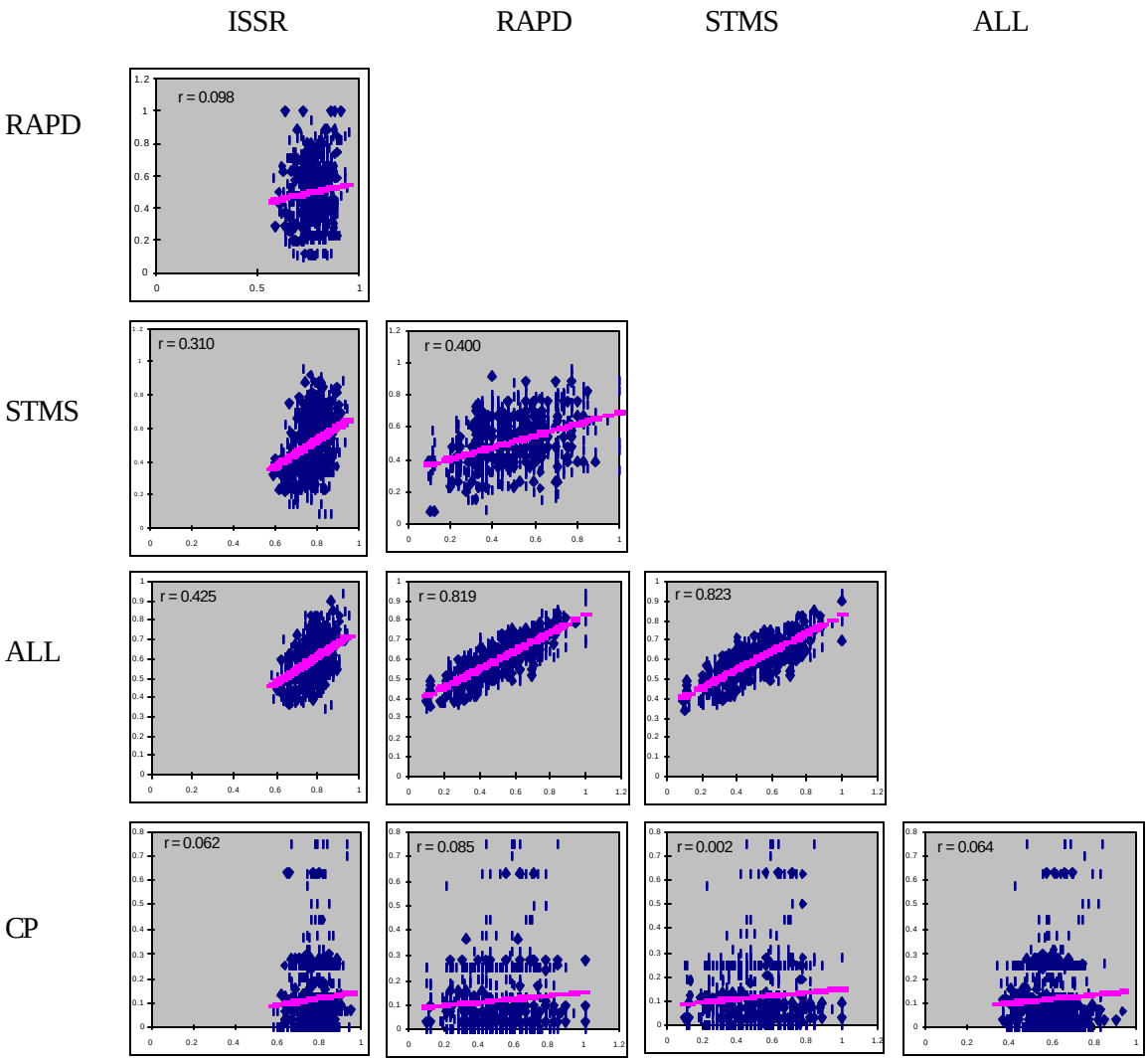


Figure 2.5: Regression analysis for comparison of all the systems used to generate similarity matrices. The term ALL denotes the matrix generated using all the three molecular marker systems viz. RAPD, ISSR and STMS, while CP denotes coefficient of parentage.

majority of the chromosomes and map both proximal and distal to the centromeres (Parsons *et al*, 1997). Also, RAPD markers have been able to map novel chromosomal regions (Monna *et al*., 1994) and, therefore, may be able to target gene-poor regions of the chromosomes, while ISSR markers have been postulated to be useful for marking gene-rich regions (Kojima *et al*., 1998). The genomic distribution of microsatellites in rice seems to be random, with no obvious bias for particular regions or clustering in particular regions (Panaud *et al*, 1996; Chen *et al*, 1997a). Also, STMS markers map to the same locations in both intra- and inter-specific populations because of their locus specific nature, demonstrating that they lie in conserved, evolutionarily stable regions of the genome and may be confidently used for analysis of distantly related germplasms (Panaud *et al*, 1996). The high level of polymorphism observed using such primers is due to variation in the number of tandem repeat motifs at that specific locus (Schlotterer *et al*, 1997).

Primers containing dinucleotide repeats, $(GA)_n$, $(AG)_n$, $(CA)_n$, $(CT)_n$ and $(GT)_n$ and STMS primers amplifying dinucleotides $(GA)_n$ and $(CT)_n$ have been used in our ISSR and STMS analyses, respectively. The reason for the choice of these primers in our study is that, dinucleotide SSRs are more common than any other SSRs in rice (Wu and Tanksley, 1993; Panaud *et al*, 1995; 1996; Chen *et al*, 1997a;), with $(GA)_n$ microsatellite being the most abundant with 1360 poly-(GA) microsatellites followed by 1230 poly-(GT) and 1000 poly-(AC) sites (Panaud *et al*, 1995) in a genome size of 0.45×10^9 bp (Arumunagathan and Earle, 1991; Deshpande and Ranjekar, 1980). Secondly, the $(GA)_n$ and $(GT)_n$ blocks are not found to be clustered together in rice(Wu and Tanksley, 1993). In ISSR analysis Blair *et al*, (1999), poly-(GA) containing, 3'- anchored primers have produced on an average five times as many bands as those with the poly-(GT) motif. They have further shown that dinucleotide motif primers provided the highest rates of polymorphism in rice, while primers based on tri- and tetra-nucleotide motifs have produced fewer bands (Blair *et al*, 1999). Recently, Joshi *et al* (2000) have shown that the most polymorphic patterns are obtained by using $(GA)_n$ or $(AG)_n$ microsatellite-containing primers for ISSR analysis.

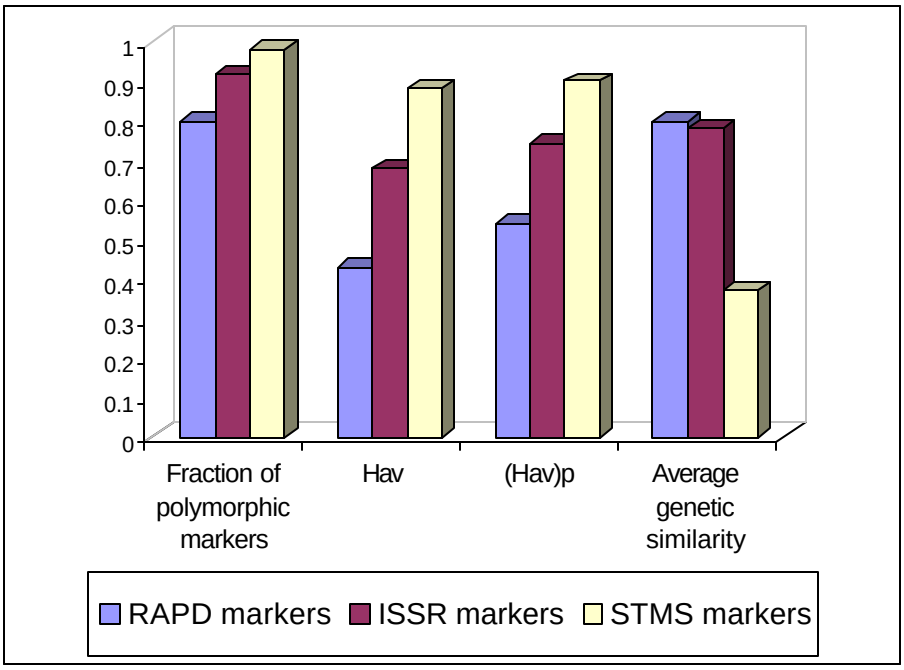


FIGURE 2.6: Histogram showing the polymorphism and average genetic similarity obtained using each of the three DNA marker systems, RAPD, ISSR and STMS. H_{av} is the arithmetic mean heterozygosity across all loci studied while $(H_{av})_p$ is the arithmetic mean heterozygosity at polymorphic loci.

2.4.2 The three marker systems provide a more complete picture of genetic relationship

A comparison among several molecular marker systems for various technical attributes has been detailed by Staub *et al* (1996). It has been demonstrated in rice that microsatellites represent an efficient and suitable marker system for discriminating among closely related germplasm collections and cultivars with a narrow genetic base (Yang *et al*, 1994; Panaud *et al*, 1996; Akagi *et al*, 1997; Olufowote *et al*, 1997). Our study also demonstrates the discriminatory power of microsatellites where STMS markers produced the lowest average genetic similarity coefficient of 0.37 as compared to 0.80 for RAPD and 0.78 for ISSR (Table 2.2, Fig. 2.6). The arithmetic mean heterozygosity (H_{av}) across all loci and marker index are the highest for STMS at 0.88 and 5.49, respectively, followed by 0.68 and 4.05, respectively for ISSR and 0.43 and 2.21, respectively for RAPD markers. Even when only polymorphic loci were considered, the H_{av} values were 0.90, 0.74 and 0.54 for STMS, ISSR and RAPD markers, respectively (Table 2.2, Fig. 2.6) reflecting the hypervariability and higher resolving power of SSR loci.

Combined analysis of rice genotypes by using all three marker systems should give a more holistic picture of the genetic relationships among different elite cultivars. The average genetic similarity using ISSR, RAPD and STMS data is 0.70 and the probability of obtaining an identical match by chance between any two cultivars was observed to be 3.04×10^{-11} as against $2.20X10^{-3}$ for RAPD, $6.40X10^{-4}$ for ISSR and $3.46X10^{-5}$ for STMS markers, indicating that the three marker systems together are able to differentiate all the elite cultivars better. In *Elymus caninus*, microsatellites have detected the highest amount of polymorphism, however, a dendrogram which takes into account all the fragments produced by isozymes, RAPDs and microsatellites has reflected a better relationship than the dendrograms based on only one type of marker (Sun *et al*, 1999). Beer *et al* (1993) have reported that neither RFLP nor allozyme proximities are sufficiently correlated with morphological proximities to warrant their exclusive use for selecting morphologically diverse sets of accessions. However, broad patterns of variation revealed by different types of traits are similar in the *Avena sativa* genotypes surveyed, while differences in pairwise estimates of relationships are sufficiently great to

question the exclusive use of one type of trait for sampling and management of plant germplasm collections. They have also postulated that sampling based on diversity measures related to DNA sequence divergence should optimize the value of the sample, provided that the changes in DNA that are sampled are representative of loci affecting agronomic traits.

In the dendrogram based on STMS marker data as well as that based on combined data of all three marker systems, we have observed that more than 70% of the irrigated rice genotypes of medium duration, cluster together, while in the dendrograms based on RAPD and ISSR data, few of these lines are found clustering together. The deep water rices Jalpriya and one of its parents, Jalmagna, cluster together in the ISSR based dendrogram as well as in the combined dendrogram of all the markers. Thus, the combined dendrogram gives a more complete picture of the relationships among these elite rice cultivars than the dendrograms based solely on a single marker system.

2.4.3 Correlation between molecular data and coefficient of parentage data: a complex issue

From Fig. 2.5 and Table 2.3, it is clear that very poor correlation (correlation coefficient < 0.1) is observed between the similarity matrices generated using molecular markers and the matrix generated using coefficients of parentage. Low correlations have been observed between coefficients of parentage and genetic similarities in wheat, barley and oat cultivars (Graner *et al*, 1994; Plaschke *et al*, 1995; Schut *et al*, 1997; Bohn *et al*, 1999). These results alongwith ours suggest that there are fundamental differences in the concepts underlying both these measures of genetic diversity between two genotypes. The coefficient of parentage or coancestry is an indirect measure of genetic diversity, which quantifies the degree to which two genotypes are "identical by descent". In contrast, the rationale for using genetic similarity estimates based on molecular marker data is that the proportion of bands shared between two genotypes is an indicator of their resemblance in the DNA sequence across the entire genome (Nei, 1987). Consequently, genetic similarity is a direct measure and should reflect the proportion of "genes alike in state"; irrespective of whether the identity is caused by alleles "identical by

descent" or only those "alike in state". Genetic similarity based on molecular markers also assumes that all co-migrating bands are identical (Bohn *et al*, 1999).

As reviewed by Cox *et al* (1985), Graner *et al* (1994) and Messmer *et al* (1993), practical assumptions required for the computation of coefficients of parentage maybe a source of poor relationship between the two measures. The alleles are not always transmitted equally from female and male parents to the progeny and, therefore, in reality each parent may not contribute equally to the cross (Kim and Ward, 1997). For example, the transmission of alleles controlling quantitative traits with high heritability is clearly influenced by intense selection pressure in the breeding program, which results in a biased contribution from one parent with favorable alleles (parental superiority) to the progeny generation (Cox *et al*, 1985; Souza and Sorrells, 1989). However, St Martin (1982) has shown that such shifts are unlikely to be large and that with a biparental cross, most offspring would receive between 40 and 60% of their alleles from each parent. In addition to selection, genetic drift during self-pollination of cultivar development causes deviations from expected genetic relationships (Cox *et al*, 1985).

Cox *et al* (1985) and Graner *et al* (1994) have postulated that poor relationships between the two measures of genetic diversity may result partially from the high levels of genetic similarity between unrelated lines. In our study, mean coefficient of parentage and coefficient of similarity over the 861 cultivar pairs are 0.10 and 0.70, respectively. Thus, the genetic diversity estimates based on pedigree are on an average about seven times higher than estimates using any other marker system. Prabhu *et al* (1997) have observed in soybean that genetic diversity estimates using pedigree are on an average about two times higher than that estimated using any marker method. The lower similarity estimates or higher diversity estimates based on pedigree could result from the assumption that within a pedigree, any variety whose parents are not known has no relationship (coefficient of parentage =0) with any variety in the set. The high diversity estimates could, therefore, be due to a relative lack of accurate (if at all) pedigree information of many of the cultivars (Huang *et al*, 1996; Prabhu *et al*, 1997).

In addition, the breeding parents may not always be homogeneous and homozygous (Kim and Ward, 1997). In case of many cultivars, the lack of any relationship between coefficient of parentage and genetic similarity estimates using molecular markers could be a consequence of the heterogeneous nature of the ancestors, which are mostly landraces rather than pure-line cultivars. It is unlikely that the samples which have been analyzed have exactly the same patterns as the individual plants involved in selection and hybridization leading to the development of modern cultivars. It is also possible that different breeding programs maybe using somewhat different variants of cultivars which are assumed to be identical as observed by Skorupska *et al* (1994) in various Peking soybean cultivars obtained from different countries / regions. However, in computing coefficients of parentage it was assumed that all plants of an ancestor are identical (Cox *et al*, 1985). The inconsistent relationships revealed by molecular markers and coefficients of parentage could also be due to sampling deviations and the failure of phenotypes to differentiate the genotypes precisely (Delannay *et al*, 1983; Kim and Ward, 1997).

Molecular markers directly sample the DNA composition of the genomes, regardless of known pedigree and may, therefore, provide more accurate estimates of the true similarities and differences between genomes (Prabhu *et al*, 1997). However, the accuracy of genetic similarity estimates based on molecular markers depends on the location and number of molecular markers employed (Bohn *et al*; 1999).

Sampling effects due to inappropriate genome coverage may increase the standard errors in estimating genetic similarity using molecular markers resulting in poor correlation between coefficients of parentage and genetic similarity based on molecular markers as studied by Bohn *et al* (1999) in wheat. Individual molecular marker systems may not be able to provide complete genome coverage and, hence, a combination of RAPD, ISSR and STMS provides whole genome coverage and reduces the errors in genetic similarity estimation based on any one marker system alone. Also, if we assume that by increasing the number of molecular markers, the variance of individual genetic similarity estimates is decreasing, the large number of

polymorphic bands obtained in our analyses should provide more precise genetic similarity estimates.

Although single copy and few repetitive DNA classes contain expressed sequences, the majority of the genome complexity is not associated with expressed genes (Goldberg *et al*, 1978). However, it may be possible that some banding patterns are correlated with some traits selected during elite line development, which may violate the selective neutrality. Therefore, the majority of the random molecular markers employed in our study would be associated with non-expressed and neutral sequences that should not be affected by selection and would, therefore, not violate selective neutrality. The maker data as a whole should, therefore, be reflective of the genetic distance among the parents.

2.4.4 Potential of DNA data in broadening the genetic base of Indian elite rice cultivars

IR8, T(N)1 and TKM6 have been used most frequently in the Indian breeding program to develop the elite rice varieties used in the present study. Analysis of the crosses suggests that even though many combinations have been produced, only limited genetic variability is available within these combinations. For example, IR8 and T(N)1 have been individually used as one of the parents in developing more than 40% of the elite cultivars studied. Even though our study indicates the presence of a narrow genetic base in all the 42 Indian elite rice varieties, they still exhibit extensive genetic diversity for various morphological, biological and physical characters. However, if the clustering data obtained in Fig. 2.3 can be used along with morphological and biological characters, then it would assist breeders in selecting suitable genetically diverse parents for their crossing programs. One potential strategy is to select genotypes initially by useful agronomic characteristics, and then select from those genotypes' subsets that are mutually dissimilar, on the basis of molecular marker data. This could maximize opportunities for transgressive segregation, because there is a higher probability that unrelated genotypes would contribute unique desirable alleles at different loci (Beer *et al*, 1993). For example, Salivahana in cluster 5 (Fig. 2.3), with a yield of 6-7 tons/hectare and having resistance to blast and sheath blight (personal

communications with rice breeders, Directorate of Rice Research, Hyderabad, India) can be crossed with Phalguna, which lies between clusters 2 and 3 and is resistant to gall midge (Breeding records of Directorate of Rice Research, Hyderabad, India). Both these rice varieties are rainfed lowland rice cultivars, with diverse parents and group in different clusters using coefficient of parentage data (Fig. 2.4). Similarly, Vibhava in cluster 3 (Fig. 2.3) with a yield potential of 6.5-7 tons/hectare and resistance to gall midge and sheath blight can be crossed with Prasanna, which totally outgroups from all the clusters and is resistant to blast. These two elite cultivars cluster in different groups based on coefficient of parentage data as both of them have diverse parents. Thus, either of these crosses could be performed with a goal of creating progeny that are resistant to blast, gall midge and sheath blight.

In conclusion, the use of a combination of molecular markers gives a better understanding of the genetic relationships among closely related rice cultivars and also gives a tool to rice breeders in broadening the genetic base in rice breeding populations.

CHAPTER 3

Potential of $(GATA)_n$ Microsatellites from Rice for Studying Inter- and Intra-Specific Variability

ABSTRACT

Although use of $(GATA)_n$ microsatellite has been reported in fingerprinting of rice, very few attempts have been made to study the $(GATA)_n$-containing microsatellite loci in this crop. Three polymorphic $(GATA)_n$ harboring loci viz. OS1A6, OS1H10 and OS2E7, containing 7-13 repeat motifs were identified from the genomic library of a cultivated rice, *Oryza sativa* var. Basmati-370 using an oligonucleotide probe $(GATA)_4$. Primers flanking GATA repeats were used to screen a large germplasm of rice including 26 wilds (representing all the genomes of rice), 16 cultivars, 47 elite rice varieties of India and 37 lines resistant / susceptible to bacterial leaf blight. Upto 22 alleles were obtained at an individual locus, in the germplasm analyzed, revealing the potential of $(GATA)_n$-containing loci as powerful tools to detect simple sequence length polymorphism (SSLP). The $(GATA)_n$ flanking primers were not only useful in distinguishing closely related genotypes, but they could also be used for cross species amplification. For example, an average of 5.67 alleles were obtained per locus when only closely related Indian elite rice varieties were analyzed, while an average of 14.5 alleles were obtained per locus when wild and cultivated genotypes were considered. Another interesting observation was the clustering of bacterial blight resistant lines into a separate group from the remaining rice genotypes, when a dendrogram was constructed based on the polymorphism obtained at the three loci. This may be due to the partial homology of the clones, OS1H10 and OS2E7 to regions encoding for *O.longistaminata* receptor kinase-like protein and pathogenesis-related protein. The ability of these *O.sativa* flanking primers to amplify DNA of maize, wheat, barley and oat, indicated the conservation of these $(GATA)_n$-containing loci across different cereal genera.

3.1 INTRODUCTION

Microsatellites (Litt and Luty, 1989) or simple sequence repeats (SSRs) (Tautz, 1989; Weber and May, 1989), consist of simple tandemly repeated di- to penta-nucleotide sequence motifs. Since they are abundant, hypervariable, multi-allelic and evenly distributed throughout nuclear genomes of eukaryotes (Tautz and Renz, 1984; Epplen, 1988; Goodfellow, 1992; Wu and Tanksley, 1993; Bell and Ecker, 1994) as well as some prokaryotic genomes (Field and Willis, 1996), they provide a valuable source of polymorphism making them an important class of genetic markers. The exceptionally high levels of polymorphism detected by SSRs, referred to as simple sequence length polymorphism (SSLP) is due to the variability in the number of tandem repeats at a particular locus (Schlotterer *et al*, 1997).

SSLP can be easily, accurately, rapidly and economically assayed by developing sequence tagged microsatellite markers, which can be useful in PCR analysis (Beckman and Soller, 1990; Love *et al*, 1990; Wu and Tanksley, 1993; Bell and Ecker, 1994). Development of these markers requires screening a genomic library with microsatellites followed by selection and sequencing of the clones containing the microsatellite inserts and synthesizing primers flanking the SSR motif (Weber and May, 1989; Litt and Luty, 1989), which can then be used to amplify locus specific DNA fragments. These second generation markers are somatically stable and inherited in a co-dominant Mendelian manner and can, therefore, distinguish between heterozygotes and homozygotes (Weber, 1990b; Morgante and Olivieri, 1993; Thomas and Scott, 1993). All these factors together have made SSRs ideal markers for plant breeding, genetic linkage analysis, gene mapping, paternity analysis, forensic identification and population genetics (Weber, 1990a; Rafalski and Tingey, 1993; Gupta *et al*, 1996; Jarne and Lagoda, 1996).

In rice, about 50% of the genome consists of repetitive DNA sequences (Deshpande and Ranjekar, 1980), with presence of approximately 5,700 to 10,000 SSRs (McCouch *et al*, 1997). Totally only 323 SSRs have been mapped on the 12 chromosomes of rice, with an average distance of one SSLP per 6 cM (Cho *et al*, 2000; Temnykh *et al*, 2000). Most of the SSR loci studied so far mainly contain dinucleotide motifs and few contain trinucleotide

motifs while very few tetranucleotide-containing microsatellite loci have been identified.

Our laboratory has earlier reported the usefulness of tetranucleotide microsatellite $(GATA)_4$ in DNA fingerprinting of different rice genotypes (Ramakrishna *et al*, 1994). The work on other crop plants at our laboratory has further revealed that the oligonucleotide $(GATA)_4$ is not only important in fingerprinting cultivars of pearl millet (Chowdari *et al*, 1998a; 1998c) and isolates of its pathogen, *Sclerospora graminicola* (Sastry *et al*, 1995) but is also useful in detecting sex specific differences in papaya (Parasnis *et al*, 1999). In the present work, I have cloned and sequenced three $(GATA)_n$-containing loci from *Oryza sativa* var. Basmati-370, and used these $(GATA)_n$ locus specific microsatellite markers to study the allelic variation revealing the intra- and inter-specific relationships in rice.

3.2 MATERIALS AND METHODS

3.2.1 Library screening and selection of clones containing $(GATA)_n$ repeats

A genomic library, constructed using *EcoR*I digest of *Oryza sativa* var. Basmati-370 in Charon 40 phage vector (PFU $4.8X10^{-8}$), was screened with the oligonucleotide $(GATA)_4$, by plaque hybridization (Sambrook *et al*, 1989). The tetranucleotide, $(GATA)_4$, was end-labeled using γ^{32}P ATP and T4 polynucleotide kinase (U.S. Biochemical Corp., U.K.), according to Sambrook *et al* (1989). To 20 pmoles of $(GATA)_4$, 50µCi γ^{32}P ATP, one-tenth volume of 10X kinase buffer for T4 polynucleotide kinase [670mM Tris-HCl (pH 8.0), 100mM $MgCl_2$, 100mM DTT] and 5U T4 polynucleotide kinase (U.S. Biochemical Corp. U.K.) were added and the reaction was incubated at 37°C for 45 min. The reaction was stopped by adding one-tenth volume of 0.5M EDTA and the labeled probe was separated from unincorporated γ^{32}P ATP by ion exchange chromatography on DE-52 column. Labeled probe was loaded on 200µl of DE-52 column packed in a 1.5ml eppendorf tube and the column was washed with 4ml of TE buffer and 4ml of 0.2M NaCl in TE buffer. 5' end-labeled oligonucleotide was finally eluted by two washes of 500µl each of 1M

NaCl in TE buffer and stored at -20^{o}C to be used later as a probe for hybridization.

Hybridization was carried out at (Tm-5)^{o}C for 16 hours in 15ml hybridization buffer containing 5X SSPE [0.75M NaCl, 0.05M NaH_2PO_4 and 0.01M Na-EDTA], 0.1% SDS, 5X Denhardt's solution [0.1% Ficoll, 0.1% polyvinyl pyrrolidone] and 0.2X BLOTTO [1% defatted milk powder in water]. Post hybridization washes were given twice for 10min each at room temperature and for 5min at hybridization temperature with 3X SSPE and 0.1% SDS. The filters were exposed to X-ray films at –70^{o}C for 12-24 hours (depending on the signal) with intensifying screens. To reduce the number of false positives, three rounds of screening were performed, wherein the clones that hybridized with $(GATA)_4$, probe in the first round were transferred onto new filters and re-hybridized with the probe, $(GATA)_4$.

3.2.2 Subcloning of the positive clones

Phage DNA, isolated from the selected putative positive clones by lambda DNA isolation kit (QIAGEN, U.S.A.), was digested with *Taq*I and *Msp*I restriction endonucleases (Promega, U.S.A.) according to manufacturer's instructions. The DNA fragments thus obtained were ligated into the *Acc*I site of pUC18 (Yanisch-Peron *et al*, 1985) or the *Cla*I site of pBluescript cloning vectors according to manufacturer's instructions. These recombinant plasmids were used to transform *Escherichia coli* host strain XL1-Blue competent cells and the transformed cells were plated on Luria Bertani (LB) agar medium [1% Bacto-tryptone, 0.5% NaCl and 0.5% Bacto-yeast extract (pH 7.2), 2% Bacto-agar]. The white colonies obtained were transferred into microtitre plates containing LB medium [1% Bacto-tryptone, 0.5% NaCl and 0.5% Bacto-yeast extract (pH 7.2)] with 100µg/ml of Ampicillin. Colony hybridization was performed using end-labeled $(GATA)_4$ probe (Sambrook *et al*, 1989) in order to select the recombinant clones containing $(GATA)_n$ repeat.

A single colony with the recombinant plasmid was inoculated in LB medium containing 100µg/ml of Ampicillin and was grown to saturation at 37^{o}C for 16 hours with shaking at 175 rpm. Plasmid DNA was isolated using alkaline lysis method (Sambrook *et al*, 1989). The cells were pelleted by

centrifugation at 5000 rpm for 10 min. GTE buffer [50mM Glucose, 25mM Tris-HCl (pH 8.0) and 10mM EDTA (pH 8.0)] was added to the pellet, vortexed and incubated at room temperature for 10min. Then double the volume of solution II [1% SDS and 0.2N NaOH] was added to the above cell suspension, the contents were mixed well and kept on ice for 10min. The suspension was neutralized by adding half volume of potassium acetate [60ml 5M Potassium acetate, 11.5ml glacial acetic acid and 28.5ml water], mixed well by inversion, kept on ice for 10 min and centrifuged at 10,000 rpm for 10min to pellet out cell debris and chromosomal DNA. To the supernatant, equal volume of phenol was added which was then centrifuged as above, followed by treatment with equal volume of a mixture of phenol : chloroform : isoamyl alcohol (25:24:1) and lastly with chloroform : isoamyl alcohol (24 : 1). The DNA was then precipitated by adding one-tenth volume of 3M sodium acetate (pH 5.2) and two and a half volumes of absolute ethanol. The DNA precipitate was washed with 70% ethanol, centrifuged and redissolved in TE buffer [10mM Tris-HCl (pH 8.0) and 1mM EDTA (pH 8.0)]. For removal of RNA, the DNA was incubated with RNase A (DNase free) at 37^{o}C for 2 hours. The plasmid DNA thus obtained was purified by polyethylene glycol method (Sambrook *et al*, 1989).

3.2.3 Sequencing of plasmid DNA

The purified plasmid DNA was sequenced by Sanger's dideoxy chain termination method using Sequenase version 2.0 DNA sequencing kit (U.S. Biochemical Corp., U.K.). 5µg of purified plasmid DNA was denatured at 37^{o}C for 30 min after adding 4µl of denaturing solution [2M NaOH and 2mM EDTA] and making up the volume to 40µl with sterile distilled water. One-tenth volume of 3M sodium acetate (pH 5.2) and two and a half volumes of chilled absolute ethanol were added and the mixture was kept at -70^{o}C for 30 min. DNA was precipitated by centrifugation at 10,000 rpm for 10 min at 4^{o}C and the DNA pellet was washed with 70% ethanol and dissolved in 7µl sterile distilled water.

The second step was primer annealing in which 1µl (0.5 pmoles) of primer (pUC sequencing or reverse sequencing or specific primer) and 2µl of

5X reaction buffer were added, incubated at 65°C for 2 min and then allowed to cool gradually to 37°C over 15-30 min. To this annealed template-primer mix which was then chilled on ice, 1µl of 0.1M DTT, 2µl labeling mix, 1µCi α^{35}S dATP or α^{32}P dATP and 2µl of diluted T7 DNA polymerase (3.7U) were added and incubated at 21°C for 2-5 min. Then 3.5µl of the mix was added to four tubes containing 2.5µl of the respective dideoxynucleotide and incubated at 37°C for 5 min for chain termination. Finally, 4µl of stop solution was added to terminate the reaction. The sequencing reactions were heated at 75-80°C for 2 min to denature and were then loaded on 6% polyacrylamide gels containing 7M urea and 0.5X TBE and electrophoresed at 2000V. Successive loadings were performed so that the entire sequence could be read. The gels were dried on a vacuum gel dryer at 80°C for 2 hours and exposed to X-ray films overnight.

3.2.4 Homology search and design of primers

In order to determine if the sequenced clones had any homology to the known sequences of rice and Arabidopsis and to explore the abundance of repeats in these two model plants, database searches were carried out using $(GATA)_4$ as a query sequence in BLAST (www.ncbi.nlm.nih.gov/BLAST). PCR primers, flanking the $(GATA)_n$ repeat motif, were designed and were synthesized partly by Research Genetics Inc (USA) and partly at the primer synthesis facility at NCL (India).

3.2.5 Plant material

The PCR primers synthesized above were used to screen wild and cultivated rice genotypes, along with four representative cereal genotypes. 26 wilds, 6 cultivated rice varieties and related genera (Table 3.1) were obtained from the International Rice Research Institute, Manila, Philippines, while the remaining two wilds and all the other cultivated rice genotypes (Table 3.1), along with 47 Indian elite rice varieties (Table 3.2) and 37 bacterial blight resistant and susceptible lines (Table 3.3) were obtained from Directorate of Rice Research, Hyderabad, India. Seeds of maize were procured from Indian Agricultural Research Institute, New Delhi, while seeds of wheat were

Table 3.1: Number of alleles obtained in wild and cultivated rice genotypes as well as four cereal genotypes at the $(GATA)_n$-containing loci

Sr. No.	List of samples used	Accession	Genome	Origin	Number of alleles obtained with		Total number of alleles
					OS1A6	OS1H10	
	Wild *Oryza* species						
1	*O.glaberrima*	TOG 6216	AA	Africa via Warda	1	2	3
2	*O.glaberrima*	TOG-6229	AA	Africa via Warda	1	3	4
3	*O.longistaminata*	-	AA	Africa via CRRI	2	2	4
4	*O.perennsis*	104823	AA	Thailand	1	2	3
5	*O.rufipogon*	106424	AA	Vietnam	2	1	3
6	*O.rufipogon*	105908	AA	Thailand	1	1	2
7	*O.punctata* (2n)	105980	BB	Cameroon	1	2	3
8	*O.punctata* (4n)	100884	BBCC	India	1	2	3
9	*O.minuta*	101141	BBCC	Philippines	1	3	4
10	*O.minuta*	101125	BBCC	Philippines	1	1	2
11	*O.rhizomatis*	105432	CC	Sri Lanka	1	2	3
12	*O.officinalis*	100896	CC	Thailand	1	1	2
13	*O.officinalis*	101399	CC	Vietnam	1	-	1
14	*O.alta*	105143	CCDD	Guyana	1	1	2
15	*O.grandiglumis*	105669	CCDD	Brazil	1	1	2
16	*O.latifolia*	100167	CCDD	Costa Rica	1	-	1
17	*O.latifola*	100965	CCDD	Costa Rica	1	2	3
18	*O.australiensis*	100882	EE	Australia via CRRI, India	1	-	1
19	*O.australiensis*	T-1434	EE	Australia via CRRI, India	1	2	3
20	*O.brachyantha*	101232	FF	Sierra Leone	1	-	1
21	*O.brachyantha*	B98-8025	FF	Africa	1	-	1
22	*O.granulata*	104986	GG	via CRRI, India	1	1	2
23	*O.granulata*	106448	GG	Nepal	1	7	8
24	*O.longiglumis*	105148	HHJJ	Indonesia	-	-	-
25	*O.ridleyi*	100821	HHJJ	Thailand	1	1	2
26	*O.ridleyi*	101453	HHJJ	Malaysia	1	2	3
27	*O.nivara*	-	AA	India	2	2	4
28	*O.malampuzhaensis*	-	BBCC	India	1	-	1
	***O.sativa* cultivars**						
29	IR 36	Indica	AA	Cultivated variety	1	2	3
30	IR 64	Indica	AA	Cultivated variety	1	1	2
31	BG-90-2	Indica	AA	Cultivated variety	1	2	3
32	Basmati-370	Aromatic rice	AA	Cultivated variety	1	1	2
33	Azucena	Japonica	AA	Cultivated variety	1	-	1
34	Indrayani	Indica	AA	Cultivated variety	1	2	3
35	Ambemohr	Indica	AA	Cultivated variety	1	1	2
36	Intan	Indonesian	AA	Cultivated variety	1	1	2
37	Khadkya	Landrace	AA	Maharashtra India	2	1	3
38	Waseasahi	Japonica	AA	Cultivated variety	1	-	1
39	Hakkoda	Japonica	AA	Cultivated variety	1	1	2
	Related genera						
40	*Porteresia coarctata*	Related genera	Unknown	Bangladesh	1	3	4
41	*Leersia*	Related genera	Unknown	Madagascar	1	-	1
42	*Rhynchoryza subulata*	Related genera	Unknown	Argentina	1	1	2
	Cereal genera						
43	*Zea diploperensis*	-	-	-	1	2	3
44	*Triticum durum* landrace Narsimgarh III	1633	-	-	-	2	2
45	*Hordeum marinum*	PI 200341	-	-	-	1	1
46	*Avena vaviloviana*	PI 412733	-	-	-	3	3

Table 3.2: Number of alleles obtained in Indian elite rice cultivars at the $(GATA)_n$-containing loci

Sr. No.	Name	Number of alleles obtained with			Total number of alleles obtained
		OS1A6	OS1H10	OS2E7	
1	Hema	1	1	2	4
2	Tellahamsa	1	1	1	3
3	Rambha	1	1	1	3
4	Jallahari	3	1	2	6
5	Phalguna	2	3	3	8
6	Intan	1	-	-	1
7	Rajeshwari	1	1	1	3
8	Jaya	1	-	1	2
9	Indrayani	1	1	1	3
10	Bhavani	1	-	-	1
11	Rasi	1	1	3	5
12	Lalat	2	1	1	4
13	Prakash	1	-	2	3
14	Gauri	1	1	1	3
15	Shankar	1	1	1	3
16	Ananga	1	-	2	3
17	Akashi	1	-	3	4
18	Sugandha	1	1	1	3
19	Madhukar	1	1	1	3
20	Subhadra	1	1	1	3
21	Vibhava	1	1	1	3
22	Rudra	1	1	2	4
23	Mangala	1	1	3	5
24	Varsha	1	1	2	4
25	Jalpriya	1	1	1	3
26	Pothana	1	1	3	5
27	Ratna	1	-	-	1
28	Jalnidhi	1	1	1	3
29	Salivahana	1	1	1	3
30	Surekha	1	1	1	3
31	Daya	1	1	1	3
32	Narmada	1	1	1	3
33	Parijat	1	1	2	4
34	Jalmagna	1	1	2	4
35	Vani	1	2	1	4
36	Vikas	1	1	1	3
37	Mandyavani	1	1	1	3
38	Chandana	1	2	1	4
39	Prasanna	1	1	3	5
40	Karjat-1	1	1	3	5
41	Karjat-2	1	1	1	3
42	Karjat-18	1	1	1	3
43	Karjat-23	1	2	2	5
44	Karjat-35-3	1	1	1	3
45	Karjat-184	1	-	1	2
46	Ratnagiri-1	1	1	1	3
47	Ratnagiri-2	1	1	2	4

Table 3.3: Number of alleles obtained in bacterial blight resistant and susceptible rice genotypes at the $(GATA)_n$-containing loci

Sr. No.	Lines analyzed		Number of alleles obtained with			Total number of alleles
			OS1A6	OS1H10	OS2E7	
1	IRBB-3	Near isogenic line	1	-	2	3
2	IRBB-4	Near isogenic line	1	1	1	3
3	IRBB-5	Near isogenic line	1	1	4	6
4	IRBB-7	Near isogenic line	1	1	1	3
5	IRBB-8	Near isogenic line	1	1	6	8
6	IRBB-10	Near isogenic line	1	1	6	8
7	IRBB-13	Near isogenic line	1	1	6	8
8	IRBB-21	Near isogenic line	1	1	5	7
9	BJ-1	Tall BLB resistance donor	1	2	2	5
10	DV-85	Tall BLB resistance donor	1	2	-	3
11	Camor (AC 17366)	Tall BLB resistance donor	1	2	1	4
12	Java-14	Tall BLB resistance donor	1	1	1	3
13	AC-19-1-1	BLB resistance donor	1	2	1	4
14	Kalimekri	BLB resistance donor	1	2	1	4
15	Cemposelak	Tall BLB differential	1	2	1	4
16	Hashikalmi	BLB resistance donor	1	3	2	6
17	Tadukan	BLB susceptible but blast resistance donor	1	2	9	12
18	PR-113	BLB resistant variety, Ajaya derivative	1	2	3	6
19	T(N)-1	BLB susceptible variety (semi-dwarf)	1	2	6	9
20	Jaya	BLB susceptible variety (dwarf)	1	-	2	3
21	Rasi	BLB susceptible variety (dwarf)	1	2	2	5
22	BPT- 5204	BLB susceptible variety (dwarf)	1	2	6	9
23	IR-54	BLB resistant variety	1	2	3	6
24	IR-64	BLB resistant variety	1	3	6	10
25	IR-22	BLB resistant variety (semi-dwarf)	1	3	3	7
26	Swarna	-	1	4	1	6
27	IR-20	BLB resistant variety	1	3	3	7
28	Kuntalan	BLB resistance donor	1	2	3	6
29	Nigeria-5	BLB resistance donor	1	2	2	5
30	Zenith-BB	BLB resistant line	2	4	1	7
31	Chugoku-45	Dwarf line	1	-	6	7
32	IET-8320	BLB resistant dwarf pyramid	2	2	1	5
33	IET-8585 (Ajaya)	BLB resistant dwarf pyramid	1	4	1	6
34	RP-2151-222-4	BLB resistant dwarf pyramid	1	2	4	7
35	IET-14444	Dwarf pyramid, Ajaya derivative	1	2	4	7
36	CRRI-1	BLB resistant pyramid from CRRI	1	1	4	6
37	CRRI-3	BLB resistant pyramid from CRRI	1	1	2	4

obtained from Agharkar Research Institute, Pune, India and those of barley and oat were made available by U.S.D.A. A.R.S., National Small Grains Collection, Aberdeen, U.S.A.

DNA was extracted from leaf tissue by CTAB method as described earlier in chapter 2 (section 2.2.1).

3.2.6 Polymerase chain reaction using primers flanking the $(GATA)_n$ microsatellite motifs

The PCR volumes, amplification conditions, sample preparation for loading and electrophoresis conditions were as described in chapter 2 (section 2.2.3) with some modifications. α^{32}P dATP was used instead of α^{32}P dCTP, and, therefore, 25μM dATP and 200μM dCTP were added in the reaction. 50ng of template DNA was used per reaction and PCR was performed using Taq DNA polymerase from Boehringer Mannheim, Germany. Forty thermal cycles were performed instead of 35 and DNA amplifications were carried out at specific annealing temperature depending on the Tm of the primer pair.

ϕX174 DNA digested with restriction enzyme, *Hae*III (Bangalore Genei, India) was dephosphorylated using calf intestinal phosphatase (CIP) (Boehringer Mannheim, Germany) and end-labeled using γ^{32}P ATP according to Sambrook *et al* (1989) and loaded as molecular weight standard.

3.2.7 STMS analysis using primers flanking mapped microsatellite loci

Primers flanking mapped microsatellite loci were obtained from Research Genetics, U.S.A. (Table 3.4) and STMS reactions, sample preparation, loading and electrophoresis were carried out as described above. However, 20ng template DNA was used per reaction instead of 50ng and 30 thermal cycles were performed wherein the annealing was reduced to 1 min.

3.2.8 Data analysis

Band sizes of each of the alleles were calculated in base pairs, using the program, SeqAid II (tm) version 3.5 (Schaffer and Sederoff, 1981), wherein the distance of migration of the bands in the marker lane was

TABLE 3.4: List of mapped STMS markers used for analysis of bacterial blight resistant / susceptible genotypes

Sr. No.	Gene	Located on Chromosome	Linked DNA marker	Type of marker	Sequence of primers	Number of alleles	Reference
1	*xa*-5	5	RM122	STMS	F GAGTCGATGTAATGTCATCAGTGC R GAAGGAGGTATCGCTTTGTTGGAC	6	Wu & Tanksley, 1993
2	*xa*-5	5	RM390	STMS	F CCCTTGTTTCAGTGGCTCAG R CCAAGATCAAGAACAGCAGGAATC	2	Blair & McCouch, 1997
3	*xa*-5	5	RM13	STMS	F TCCAACATGGCAAGAGACAG R GGTGGCATTCGATTCCAG	6	Panaud *et al*, 1996
4	*Xa*-10	11	RM206	STMS	F CCCATGCGTTTAACTATTCT R CGTTCCATCGATCCGTATGG	10	Chen *et al*, 1997a
5	*Xa*-10	11	RM254	STMS	F AGCCCCGAATAAATCCACCT R CTGGAGGAGCATTTGGTAGC	4	Chen *et al*, 1997a
6	*xa*-13	8	RM230	STMS	F GCCAGACCGTGGATGTTC R CACCGCAGTCACTTTTCAAG	4	Chen *et al*, 1997a
7	*Xa*-21	11	RM21	STMS	F ACAGTATTCCGTAGGCACGG R GCTCCATGAGGGTGGTAGAG	11	Panaud *et al*, 1996
8	*Xa*-3 / *Xa*-4	11	RM224	STMS	F ATCGATCGATCTTCACGAGG R TGCTATAAAAGGCATTCGGG	8	Chen *et al*, 1997a
9	*Xa*-3 / *Xa*-4 / *Xa*-10 / *Xa*-21	11	RM167	STMS	F GATCCAGCGTGAGGAACACGT R AGTCCGACCACAAGGTGCGTTGTC	6	Wu & Tanksley, 1993
10	*Xa*-7/*xa*-8	3	RM251	STMS	F GAATGGCAATGGCGCTAG R ATGCGGTTCAAGATTCGATC	8	Chen *et al*, 1997a
11	*xa*-8	2	RM263	STMS	F CCCAGGCTAGCTCATGAACC R GCTACGTTTGAGCTACCACG	6	Chen *et al*, 1997a

considered as a standard, against which the sizes of the fragments of interest in the remaining lanes were calculated.

The fragments amplified using these primer pairs were scored as present (1) or absent (0) and the data entered into a binary matrix was analyzed using Winboot program (Yap and Nelson, 1996). Coefficients of similarity were calculated using Dice coefficient and cluster analysis was performed by agglomerative technique using the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) method. The relationships between the different rice lines studied were portrayed graphically in the form of a dendrogram.

3.3 RESULTS

3.3.1 Isolation and cloning of loci containing $(GATA)_n$ repeats

Screening of a genomic library of *Oryza sativa* var. Basmati-370 in Charon-40 vector with the probe $(GATA)_4$ resulted in 150 positive signals in the first round of screening. Of these, approximately 40 strongly hybridizing plaques were selected for subsequent screening and six clones that gave strong signals on hybridization with $(GATA)_4$, were randomly selected for further analysis. Since the six putative positive phage clones contained inserts greater than 10kb in size, further subcloning into pUC18 or pBluescript vector was performed, in order to obtain subclones for sequencing.

3.3.2 DNA sequence analysis

Sequence analysis revealed that the three subclones viz. OSIA6, OSIH10 and OS2E7 contained $(GATA)_n$ repeats and hence they were deposited in GenBank, with accession numbers, AF294726 (BankIt 351397), AF285877 (BankIt 345999) and AF 284424 (BankIt 345056), respectively. A close observation of the sequences of these three subclones revealed that the GATA repeat regions were of varying complexity and length, with 7 to 13 repeats which may be perfect or imperfect, according to the definition of Weber (1990b). The subclone OS1H10, of the phage clone GATA 3.1, contained seven perfect GATA repeats, while the subclone OS1A6, of the phage clone GATA 7.1, contained 7 imperfect repeats, with interruption of the repeat motif by TA (Table 3.5). The subclone OS2E7, of the phage clone

TABLE 3.5: Analysis of the three $(GATA)_n$-containing subclones with respect to repeat number, allele length and number of alleles

Sr. No.		OS1A6	OS1H10	OS2E7
1	Original phage clone number from which derived	GATA 7.1	GATA 3.1	GATA 8.1
2	Restriction endonuclease used to digest phage DNA	*Msp*I	*Taq*I	*Taq*I
3	Vector in which sub-cloned	pUC-18	pUC-18	pUC-18
4	Cloning site	*Acc*I	*Acc*I	*Acc*I
5	GenBank accession number	AF294726	AF285877	AF284424
6	Number of GATA motifs	7 imperfect **GATA TA $(GATA)_6$**	7 perfect **$(GATA)_7$**	13 imperfect **AATA $(GATA)_3$ GAA $(GATA)_4$ AGTA GAAA GATG AGTA**
7	Primer sequences used for PCR amplification	F GCTGCGCAGTGTATGCGTCT R ACCATTTAACAGCTAATCGA	F GAGTGGTCTCAGTTTGGTGC R TGGCGGCTGGAGAGCGTTTG	F AAGTGGCGACGTGAGAGCG R GTTAAGACAGTCCCCAGTG
8	Annealing temperature in PCR for primer annealing	55°C	67°C with 2% formamide	57°C with 2% formamide
9	Size of expected PCR product (from genomic subclone sequence)	184bp	263bp	141bp
10	Number and size range of alleles detected in			
	(A) Wild and cultivated rice genotypes	7 (184bp-208bp)	22 (116-521bp)	Unscorable, ladder-like pattern obtained
	(B) Indian elite rice varieties	3 (174-197bp)	10 (218-292bp)	4 (157-179bp)
	(C) Bacterial blight resistant and susceptible lines	5 (171-198bp)	19 (140-535bp)	6 (117-143bp)
	(D) Cereal genotypes	1 (184bp)	6 (148-270bp)	-
11	Maximum number of bands per lane in			
	(A) Wild and cultivated rice genotypes	2	7	-
	(B) Indian elite rice varieties	3	3	3
	(C) Bacterial blight resistant and susceptible lines	2	9	4
	(D) Cereal genotypes	1	3	-
12	Minimum number of bands per lane in			
	(A) Wild and cultivated rice genotypes	0	0	-
	(B) Indian elite rice varieties	1	0	0
	(C) Bacterial blight resistant and susceptible lines	1	0	0
	(D) Cereal genotypes	0	1	-

GATA 8.1, contained 13 imperfect repeats, with interruptions within the repeat motif (Table 3.5). In addition to GATA repeats, the subclone OS2E7 showed the presence of GATG, TATA and AATA repeat motifs, which may represent degenerated GATA sequences. Most of the degeneration could be explained by mutations of a single nucleotide within the GATA core sequence in the process of evolution. The OS2E7 subclone thus contained six GATA motifs, 5 GATA-like repeats, with one mismatch and 2 AGTA motifs.

The remaining three subclones did not contain GATA or its complementary CTAT repeat. Although dissociation temperature ($T_d = T_m - 7.6^oC$) (Rychlik and Rhoads, 1989) was useful to predict the appropriate washing temperature, it was found to be least reliable for motifs with a high AT or GC content (Panaud *et al*, 1995). In rice, Panaud *et al* (1995) observed that 11 out of 20 putative CGG-containing clones and 3 out of 20 TCT-containing clones did not contain the respective repeat motifs, although they did contain CG and AT rich regions, respectively. In our study too the three putative GATA clones contained AT rich regions, which could explain the positive hybridization signals at $(T_m - 5)^oC$ at these loci.

Sequence similarity search using BLAST (Altschul *et al*, 1990) through the NCBI, BLAST email server revealed that two of the subclones, OS1H10 and OS2E7, showed significant but quite short homology to nucleotide sequences in the GenBank database. The subclone, OS1H10, had 95% homology to a 43bp segment of the *Oryza longistaminata* receptor kinase-like protein family member D (Acc # U72726) and 90% homology to a 39bp fragment of the *Oryza sativa* pathogenesis-related protein class 1 (PR-1) gene (Acc # U89895). The subclone, OS2E7, showed 84% similarity to a 56bp region of the *Oryza sativa* pathogenesis-related protein class 1 (PR-1) gene (Acc # U89895) and 94% homology to a 35bp stretch of the gene encoding for *Oryza longistaminata* receptor kinase-like protein family member D (Acc # U72726). In both the subclones, the segments showing homology to genes for receptor kinase-like protein and pathogenesis-related protein were found to be overlapping each other, indicating similar regions within these two genes. The BLAST searches were based on full-length sequences from which the primer pairs were designed and were not just from the amplified sequences. When the sequences of these two subclones were compared,

only a 67bp region towards the end of subclone, OS1H10 was similar to a 66bp region in the middle of subclone, OS2E7. Since the remaining sequence of both these subclones was different and the region showing similarity was in the middle of OS2E7 subclone, these two subclones might be entirely different. Primers flanking the $(GATA)_n$ repeat were synthesized (Table 3.5) and used to amplify various wild and cultivated genotypes of rice, Indian elite rice varieties and bacterial blight resistant and susceptible lines of rice as detailed earlier.

3.3.3 $(GATA)_n$ motifs in the rice and *Arabidopsis* genome

When $(GATA)_4$ was used as a query sequence in BLAST, GATA motifs were found distributed on all the chromosomes sequenced in rice and *Arabidopsis* genomes. The largest stretches of $(GATA)_{13}$ and $(GATA)_6GAA(GATA)_2$ were found on the *adh1-adh2* BAC (AF172282) in the intron of leucine rich repeat containing protein kinase gene, while another region on the rice chromosome 10 BAC (AC016780) had eight CTAT motifs. Several other regions had smaller number of GATA repeats. Six rice ESTs were found to have $(GATA)_n$ stretches, the largest ones being $(GATA)_{12}$ and $(GATA)_7$, present in rice seed (BE229156) and panicle (AU094720) ESTs, respectively. Since the rice genome sequencing is not completed as yet, it is possible that many other regions of the genome could be harboring GATA repeats. A search of the *Arabidopsis* genome, which is the only plant genome that has been completely sequenced (The Arabidopsis Genome Initiative, 2000), showed a similar pattern as compared to rice with a continuous stretch of $(GATA)_{14}$ in chromosome 4 BAC (AL161581) and $(CTAT)_7TTAT(CTAT)_8$ in a different chromosome 4 BAC (AL0225371). *Arabidopsis* also showed $(GATA)_4$ atleast once in six ESTs, the largest stretch being $(GATA)_5$ (AV536635). The longest stretch of GATA repeats found in plants in the EST database was from *Sorghum* (BE356663), which had $(CTAT)_{15}$.

3.3.4 Germplasm analysis using primers flanking the $(GATA)_n$ repeat region

When different rice genotypes were amplified using the primers flanking the three different (GATA)n-containing subclones, it was observed

that these primer pairs generated large amount of polymorphism even in closely related elite rice germplasm of India. Amplification was performed at 55^{o}C without formamide and 67^{o}C and 57^{o}C with 2% formamide using primers flanking the subclones, OS1A6, OS1H10 and OS2E7, respectively.

[1] Analysis of distantly related germplasm

Using OS1A6 flanking primers, seven alleles ranging in size from 184bp to 208bp (Fig. 3.1) were observed in the wild and cultivated rice germplasm set (Fig. 3.1) possibly suggesting a difference of six tetranucleotide repeats between the smallest and the largest alleles (Table 3.5). 22 alleles (116bp - 521bp) were detected with OS1H10 flanking primers on the above germplasm (Table 3.5). On the other hand, OS2E7 flanking primers detected a ladder-like pattern with large number of bands, suggesting the presence of many alleles, which differ in the number of GATA repeats. When representative cereal genotypes of oat, barley, maize and wheat were studied, amplification was obtained in all the cereals using OS1H10 flanking primers while only the maize genotype, *Zea diploperensis*, gave amplification with OS1A6 flanking primers. When the wild and cultivated germplasm of rice and other cereals were considered together (Table 3.1), an average of 15.5 alleles was detected over the two scorable loci, OS1A6 and OS1H10. However, when the rice genotypes were considered alone, an average of 14.5 alleles were detected, while an average of 3.5 alleles was observed in the cereal genotypes studied, considering these two loci.

[2] Analysis of closely related Indian elite rice varieties

When the $(GATA)_n$ flanking primers were used to amplify 47 Indian elite rice varieties (Table 3.2), 3 (174bp -197bp), 10 (218bp -292 bp) and 4 (157bp - 179bp) alleles were detected at the OS1A6, OS1H10 and OS2E7 (Fig. 3.2) loci, respectively (Table 3.5). Thus, even in the case of closely related germplasm, the three microsatellite loci were found to be quite polymorphic with an average of 5.67 alleles per locus. Although the 47 Indian elite rice varieties have a narrow genetic base (section 2.4.4 of chapter 2), the three microsatellite loci, when used together, could distinguish about three-fourths of the elite rice cultivars unambiguously.

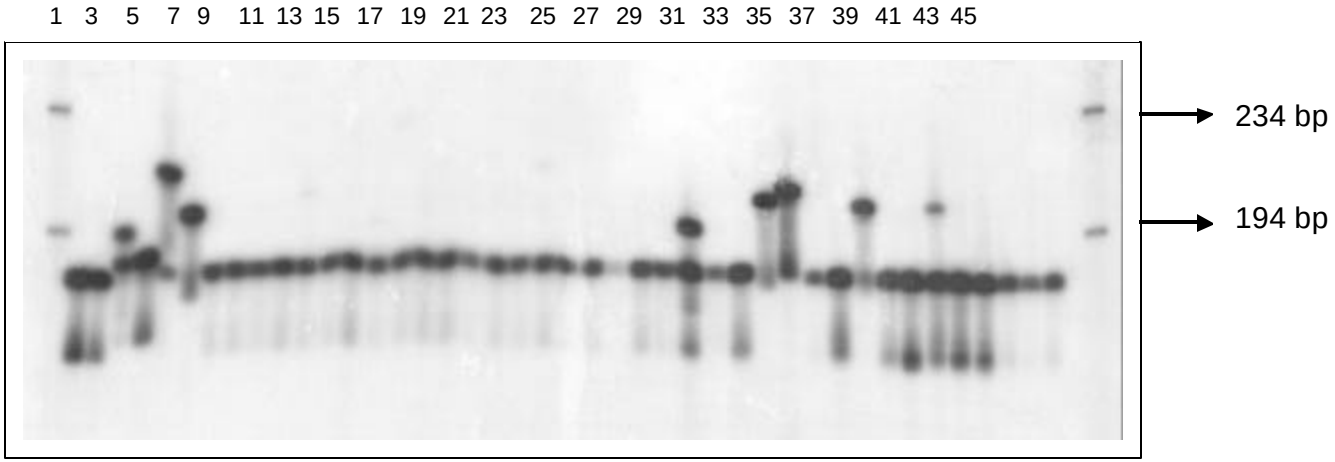


Figure 3.1: DNA of wild and cultivated rice genotypes amplified with OS1A6 flanking primers. Lanes 2-43 are samples 1-42 as listed in Table 3.1. Lane 44 is the negative control without DNA. Lanes 1 and 45 are ϕX174 DNA / *Hae*III marker.

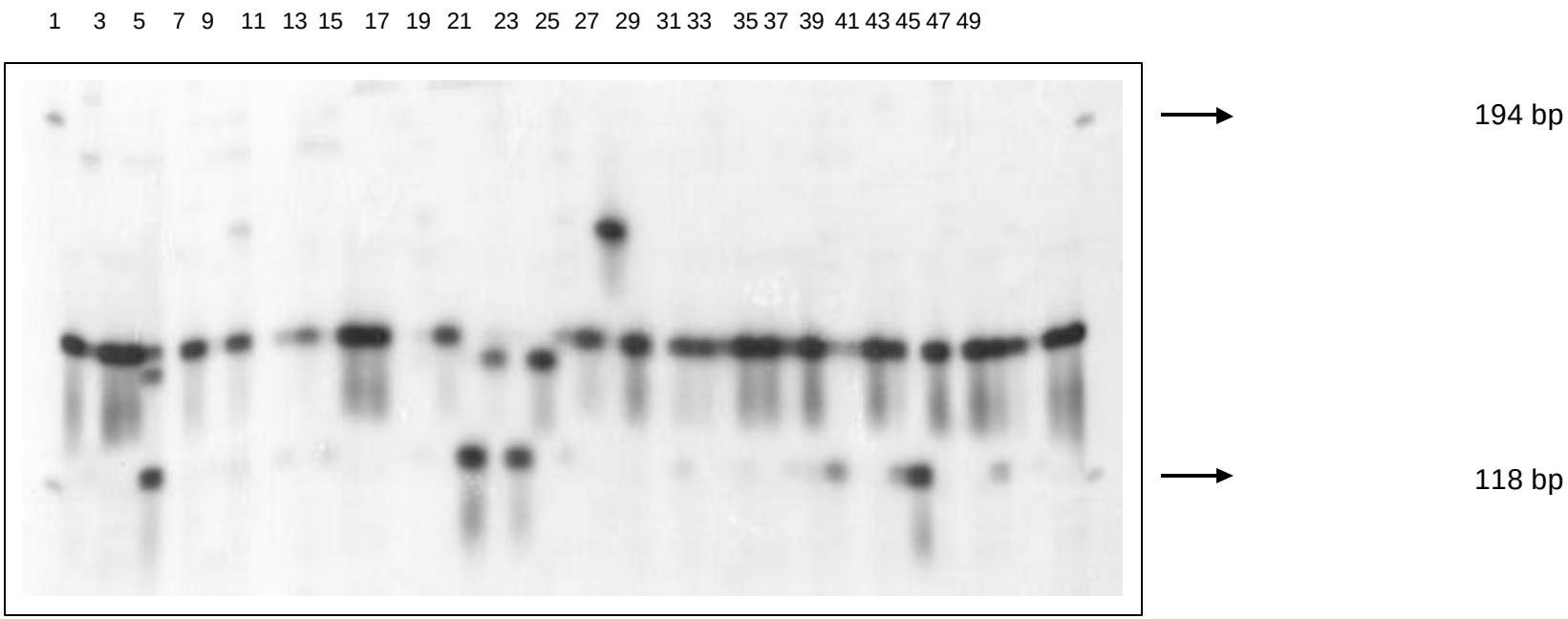


Figure 3.2: DNA of Indian elite rice cultivars amplified with OS2E7 flanking primers. Lanes 2-48 are samples 1-47 as listed in Table 3.2. Lanes 1 and 49 are ϕX174 DNA / *Hae*III marker.

[3] Analysis of bacterial blight resistant / susceptible germplasm

Thirty-seven bacterial blight resistant and susceptible lines (Table 3.3) resulted in 5, 19 and 6 alleles ranging in size from 171-198bp, 140-535bp and 117-143bp, respectively, using primers flanking the OS1A6, OS1H10 and OS2E7 loci, respectively (Table 3.5), with an average of 10 alleles per locus. In all the cases, the number of alleles detected was excluding null alleles.

3.3.5 Cluster analysis based on the $(GATA)_n$-containing loci

A dendrogram was constructed based on the polymorphism data obtained at the three $(GATA)_n$ microsatellite loci in a few representative genotypes of the three sets of germplasm analyzed. The dendrogram revealed that all the bacterial blight resistant lines clustered together, while the remaining genotypes clustered separately (data not shown). Since the bacterial blight resistant lines were found clustered in a separate group, another dendrogram was constructed using the patterns obtained at the three microsatellite loci for all the 37 lines which were either resistant or susceptible to bacterial blight. The dendrogram thus obtained (Fig. 3.3) revealed the presence of six clusters, wherein the lines grouped together based on the resistance genes present in them. In cluster 1, with the exception of IR22, the remaining three lines were near isogenic to IR-24. The first subgroup of cluster 2 also contained all near-isogenic lines with the exception of Rasi, which joined the cluster separately. The second and third subgroups of the second cluster were heterogeneous, containing lines with *Xa*-4 and *xa*-5 resistance genes and one line BPT-5204, which has no resistance genes. Except Swarna and Nigeria-5, in clusters 3 and 6, respectively, all the lines in the third and sixth clusters contained the *xa*-5 gene. All the *Xa*-4 containing lines with the exception of T(N)1 in cluster 4 and Kuntalan in cluster 5, grouped together in clusters 4 and 5. In cluster 5, in addition to *Xa*-4 gene, the lines CRRI-1, CRRI-3 and IET-14444 also contained the *xa*-5 gene. Interestingly, Jaya, which contains no resistance genes totally outgrouped from the remaining lines.

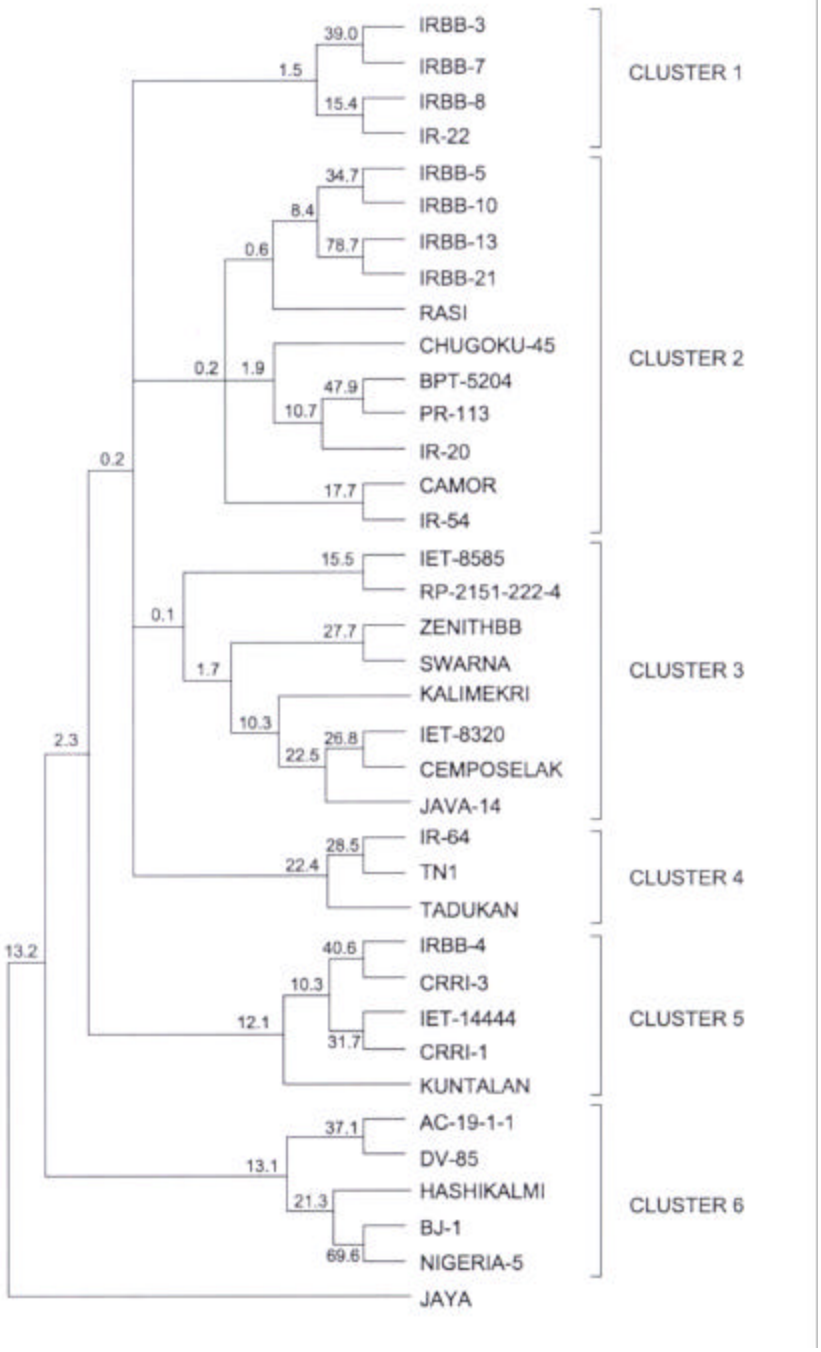


Figure 3.4: Dendrogram showing the genetic relationships between different near-isogenic lines, bacterial blight resistant and susceptible lines based on the polymorphism data obtained at 11mapped microsatellite loci which are either linked or close to the bacterial blight resistance genes. The numbers at the forks indicate the confidence limits for the grouping of those species, which are to the right of that fork.

3.3.6 Analysis of bacterial blight resistant / susceptible lines using mapped STMS markers

Since the bacterial blight resistant lines were observed to group together based on the resistant genes present in them (Fig. 3.3), I analyzed these lines using 11 mapped STMS markers (Chen *et al*, 1997a). Of these, three microsatellite markers RM122, RM390 and RM13 have been shown to be linked to *xa-5* gene at distances of 0.4cM, 0.8cM and 14.1cM, respectively (Blair and McCouch, 1997; Sanchez *et al*, 1997), while the remaining eight STMS markers were close to the bacterial blight resistance genes in the linkage map but not studied for co-segregation analysis so far. The 11 STMS markers represented five different rice chromosomes and targeted eight different bacterial blight resistance genes (Table 3.4). The microsatellite marker RM21 located on chromosome 11, had the maximum number of 11 alleles [ranging in size from 186bp to 152bp] followed by RM206 on the same chromosome with 10 alleles [from 231bp to 197bp] and RM224 and RM251 on chromosomes 11 and 3, respectively with eight alleles each (Tables 3.4). The remaining STMS markers resulted in 6 (STMS markers: RM13, RM122, RM167, and RM263), 4 (STMS markers: RM230, RM254) and 2 (STMS marker: RM390) alleles, respectively (Table 3.4). Most of the rice lines studied resulted in a single allele per microsatellite primer pair, however, a few of the lines amplified two alleles with a single STMS marker and IR-22 resulted in 4 bands on amplification with RM206.

A dendrogram based on the data generated by using 11 STMS markers (Fig. 3.4) revealed that all the near-isogenic lines, except IRRB-5, clustered together in one group. IRBB*5* grouped together in the second cluster with CRRI-1, CRRI-3, RP-2151-222-4, IET-8320, IET-14444 and IET-8585, all of which contain multiple genes, with one of them being *xa-5*. In addition, all the lines in cluster 5 also contained the *xa-5* gene with the exception of Kuntalan. Except IR-22 and PR-113, all the lines in the first and second sub-groups of the third cluster, 3A and 3B, were susceptible to bacterial blight, while the lines in the third sub-group of the same cluster, 3C, mainly contained *Xa-4* gene with the exception of Swarna. Thus, the dendrogram obtained revealed that the lines clustered together on the basis of the resistance genes present in them.

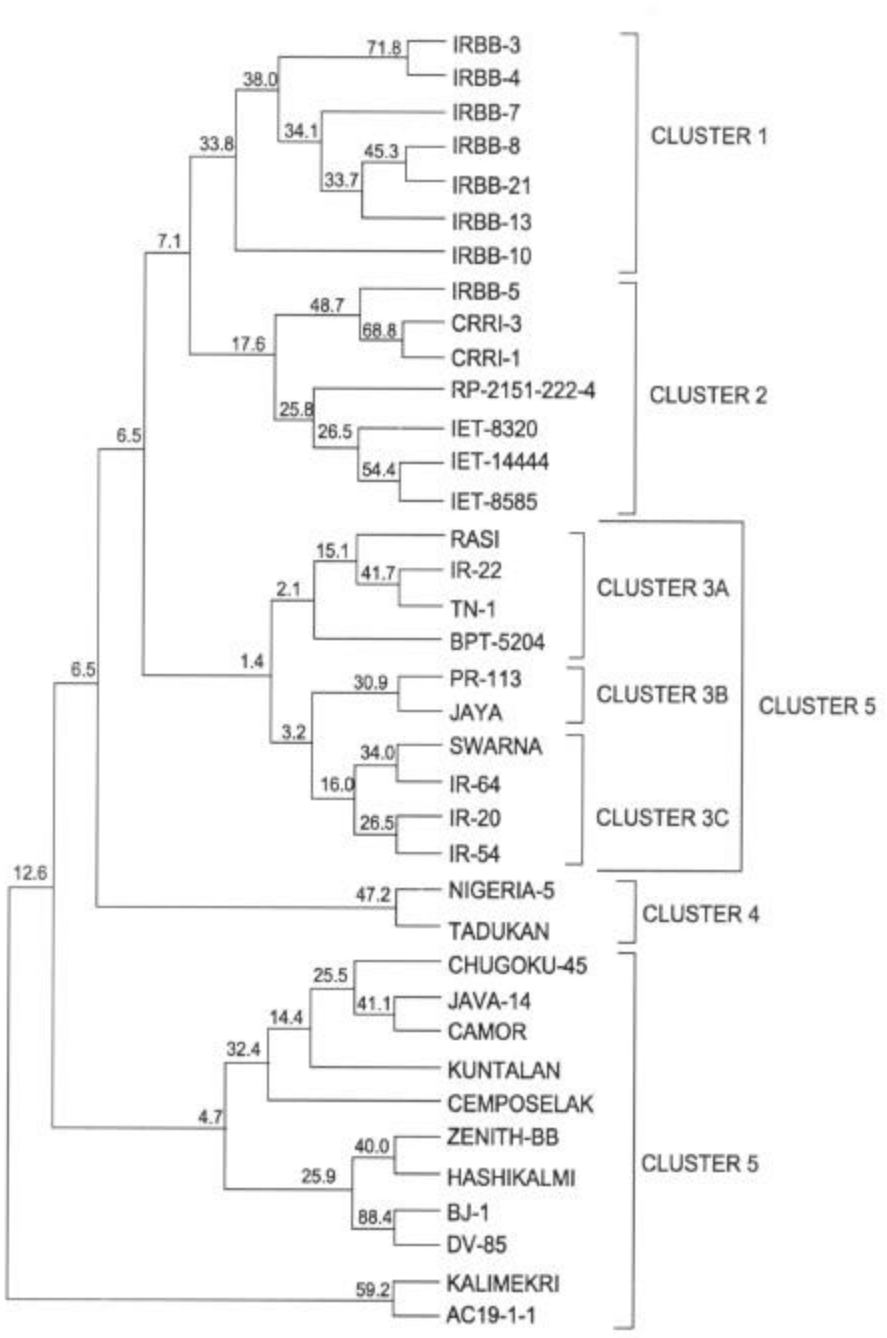


Figure 3.4: Dendrogram showing the genetic relationships between different near-isogenic lines, bacterial blight resistant and susceptible lines based on the polymorphism data obtained at 11mapped microsatellite loci which are either linked or close to the bacterial blight resistance genes. The numbers at the forks indicate the confidence limits for the grouping of those species, which are to the right of that fork.

3.4 DISCUSSION

3.4.1 Tetranucleotide $(GATA)_n$: a source of new STMS markers in rice

Rice has one of the smallest genome sizes among plants with a lower frequency of repeated sequences which necessitates the use of a variety of different SSRs to completely saturate its linkage map. Till now, three hundred and twenty three microsatellite markers, identified by library screening (Zhao and Kochert, 1992; 1993b; Wu and Tanksley, 1993; Akagi *et al*, 1996a; Panaud *et al*, 1996; Chen *et al*, 1997a; Cho *et al*, 2000; Temnykh *et al*, 2000) and GenBank searches of rice sequences (Wu and Tanksley, 1993; Wang *et al*, 1994c; Akagi *et al*, 1996a; Cho *et al*, 2000; Temnykh *et al*, 2000), have been mapped on the rice genetic map. Here, most of the markers contain dinucleotides and trinucleotides and only seven loci contain tetranucleotide repeats such as $(AATT)_n$, $(TTCC)_n$, $(GATG)_n$, $(ATGT)_n$, $(GTAT)_n$, $(ATTT)_n$, $(TTTG)_n$ (Wu and Tanksley, 1993; Akagi *et al*, 1996a; Panaud *et al*, 1996; Chen *et al*, 1997a; Cho *et al*, 2000; Temnykh *et al*, 2000). $(GATA)_n$ sequences which have not been mapped so far in rice, are the most frequent among tetranucleotide repeats with 270 poly $(GATA)_n$ motifs in the entire rice genome (Panaud *et al*, 1995). The degree of polymorphism shown by microsatellites rather than their abundance in the genome contributes to their usefulness in genome analysis. Considering the high utility of $(GATA)_n$ repeats for fingerprinting plant genomes (Epplen, 1992), I thought it would be interesting to isolate these $(GATA)_n$-containing regions from rice and convert them into PCR amplifiable microsatellites to determine variations at these loci.

$GA^T/_CA$ repeats are found to be sex-specifically accumulated on the Y chromosome in mouse, at the base of the X chromosome in *Drosophila melanogaster* and in short stretches on the human Y chromosome (Epplen, 1988). Simple $GA^T/_CA$ repeats have been first isolated and sequenced from a female-specific satellite DNA library of the colubrid snake, *Elaphe radiata* (Epplen *et al*, 1982). Thereafter, the GATA-containing regions from mouse (Singh *et al*, 1984), humans (Erickson *et al*, 1988), *Drosophila melanogaster* (Simpson, 1990) and tomato (Phillips *et al*, 1994; Vosman and Arens, 1997) have been sequenced. In most of these cases, the GATA repeat seems to be degenerate and contains tetranucleotides that deviate from GATA by a single point mutation as seen in the case of the $(GATA)_n$-containing clone OS2E7 in

the present work (Table 3.5). The accumulation of a large number of apparent point mutations in these sequences indicates a possibility that the GATA repeat clusters are evolutionarily very old.

3.4.2 $(GATA)_n$ loci: tools for intraspecific variability in rice

Several researchers in their allelic diversity studies of rice (Wu and Tanksley, 1993; Yang *et al*, 1994; Panaud *et al*, 1996; Chen *et al*, 1997a; Cho *et al*, 2000) have revealed the presence of 2-25 alleles per SSLP locus as compared to 2-4 alleles per RFLP locus in cultivated indica and japonica germplasm (McCouch *et al*, 1997). In the present study too, upto 19 alleles are detected in *Oryza sativa* cultivars at the three $(GATA)_n$-containing loci. Even in the closely related Indian elite rice cultivars, 3-10 alleles are obtained indicating that microsatellite based PCR markers are very useful in studying variability among genotypes having a narrow genetic base. The main reason for this high degree of polymorphism at the intraspecific level is due to the high mutation rate at microsatellite loci, which results in additions and deletions of repeat motifs.

My studies on cultivated germplasm of rice could not detect any obvious correlation between the repeat length and polymorphism rate, which was also observed by Bell and Ecker (1994) in *Arabidopsis*. I have observed that the subclone OS1H10, with 7 perfect repeats results in the maximum number of alleles viz. 10, while OS1A6 and OS2E7 with 7 and 13 imperfect repeats, respectively result in 3 and 4 alleles, respectively, indicating that the higher mutation rate within perfect repeats could be responsible for generating greater number of alleles at such loci. Several workers have shown that pure repeats are unstable, with stepwise mutations and deletions increasing with the size of the pure repeat (Amos *et al*, 1996; Primmer *et al*, 1996b; Wierdl *et al*, 1997). Impure repeats, on the other hand, are possibly stabilized due to interruptions within the repeat motif as observed by Petes *et al* (1997) in yeast, where interruptions in the GT repeats decreased mutation rates five fold. My study shows that pure repeats may represent a transient evolutionary state supporting the hypothesis of Peakall *et al* (1998).

In humans, SSRs with fewer than 10 repeat motifs tend to show little if any polymorphism (Weber, 1990a; Garza *et al*, 1995). However, my studies

on rice have revealed that SSRs with less number of repeat motifs also can be extremely informative in plants. In many other crop systems like *Beta*, *Arabis*, *Lycopersicon* and soybean, polymorphism is observed, despite the small number of repeats per SSR locus (Morchen *et al*, 1996; van Truren *et al*, 1997; Smulders *et al*, 1997; Peakall *et al*, 1998), suggesting that microsatellite loci with small number of repeats can be potentially informative in plants and should not be ignored.

3.4.3 Cross species amplification at $(GATA)_n$ loci

Wu and Tanksley (1993) and Panaud *et al* (1996) have used STMS markers to amplify DNA of wild *Oryza* species, indicating their utility in analyzing inter-specific crosses. In my work, I have observed that STMS markers resulted in a higher level of variation as revealed by a greater number of alleles, when used to amplify genomic DNA of wild rice accessions as compared to *O. sativa* genotypes. When wild germplasm of rice has been analyzed 7-22 alleles have been detected at two loci viz. OS1A6 and OS1H10, while the locus OS2E7 resulted in a ladder-like pattern indicating the extremely high level of polymorphism detected at microsatellite loci. Provan *et al* (1997) have observed that amplification of wild rice accessions with primers flanking mononucleotide repeats in rice chloroplast genome, which are analogous to nuclear SSRs, has resulted in higher diversity values as compared to *O. sativa* accessions.

There are several examples of cross species amplification of SSR loci in animal studies, where transferability of these loci in divergent taxa has been reported. For example, Rico *et al* (1996) have amplified 17 SSR loci across fish that had diverged as long as 470Myr while Fitzsimmons *et al* (1995) have observed that the SSR loci from marine turtles could amplify in freshwater turtles separated by 300Myr. Schlotterer *et al* (1991) have observed that the cetacean SSR loci are conserved over 40Myr. Such a phenomenon has also been observed in seals (Coltman *et al*, 1996; Gemmell *et al*, 1997), birds (Primmer et al, 1996a), primates (Blanquer-Maumont and Crouau-Roy, 1995; Garza *et al*, 1995), rodents (Kondo *et al*, 1993) and artiodactyls (Engel *et al*, 1996). Conservation of microsatellite-containing loci in diverse mammalian genomes (Stallings, 1995) has facilitated the use of these heterologous primer

pairs for amplification of closely related species (Moore *et al*, 1991). In plants also, although reports on cross-species amplification are fewer, similar observations are available in different species of *Cicer* (Udupa et al, 1999) and *Glycine* (Peakall et al, 1998). Using soybean SSR flanking primers, Peakall *et al* (1998) have amplified DNA from several legume genera such as *Albizia*, *Vigna*, *Trifolium*, etc. Weising and Gardner (1999) have observed chloroplast DNA variation in members of Solanaceae and most other angiosperm species when they used primers flanking microsatellite motifs in tobacco. The *O.sativa* primers obtained in the present study show amplification not only in different *Oryza* species but also in representative cereal genera viz. maize, wheat, barley and oat, indicating that they may be located in the conserved regions of the genome. Very few primer sequences from rice have been shown to be conserved enough to amplify in related genera, although some have been observed to amplify DNA of wheat, oat, barley, maize, *Sorghum*, millet, sugarcane and bamboo reliably (Zhao and Kochert, 1993b; McCouch *et al*, 1997). Roder *et al* (1995), have used microsatellites from wheat to amplify barley and rye DNA. Recently, Ishii and McCouch (2000) have observed microsynteny at microsatellite loci in the chloroplast genomes of rice and maize. Primers flanking the above loci have been used to amplify wild and cultivated accessions of rice and eight other members of the family Gramineae viz. maize, teosinte, wheat, oat, barley, pearl millet, *Sorghum* and sugarcane. All the above examples indicate that microsatellite loci could be useful for comparative genome mapping in cereals.

Microsatellite-containing loci with greater number of repeats are more likely to be polymorphic (Weber, 1990a) and the high level of polymorphism at such loci can be attributed to two molecular mechanisms, viz. replication slippage or unequal crossing over. Levinson and Gutman (1987), have proposed that if replication slippage is an important mechanism, a longer repeat would tend to show more variation, since the chance of replication errors is higher for a longer stretch of repeated sequence. My studies on wild genotypes have revealed a significant positive correlation between the average number of repeats and the amount of variation. The OS2E7 locus with 13 imperfect repeats, has resulted in a large number of alleles, which

could not be scored. Similarly, 22 alleles have been obtained at OS1H10 locus with 7 perfect repeats and OS1A6 with 7 imperfect repeats has resulted in 7 alleles. All this data indicates that replication slippage must be the molecular mechanism responsible for generating new alleles at these SSR loci (Levinson and Gutman, 1987; Wolff *et al*, 1991; Innan *et al*, 1997). Such a phenomenon has also been observed in the case of $(TAA)_n$-containing loci in chickpea (Udupa *et al*, 1999).

3.4.4 Utility of microsatellites in coding and non-coding regions

Two of the loci reported here have shown significant homology to short stretches of the genes encoding for *Oryza longistaminata* receptor kinase-like protein and *Oryza sativa* pathogenesis related-protein. Further, these homologous regions lie within the OS1H10 and OS2E7 loci that are amplified in PCR, 4bp prior to the $(GATA)_n$ repeat in case of OS2E7 and 5bp after the $(GATA)_n$ repeat in OS1H10. Also, *Xa21*, which is one of the genes responsible for imparting resistance to bacterial blight in rice, encodes a putative receptor kinase (Song *et al*, 1995; Ronald, 1997). $(CGG)_n$ repeats have been found in the exonic region of the receptor kinase-like protein, *Xa21* (Cho et al, 2000). Considering these factors, I conclude that the reason for the bacterial blight resistance lines grouping together in one cluster, distinct from the other clusters containing the remaining genotypes, is the short but significant homology of these $(GATA)_n$ loci to the *Oryza longistaminata* receptor kinase-like protein.

Using primers flanking the $(GATA)_n$-containing microsatellite loci, I also observe that the bacterial blight resistant lines cluster into different groups based on the resistance genes present in them (Fig. 3.3). In order to test if this is an isolated phenomenon, I have further analyzed the same set of 37 rice genotypes using 11 previously mapped STMS markers either linked or close to the bacterial blight resistance genes. Interestingly, in this analysis too, all the lines group together according to the resistance genes present in them (Fig. 3.4). The present analysis probably suggests that these $(GATA)_n$-containing microsatellite loci maybe associated with disease resistance and could probably be responsible for selection of this resistance gene function. I further postulate that these loci may have evolved along with the *Xa* genes

leading to clustering of the lines according to the bacterial blight resistance genes present in them and suggest a possible future diagnostic potential of these $(GATA)_n$ microsatellite loci.

Tautz *et al* (1986) have shown that there is a frequent occurrence of nearly all possible SSR motif arrays in both coding and non-coding regions of eukaryotic genomes. Microsatellite repeats have been found in several cloned and completely sequenced genes of rice and many of these microsatellite loci have been mapped (Cho *et al*, 2000; Wu and Tanksley, 1993). Dinucleotide repeats of RM120 are present in the 3' untranslated region (UTR) of the rice *Adh-2* gene for alcohol dehydrogenase (Xie and Wu, 1989), while those of RM203 are found upstream of the *phy-18* gene for phytochrome in rice (Kay *et al,* 1989). Genetic mapping of these two microsatellite loci has resulted in simultaneous mapping of these two genes onto the rice genetic map (Wu and Tanksley, 1993), indicating that flanking microsatellite polymorphisms can be used to map known genes even when there may not be determinable polymorphism using coding sequences as probes. The screening of a cDNA library has revealed the presence of seven out of twelve surveyed SSRs in the transcribed regions in rice (Panaud *et al,* 1995). SSRs have also been found linked to several genes of agronomic importance in rice. Expansion of $(CT)_n$ repeats at the microsatellite locus, RM190, located at the 5' UTR of the *waxy* gene (Bligh *et al,* 1995) has been implicated in phenotypic variation for starch quality in commercial US rice cultivars (Ayres et al, 1997). Two microsatellite markers have been found to be closely linked to the giant embryo gene ge^s (Koh *et al,* 1996), responsible for controlling enlarged embryo size and high protein, oil and vitamin content in rice, while three markers have shown linkage to the *xa-5* gene, responsible for imparting resistance to bacterial blight in rice (Blair and McCouch, 1997). On chromosome 1, two STMS markers are found to be located in the vicinity of a QTL for rice yield (Xiao *et al,* 1996a).

The availability of greater number of SSLP markers will help to saturate the existing RFLP, SSLP and AFLP maps of rice, thus increasing the power of genome analysis and to fine map different genes and QTLs of agronomic and economic importance in rice. The clustering of bacterial blight resistant genotypes using microsatellite markers in our study postulates the diagnostic

potential of such markers which can be exploited in marker-assisted breeding programs. Furthermore, the ability of a few characterized microsatellite loci to amplify in different cereal genera could be useful for synteny mapping when they are mapped onto orthologous regions of different cereal genomes.

CHAPTER 4

Expansion of $(GA)_n$ Dinucleotide at a Microsatellite Locus Associated with Domestication in Rice

The contents of this chapter have been published as a full-length paper in Biochemical Genetics (1998) 36: 323-327

ABSTRACT

Microsatellites undergo rapid changes over short evolutionary time periods, which can be phylogenetically informative in related species. When nine cultivated and seven wild species of rice were analyzed using the highly polymorphic $(GA)_n$-containing locus, expansion of this locus was observed in the process of domestication of rice from its wild ancestors. DNA sequencing of the amplified loci revealed an increase in repeat number from 14 in distantly related wild species to 20-24 in the widely grown present day indica cultivars. Expansion of $(GA)_n$ dinucleotide associated with domestication of rice is the first observation of its kind in plants in contrast to the repeat expansion associated with genetic disorders in humans.

4.1 INTRODUCTION

Microsatellites, which are considered to be the most informative genetic markers (Tautz *et al*, 1986) have become extremely popular for inter-population (Gilbert *et al*, 1990; Edwards *et al*, 1992; Chakraborty *et al*, 1992) and molecular evolution (Bowcock *et al*, 1994) studies. They seem to be maintaining themselves by their ability to replicate within the genome and play an important role in genome evolution by increasing in a concerted manner alongwith the increasing genome size (Hancock, 1996). One of the major mechanisms for microsatellite repeat expansion is replication slippage (Levinson and Gutman, 1987; Charlesworth *et al*, 1994). However, although much has been written about the probable mechanisms causing length variations within species (Kuhl and Caskey, 1993; Charlesworth *et al*, 1994, Rubinstein *et al*, 1995), little is known about the evolutionary origin of microsatellites (Valdes *et al*, 1993). Of the different SSRs, dinucleotide repeats have a special importance due to their abundant distribution and highly polymorphic nature.

The expansion of triplet repeats has been associated with genetic disorders in humans (Caskey *et al*, 1992; Morin *et al*, 1994) and is currently known to cause 14 neurological diseases (Cummings and Zoghbi, 2000). This phenomenon is, therefore, being used to predict disease occurrence in prenatal diagnosis and to isolate novel genes associated with genetic disorders. The role of microsatellite motifs in coding and non-coding regions of the rice genome has been discussed in the previous chapter (section 3.4.4 in chapter 3). I have also observed that allelic variations at $(GATA)_n$ containing microsatellite loci were responsible for grouping different rice genotypes according to the bacterial blight resistance genes present in them (section 3.3.5 in chapter 3). The earlier work in our laboratory has shown the presence of tissue-specific transcripts in rice, homologous to the triplet repeat probe, $(CAC)_5$, suggesting a possible biological role of microsatellites in plants (Gupta *et al*, 1994). However, association of allelic variation at microsatellite loci with any form of diagnostic pathology has not yet been reported in plants.

Since rice has the smallest genome size among cereals, it is considered to be an ideal crop to identify significant steps in genome evolution. The evolution of rice and the process of its domestication have

fascinated researchers all over the world for several decades. There were many controversies concerning the evolutionary process of the two cultivated species, which mainly stemmed from differences in naming and characterizing the putative progenitor. Watt (1891) was the first to postulate that the wild rice, *O.rufipogon*, may be the putative progenitor of cultivated rice, *O.sativa*, which is now a well documented fact (Chang, 1976b; Khush, 1997). Several workers have used molecular markers to show that *O.sativa* has evolved from *O.rufipogon* as reviewed in chapter 1 (section 1.3.3). Using microsatellites, Yang *et al* (1994) have established that 28% of allelic diversity has been lost during the process of cultivar development from landraces. In order to improve our understanding of the domestication process involved in the evolution of rice, it will be very interesting to study the dynamic changes occurring in microsatellite sequences and to evaluate how these changes can give useful information for phylogenetic studies. Keeping this interest in mind, I have analyzed the highly polymorphic $(GA)_n$-containing microsatellite locus, RM122 in nine cultivated and seven representative wild rice species.

4.2 MATERIALS AND METHODS

4.2.1 Plant material and DNA extraction

Seeds of nine different cultivars of rice and seven of its wild relatives (Table 4.1) were obtained from authentic sources such as Directorate of Rice Research, Hyderabad, India, Central Rice Research Institute, Cuttack, India and Agricultural Research Station, Vadgaon, India. Total genomic DNA was extracted from leaves by CTAB method as described previously in chapter 2 (section 2.2.1).

4.2.2 PCR amplification

Primers flanking the $(GA)_n$ harboring microsatellite locus, RM122 (Wu and Tanksley, 1993) were synthesized at the primer synthesis facility at National Chemical Laboratory, Pune, India. PCR was performed in 100µl volume containing 200µM of dNTP, 0.15µM of each primer, 500ng of genomic DNA, 2.5U Taq DNA polymerase (Boehringer Mannheim, Germany) in 10mM Tris-HCl (pH 8.3), 1.5mM $MgCl_2$ and 50mM KCl. DNA amplifications were

Table 4.1: Wild and cultivated genotypes of rice analyzed at the microsatellite -containing locus, RM122

Sr. No.	Genotype	Genome type	Band size in bp	Number of (GA) repeat motifs	Sub-species
1	*O.rufipogon*	AA	223	18	-
2	*O.longistaminata*	AA	215	14	-
3	*O.officinalis*	CC	219	16	-
4	*O.eichengeri*	CC	215	14	-
5	*O.alta*	CCDD	215	14	-
6	*O.australiensis*	EE	215	14	-
7	*O.brachyantha*	FF	215	14	-
8	Tambdarambhog	AA	235	24	Indica 1
9	Ambemohr	AA	235	24	Indica 1
10	Indrayani	AA	227	20	Indica 2
11	Malkolam	AA	227	20	Indica 2
12	Vargalsal	AA	227	20	Indica 2
13	Basmati-370	AA	227	20	Indica 2
14	Fujisaka	AA	231	22	Japonica
15	Norin-49	AA	231	22	Japonica
16	Taichung-65	AA	231	22	Japonica

carried out in a Perkin Elmer Cetus thermal cycler (Perkin Elmer Inc., U.S.A.) with an initial denaturation at 94^0C for 4 min followed by 30 cycles of denaturation at 94^0C for 1 min, primer annealing at 55^oC for 1 min and primer extension at 72^oC for 2 min, with a final extension at 72^oC for 5 min. An aliquot of the amplified product was checked on 1.2% agarose gel in TAE buffer (40mM Tris-acetate, 1mM EDTA, pH 8.0).

4.2.3 Sequencing of the PCR amplified product

The amplicon obtained on PCR was sequenced using the Sequenase version 2.0 PCR product sequencing kit (U.S. Biochemical Corp., U.K.). 6µl of PCR product was incubated with 1µl of exonuclease (10U/µl) at 37^oC for 15 min, following which the exonuclease was inactivated by heating to 80^oC for 15 min and transferring to ice for 5 min. The mixture was then treated with 1µl of shrimp alkaline phosphatase (2U/µl) as above, followed by inactivation of the enzyme as before. 5-10 pmoles of primer was added to 8µl of the treated PCR product and the volume was made up to 10µl using sterile distilled water. The mix was denatured by boiling for 2-3 min and then immediately chilled on ice for 5 min. To this ice-cold annealed DNA mix, 2µl of 5X reaction buffer, 1µl of 0.1M DTT, 2µl of 1:5 diluted labeling mix, 10µCi of α^{32}P dATP and 2µl of Sequenase version 2.0 T7 DNA polymerase (1.6U/µl) were added, mixed well and incubated at 16-20^oC for 5 min. Then 3.5µl of this mix was added to four tubes containing 2.5µl of the respective dideoxynucleotide and incubated at 37^oC for 5 min for chain termination. Finally, 4µl of stop solution was added to terminate the reaction. Gel loading, electrophoresis and autoradiography were performed as described for regular sequencing reaction in chapter 3 (section 3.2.3).

4.2.4 Data analysis

The $(GA)_n$ repeat region from different wild and cultivated genotypes of rice was aligned using the CLUSTAL W package (Thompson *et al*, 1994). The sequences flanking the microsatellite were conserved and were, therefore, not included in the multiple sequence alignment.

4.3 RESULTS

Nine cultivated and seven representative wild rice species (Table 4.1) were analyzed at the highly polymorphic $(GA)_n$-containing microsatellite locus, RM122, using primers flanking the locus.

4.3.1 Genetic diversity at the microsatellite locus, RM122

When the different wild and cultivated genotypes of rice were amplified using primers flanking the $(GA)_n$-containing microsatellite locus, RM122 (Wu and Tanksley, 1993), bands of varying sizes were obtained (Table 4.1). All the wild rice genotypes viz. *O.longistaminata*, *O.eichengeri*, *O.alta*, *O.australiensis* and *O.brachyantha*, resulted in a band of 215bp with the exception of *O.officinalis* and *O.rufipogon*, which showed bands of 219bp and 223bp, respectively. All the cultivated genotypes resulted in bands of higher molecular weight than those in the wild genera. The indica rices could be grouped into two different groups depending on the size of the amplicons. The first group called indica 1, included Tambdarambhog and Ambemohr, both of which had a common band of 235bp. The second group designated indica 2, with a band of 227bp included Indrayani, Malkolam, Vargalsal and Basmati-370. All the three japonica varieties used in our study viz. Fujisaka, Norin-49 and Taichung-65, had a single band of 231bp. Variations observed in the band size in indica genotypes and occurrence of a single band in japonica lines reflect the greater diversity of indicas as compared to japonicas.

4.3.2 Repeat expansion in wild and cultivated rice genotypes at the RM122 locus

In order to determine the cause of variation in allele sizes between wild and cultivated genotypes, the PCR product was directly sequenced without cloning. Multiple sequence alignment using the CLUSTAL W package revealed that the sequences flanking the $(GA)_n$ repeat region were conserved in all the rice genotypes analyzed. As seen in Fig. 4.1, all the wild rice species viz. *O.longistaminata*, *O.eichengeri*, *O.alta*, *O.australiensis* and *O.brachyantha* had 14 GA dinucleotide repeats, which increased to 16 in *O.officinalis* and 18 in *O.rufipogon*. The cultivated rice varieties, on the other hand, had a higher number of repeats, ranging from 20 to 24, suggesting that

INDICA 1	**GAGAGAGAGAGAGAA--GAGAAGAGAGAGAGAGAGAGAGAGAGAGAGA**
INDICA 2	**GAGAGAGAGAGAGAA--GAGAA--------GAGAGAGAGAGAGAGAGAGA**
JAPONICA	**GAGAGAGAGA----AGAGAGAA--GAGAGAGAGAGAGAGAGAGAGAGA**
O.rufipogon	**GAGAGAGAGAGA--AGAGAGAA----------GAGAGAGAGAGAGAGA**
O.officinalis	**GAGAGA----------AGAGAA--------GAGAGAGAGAGAGAGAGAGA**
O.longistaminata	**GAGAGAGAGA----AGAGAGAA----------------GAGAGAGAGAGA**
O.alta	**GAGAGAGAGA----AGAGAGAA----------------GAGAGAGAGAGA**
O.australiensis	**GAGAGAGAGA----AGAGAGAA----------------GAGAGAGAGAGA**
O.brachyantha	**GAGAGAGAGA----AGAGAGAA----------------GAGAGAGAGAGA**
O.eichengeri	**GAGAGAGAGA-------GAGAA---------------GAGAGAGAGAGAGA**

Figure 4.1: Multiple sequence alignment of $(GA)_n$ repeats at the RM122 locus in rice using CLUSTAL W package. The repeats in cultivated rice (*Oryza sativa*) can be classified into three types, indica 1, which represents Tambdarambhog and Ambemohr, indica 2 which contains Indrayani, Malkolam, Vargalsal and Basmati-370 and japonica which comprises of Fujisaka, Norin-49 and Taichung-65. Gaps relative to the indica sequence are shown by dashes.

microsatellites were probably duplicated in the process of evolution. Thus, the repeat number was found to increase in steps of two dinucleotide motifs. A close observation of the sequence of the five wild rice species having 14 GA repeats showed differences in the arrangement / pattern of the $(GA)_n$ dinucleotide motifs. Four of the wilds viz. *O.longistaminata*, *O.alta*, *O.australiensis* and *O.brachyantha* contained the microsatellite $(GA)_5A(GA)_3A(GA)_6$ while *O.eichengeri* had $(GA)_5A(GA)_2A(GA)_7$. However, despite differences in the arrangement of the repeat, the number of motifs was conserved.

The sequence of cultivated rice revealed the presence of three different groups based on the number and pattern of GA repeats present. The first group comprising the indica rice cultivars, Tambdarambhog and Ambemohr contained 24 GA dinucleotide motifs and was designated indica 1. The second group with 20 repeats contained the indicas, Indrayani, Malkolam, Vargalsal and Basmati-370 and was called indica 2. The third group was made up of three japonica cultivars, Fujisaka, Norin-49 and Taichung-65, all of which had 22 GA motifs. Thus, japonicas had an intermediate number of repeats which was between that of the two indica groups. Even by sequence analysis, all the cultivars formed three groups which were the same as those revealed by band sizes.

4.4 DISCUSSION

The genus *Oryza*, which includes cultivated and wild rice species, probably originated about 130 million years ago (mya) in Gondwanaland and spread as a wild grass (Chang, 1976a; 1976b). The breakup of Gondwanaland has resulted into the distribution of different species into different continents. Wild rice, *O.rufipogon* which, belongs to the AA genome and is the wild progenitor of common cultivated rice, has evolved around 15 mya whereas domestication of wild rices has started about 9000 years ago (Khush, 1997) based on evidence from fossils of *O.sativa*. Wild rice species growing under natural environmental and ecological conditions, have undergone several changes due to intervention by man and adaptation to diverse environments in the process of domestication. Cultivated rice, *O.sativa*, has two main subspecies, indica and japonica, which Yang *et al*

(1994) could differentiate using microsatellite markers harboring dinucleotide repeats such as $(GA)_n$ or $(GT)_n$.

My study of different wild and cultivated genotypes of rice at a $(GA)_n$-containing microsatellite locus, RM122, has revealed that the regions flanking the repeats are highly conserved in all the genotypes while the repeat number is found to increase, thus suggesting that repeat areas have evolved somewhat independently of the evolutionary rate of the adjacent single-copy areas. Further analysis of the repeat region has led to the conclusion that the $(GA)_n$ microsatellite within common cultivated rice, *O.sativa*, represented by the indicas (which has two different repeat types) and the japonicas seems to have arisen from *O.rufipogon* - the wild progenitor of cultivated rice. Most of the wild species analyzed in the present study have fewer number of repeat units which might have undergone expansion by replication slippage (Levinson and Gutman, 1987; Charlesworth *et al*, 1994) during the process of evolution of cultivars. These expansions seem to be occurring according to the stepwise mutation theory (Ohta and Kimura, 1973; Valdes *et al*, 1993; Charlesworth *et al*, 1994) proposed for microsatellite evolution, where short arrays increase or decrease in repeat number by one or two repeat units, probably by replication slippage. In the wilds, $(GA)_n$ dinucleotide expands very slowly on an evolutionary time scale but once the repeat unit reaches a minimum threshold number it seems to be expanding faster in cultivars.

Hancock (1996) has shown that accumulation of simple sequences in the genome is related to increase in genomic size from eubacteria to humans. Such a phenomenon has also been proposed by Messier *et al* (1996), based on their analysis of two microsatellites in a short arm of the η–globin pseudogene of primates. They have suggested that a minimum number of repeat units might be necessary before initial expansion occurs. Once a critical number of repeat units (Valdes *et al*, 1993; Armour *et al*, 1994) has arisen in a given species, that locus can become hypervariable, with mutations occurring even on a generational time scale (Campuzano *et al*, 1996). Population genetic and emperical data have also demonstrated that microsatellites are more likely to expand in length when they mutate and the rate of expansion of such sequences differs in related species (Rubinsztein *et*

al, 1999). However, studies by Xu *et al* (2000) have revealed that in animal systems, microsatellite loci are equally likely to gain or lose repeats when they mutate and the overall rate of expansion mutations does not differ from that of contractions, thus maintaining a stable allele distribution of microsatellites. In pine, duplication and / or expansion of dinucleotide repeats viz. AC or GT, has occurred only after the divergence of *P.strobus* and *P.lambertina* (10-25 mya) or *P.resinosa* and *P.sylvestris* (10-25 mya), respectively, before which there was only one or at the most two repeat units in all the species (Karhu *et al*, 2000). They have postulated that base substitutions could be responsible for providing material for replication slippage or some other inserting mechanism and, thus, could enable further expansion. It has been suggested that during plant evolution genome size increases (Bennetzen and Kellogg, 1997), which is well illustrated in the history of the grass family where genome sizes have increased upto 28 times. In case of rice, where such a dramatic increase in genomic size is not observed, an increase in one class of repeats is accompanied by a decrease in another class of sequences. Whether there is a correlation between the higher DNA content of cultivated rice versus wild rice (Arumunagathan and Earle, 1991) and the increase in repeat number as shown in our study is still an open question.

Th $(GA)_n$-harboring microsatellite locus, RM122 (Wu and Tanksley, 1993) has been shown to be tightly linked to the bacterial blight resistance gene, *xa*-5, in rice (Blair and McCouch, 1997). My observation regarding the association between $(GA)_n$ dinucleotide expansion and domestication of rice is the first of its kind in plants in contrast to the repeat expansion associated with genetic disorders in humans. A similar observation has been later reported in pine, involving expansion of AC or GT dinucleotide repeats (Karhu *et al*, 2000). This is the first step towards understanding a biologically significant role of microsatellites in plants. The study of association of microsatellite repeat number expansion with domestication in other cereals may provide valuable information about the origin and evolution of microsatellites.

CHAPTER 5

Sequence Variations at a Complex Microsatellite Locus in Rice and Its Conservation in Cereals

The contents of this chapter have been published as a full-length paper in Theoretical and Applied Genetics (2000) 101: 1291-1298

ABSTRACT

In an attempt to study changes associated with microsatellites in rice, DNAs of cultivated rice including indica and japonica varieties and wild rice genotypes were amplified by polymerase chain reaction with primers flanking the $(GATA)_n$ and $(AC)_n$ repeats at a microsatellite containing locus OS1E6 (GenBank accession number AFO16647) previously reported from *Pst*I rice (var. Malkolam) genomic library in pUC18. Eight alleles of varying sizes were obtained which were cloned and sequenced. Sequencing data indicated that the size variations of different alleles were due to differences in the repeat number as well as sequence variations in the region flanking the microsatellite motifs.

In order to study the presence of the above complex microsatellite-containing locus of rice in different cereals, cereal DNAs were amplified using primers flanking the OS1E6 locus. It was found that this locus was present in various cereal genotypes analyzed indicating its conservation across different members of cereals.

5.1 INTRODUCTION

PCR-based microsatellites are of great use in genomic, systematic, evolutionary and ecological studies and such analyses can be even extended across different laboratories. To illustrate, SSRs have been successfully amplified in different plant species such as *A.thaliana* (van Treuren *et al*, 1997), citrus (Kijas *et al*, 1995), soybean (Maughan *et al*, 1995; Powell *et al*, 1995; Peakall *et al*, 1998) and chickpea (Udupa *et al*, 1999). As seen in chapter 3 (section 3.3.4), cross-species amplification has been obtained in different species of the genus *Oryza* using primers flanking $(GATA)_n$ microsatellites.

Although different cereals have evolved from a common ancestor, there are vast differences in their genome sizes (Bennett and Smith, 1991). The massive expansion in the size of some of these cereal genomes is due to a large amplification of repetitive DNA, which evolves rapidly and hence diverges substantially with speciation (Moore *et al*, 1993). Among cereals, rice has the smallest genome size and is considered to be an ideal crop to study cereal genome evolution and trans-species crop improvement. I was, therefore, interested to study if non-coding microsatellite containing loci are also conserved / variable in different cereals. Since I have earlier observed that $(GATA)_n$-containing microsatellite loci could be amplified in four different cereal genotypes using flanking primers from rice (section 3.3.4 of chapter 3), I wanted to extend this work using more cereal genera and also a greater number of species from each genus. In the present investigation, I have analyzed inter- and intra-specific variations at a complex locus containing tetra- and di-mer microsatellites using various cultivars and a landrace of rice, a number of cereals and related wild germplasm. Homologous loci from different species may or may not contain same copy number of repeat motifs (section 4.3.2 of chapter 4) and that too at the same position. I have, therefore, further sequenced these loci from representative *Oryza* species in order to study sequence divergence in these genotypes.

5.2 MATERIALS AND METHODS

5.2.1 Plant material and DNA extraction

Seeds of different cultivated and wild genotypes of rice were obtained from various sources as mentioned in chapter 4 (section 4.2.1). Different cereal genotypes used for the analysis have been listed in Table 5.1. Seeds of pearl millet and maize were procured from Indian Agricultural Research Institute, New Delhi, while seeds of wheat were obtained from Agharkar Research Institute, Pune, India. Seeds of barley, oat, rye and *Aegilops* were made available by USDA ARS, National Small Grains Collection, Aberdeen, U.S.A.

Total genomic DNA was extracted from leaves of rice, barley, maize, oat, rye and wheat by CTAB method as outlined in chapter 2 (section 2.2.1), while pearl millet DNA extraction was according to Sharp *et al* (1988). The pearl millet leaf tissue was harvested, stored and crushed as described for CTAB method in chapter 2 (section 2.2.1). To 5g of crushed tissue, 15ml of extraction buffer [100mM Tris-HCl, 100mM NaCl, 50mM EDTA and 2% SDS] was added and mixed. Then proteinase K was added such that its final concentration was 50µg/ml and incubated at 37^{o}C for two hours. Equal volume of phenol : chloroform : isoamyl alcohol (25:24:1) mixture was added and mixed thoroughly to form an emulsion, which was centrifuged at 10,000 rpm for 15min in a SS34 rotor. The upper aqueous layer was treated with equal volume of chloroform : isoamyl alcohol (24:1) mixture and centrifuged as above. To the supernatant, one-tenth volume of 3M sodium acetate (pH 5.2) and two and a half volumes of absolute ethanol were added, mixed well and left at –20^{o}C to precipitate the DNA which was then centrifuged as above. The DNA pellet was washed with 70% ethanol, dissolved in TE buffer, treated with RNase A and concentration was estimated as described for CTAB DNA isolation in chapter 2 (section 2.2.1).

5.2.2 Locus specific PCR amplification of different rice genotypes

A plasmid clone, OS1E6 (GenBank accession number AFO16647) identified by screening partial *Pst*I genomic library of rice (var. Malkolam) in pUC18 in previous studies (Chowdari *et al*, 1998b) at our laboratory contained three

TABLE 5.1: List of different cereals used in the present study along with the number of alleles amplified from each at the OS1E6 and RM122 loci.

Sr. No.	Cereal	Genotype	Accession Number	Number of alleles at OS 1E6 locus	Number of alleles at RM122 locus
1	Wheat	*Triticum dicoccoides*	3515	1	4
2	Wheat	*Triticum dicoccum* var. Ex-33	33b	2	-
3	Wheat	*Triticum durum* landrace Narsimgarh 111	1633	1	1
4	Wheat	*Triticum aestivum* cv. Chinese spring	226	1	1
5	Wheat	*Aegilops tauschii*	PI 21098	2	-
6	-	*Chionachne koenigir* var. Chio3	DK93	3	3
7	Maize	*Zea mays* cv. Sikkim Primitive	DK94/6714	2	2
8	Barley	*Hordeum marinum*	PI 200341	2	2
9	Barley	*Hordeum bogdanii*	PI 269406	-	2
10	Barley	*Hordeum vulgare* ssp. *vulgare*	CLHO 7382	2	1
11	Oat	*Avena sativa*	CLAV 1110	2	1
12	Oat	*Avena fatua*	CLAV 2526	2	1
13	Oat	*Avena vavil oviana*	PI 412733	2	1
14	Pearl millet	*Pennisetum glaucum* cv. 841B	-	1	2
15	Rye	*Secale cereale* ssp. *ancestrale*	CLSE 107	2	1

microsatellite repeats namely $(GATA)_n$, $(AC)_n$ and $(AG)_n$. Primers flanking the $(GATA)_n$ and $(AC)_n$ repeats were synthesized and used for locus specific amplification by PCR in various rice genotypes. The sequence of the primers used in PCR was as follows:

Primer 1 - 5' AGCAGTAGAGGGAGATGA 3'

Primer 2 - 5' TCTGTCATGCAGCCATGC 3'

Each reaction was performed in 10µl volume as described in chapter 2 (section 2.2.3) using 1µCi each of α^{32}P dATP and α^{32} PdCTP and 50ng of template DNA per reaction alongwith Taq DNA polymerase (Boehringer Mannheim, Germany). Conditions for DNA amplification were the same as in chapter 2 (section 2.2.3), however, 40 thermal cycles were performed in a Perkin Elmer Cetus thermal cycler and the amplification products were electrophoresed as described in chapter 2 (section 2.2.3).

5.2.3 Cloning and sequencing of PCR products

Alleles of different sizes obtained on amplification of rice wilds and cultivars were cloned and sequenced after performing amplification reactions in a larger volume of 100µl as described earlier in chapter 4 (section 4.2.2). The PCR products were loaded on 1.4% agarose gels and the gel piece containing the amplified fragment of interest was cut. The DNA was eluted from the gel piece using Qia Quick DNA elution kit (QIAGEN, U.S.A.) and ligated into PCR product cloning vector. The ligation reaction was carried out at 16^{o}C for 16 hours in 20µl with 2U of T4 DNA ligase (U.S. Biochemical Corp., U.K.), 50ng pMOS blue T-vector (U.S. Biochemical Corp., U.K.) in 660mM Tris HCl (pH 7.6), 66mM $MgCl_2$, 100mM DTT and 660mM ATP. The ligated mixture was transformed into XL1-Blue competent cells as described in chapter 3 (section 3.2.2) and the resulting colonies were screened for recombinants by blue / white colony selection as described by Sambrook *et al* (1989). The white colonies contained positive clones with insert and were selected for further processing. Plasmid DNA was isolated from these positive clones by alkaline lysis method, purified by polyethylene glycol method and sequenced by Sanger's dideoxy chain termination method using Sequenase version 2.0 DNA sequencing kit (U.S. Biochemical Corp., U.K.) as

described in chapter 3 (section 3.2.3). The sequences of the amplified and cloned products were aligned using the CLUSTAL W (1.60) multiple alignment package (Thompson *et al*, 1994).

5.2.4 Locus specific amplification in cereals

The primers used for amplification of different rice genotypes were used to amplify DNA of barley, maize, oat, pearl millet, rye, wheat and their wild relatives. PCR was performed in 10μl volume as described in chapter 2 (section 2.2.3).

5.3 RESULTS

5.3.1 Sequence variations at a complex microsatellite containing OS1E6 locus in rice

Genetic diversity in 11 different wild and cultivated genotypes of rice, when assessed at the complex microsatellite-containing locus, OS1E6, using flanking primers, resulted in one common allele of size 220bp in all the rice genotypes studied, which was the expected size of the allele in the original OS1E6 clone obtained from the Malkolam genomic library. In addition to this one common allele, several other alleles were also obtained in different rice genotypes studied. Among the *O. sativa* varieties studied, in addition to the allele of 220bp, the rice cultivars, Basmati-370 and Taichung-65 and the landrace, Tambdarambhog resulted in alleles of sizes 226bp, 268bp and 192bp, respectively. In the wild rice species studied, *O.rufipogon* (AA genome) gave a single band of 220bp while *O.officinalis* (CC), *O.granulata* (GG), *O.minuta* (BBCC) and *O.latifolia* (CCDD) showed alleles of sizes 224bp, 227bp, 228bp and 234bp, respectively in addition to the common band at 220bp.

Since the OS1E6 locus was a complex locus containing two different microsatellite repeats, $(GATA)_n$ and $(AC)_n$, the microsatellite sequence responsible for size variations among the different alleles was determined to study if alleles of same size had the same sequence. The alleles of varying sizes from different wild and cultivated rices were eluted, cloned and sequenced and multiple sequence alignment was performed using CLUSTAL

W package (Fig. 5.1). As seen in Fig. 5.1, alleles of different sizes showed variations in both the repeat regions as well as the flanking sequences. The 220bp allele of Indrayani, Norin-49 and *O.rufipogon* showed a similar sequence composition. With the exception of one additional nucleotide (viz. T) in *O.minuta,* the sequences of *O.minuta* and *O.granulata* were also similar.

All the sequences shown in Fig. 5.1 basically contained two stretches of repeats, $(GATA)_n$, and $(AC)_n$. Analysis of the GATA repeat region revealed three types of repeats in different rice allele sequences studied. As given in Table 5.2, Indrayani, Tambdarambhog, Norin-49, Taichung-65, *O.rufipogon*, *O.officinalis*, *O.minuta* and *O.granulata* formed the first group, containing 9 GATA like repeats in a continuous stretch, while Malkolam and Basmati-370 comprised the second group containing 12 GATA like repeats in a discontinuous stretch (interrupted by CAGG motif) and *O.latifolia* with 14 GATA repeats made up the third group. In all these rice genotypes, the GATA repeat was an imperfect repeat containing motifs such as GATT, GATC and AATA. In addition, in case of Malkolam, Basmati-370 and *O.latifolia*, the GATA repeat region was made discontinuous due to a CAGG motif within the GATA repeat.

A close observation of the AC repeat region revealed the presence of three types of motifs, as seen in Table 5.2. Indrayani, Norin-49, Taichung-65, *O.rufipogon* and *O.officinalis* containing four perfect AC repeats constituted the first group of AC repeats. In addition, Tambdarambhog, containing 3 AC and 1 CA motif was also included in the first group. All the rice wilds and cultivars included in group 1 of GATA repeats, with the exception of *O.minuta* and *O.granulata*, were also present in the first group of AC repeats. The second group of AC repeats comprised Malkolam, Basmati-370, *O.minuta* and *O.granulata*, all of which contained 9 imperfect AC repeats, while *O.latifolia* containing 7 perfect AC repeats formed the third group.

In addition to two main regions viz. GATA, and AC, a third site comprising different microsatellite sequences was also observed. As seen in Table 5.2, Indrayani, Norin-49, Taichung-65, *O.rufipogon*, *O.officinalis*, *O.granulata* and *O.minuta* contained the motif GGATGA. In Taichung-65, the motif changed to AGATGA due to a point mutation while in *O.minuta*, due to insertion of T, the motif obtained was GGTATGA. All the rice genotypes

FIGURE 5.1: Multiple sequence alignment using CLUSTAL W package at the OS1E6 locus in genotypes of cultivated and wild rice. Gaps during alignment are indicated by dashes. Stars denote the nucleotides that are identical. $(GATA)_n$ repeats are highlighted in red while $(AC)_n$ repeats are in blue and the remaining repeats are in green or pink.

TABLE 5.2: Sequence variation at the $(GATA)_n$ and $(AC)_n$ repeat regions and a third site at the complex microsatellites containing OS1E6 locus.

Group	No. of repeats	Sequence at repeat region
[A] At $(GATA)_n$ repeat		
Group 1 (Indrayani, Tambdarambhog, Norin-49, Taichung-65, *O.rufipogon*, *O.officinalis*, *O.minuta*, *O.granulata*)	9	GATA GATC $(GATA)_2$ AATA $(GATA)_2$ $(GATT)_2$
Group 2 (Malkolam, Basmati-370)	12	$(GATA)_2$ GATT CAGG $(GATA)_2$ AATA $(GATA)_3$ GATTT $(GATT)_2$
Group 3 (*O.latifolia*)	14	$(GATA)_2$ $(GATA)_2$ GATT CAGG $(GATA)_2$ AATA $(GATA)_3$ $(GATA)_3$
[B] At $(AC)_n$ repeat		
Group 1 (Indrayani, Tambdarambhog, Norin-49, Taichung-65, *O.rufipogon*, *O.officinalis*)	4	$(AC)_4$ [Tambdarambhog CA $(AC)_3$]
Group 2 (Malkolam, Basmati-370, *O.minuta*, *O.granulata*)	9	$(AC)_6$ TC $(AC)_2$
Group 3 (*O.latifolia*)	7	$(AC)_7$
[C] At third motif		
Group 1 (Indrayani, Norin-49, Taichung-65, *O.rufipogon*, *O.officinalis*, *O.minuta*, *O.granulata*)	1	GGA TGA [Taichung-65 AGATGA] [O.minuta GGTATGA]
Group 2 (Malkolam, Basmati-370, *O.latifolia*	1	AGGTGGTGG

which fell into the first group of GATA repeats also contained the GGATGA motif. Malkolam, Basmati-370 and *O.latifolia*, which formed the second and third groups of GATA repeats, contained the motif AGGTGGTGG at this site.

Taichung-65 showed the presence of an additional $(GATA)_n$ motif, 37 nucleotides after the first stretch of $(GATA)_n$. This region contained 6 GATA motifs preceded by a $(TGTAG)_2$ motif (Fig. 5.1).

5.3.2 Conservation of the OS1E6 locus in cereals

Since the OS1E6 locus was found to be present in all the wild and cultivated rice species studied, it would be interesting to check if this locus was conserved across different cereals also. Hence DNA from 15 diverse genotypes of wheat, maize, barley, oat, pearl millet, rye and their wild relatives were amplified using OS1E6 flanking primers (Fig. 5.2). As seen in Fig. 5.2, a common band of 220bp was present in all the cereals used in the study except *H.bogdanii* (lane 9). In all the rice wilds and cultivars analyzed, this common band of 220bp was also found to occur indicating presence of this allele in all the cereals examined. In addition to this common allele, several other alleles were also detected. A cultivated, tetraploid wheat, *T.dicoccum*, showed the allele of 226bp (lane 2), while a wild, diploid wheat, *Aegilops tauschii*, had a band of 236bp (lane 5). Maize cultivar Sikkim Primitive and its distant wild relative, *Chionachne koenigir* var. Chio-3 showed the presence of an allele at 226bp (lanes 7 and 6, respectively) with *Chionachne* containing a third band of a very high molecular weight (lane 6). Barley genotypes *H.marinum* and *H.vulgare*, three oat genotypes *A.sativa*, *A.fatua* and *A.vaviloviana* and rye had a second allele of 236bp (lanes 8 10, 11, 12, 13 and 15, respectively). *H.bogdanii* showed no amplification, indicating presence of a null allele at this locus.

Thus, totally 4 different alleles were observed when the OS1E6 locus was amplified from 15 different cereals indicating its conservation during evolution (Table 5.1).

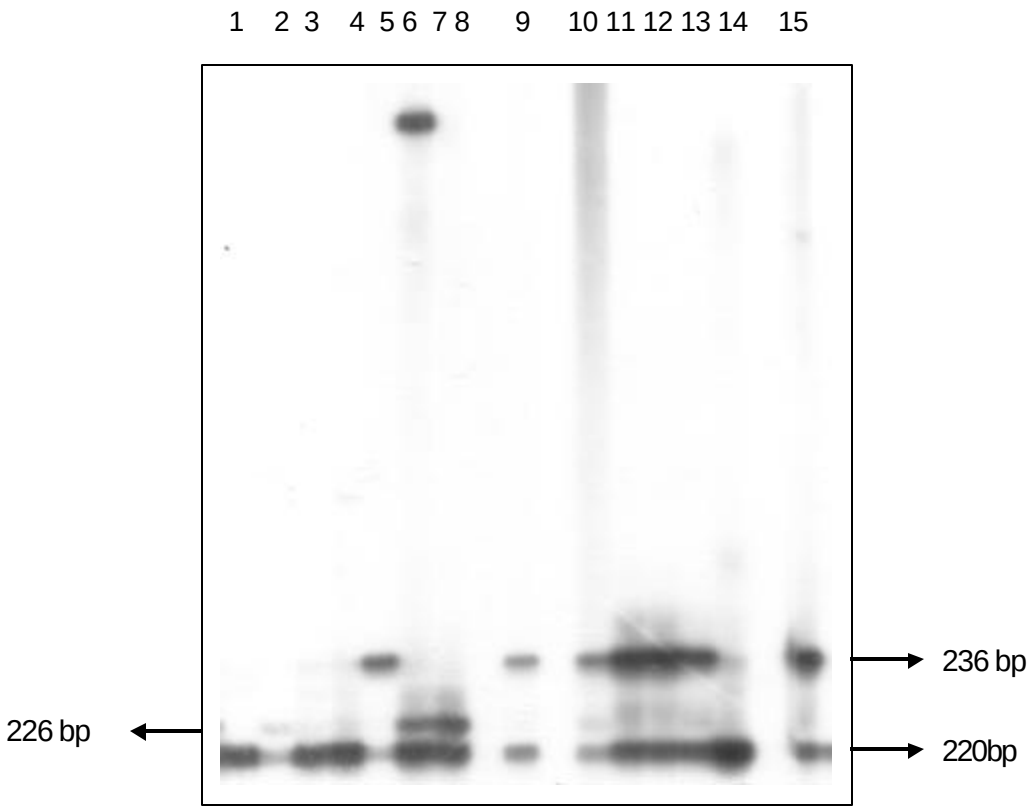


Figure 5.2: Radioactive PCR amplification of the OS1E6 locus from different cereals. Genomic DNA of several cereals was amplified and loaded on 6% denaturing polyacrylamide gel. Lanes 1 to 15 are as shown in Table 5.1.

5.3.3 Genetic diversity in cereals using primers flanking a mapped microsatellite locus in rice

Since the rice OS1E6 locus was conserved in different cereals, this analysis was further extended to check if other mapped loci of rice were also conserved in cereals. PCR amplifications were performed in diverse cereals at a $(GA)_n$ harboring microsatellite locus, RM122 (Wu and Tanksley, 1993), which I had earlier amplified and sequenced from different wild and cultivated rice genotypes (section 4.3.2 of chapter 4). The amplification of different cereal genotypes with RM122 flanking primers has been depicted in Fig. 5.3, wherein a common band of 227bp was observed in all the cereals viz. maize, barley, oat, pearl millet and rye, except wheat. In addition, maize, barley and pearl millet showed bands of higher molecular weight. In wheat, *T.dicoccoides* (lane 1) revealed presence of 4 different bands while *T.durum* landrace Narsimgarh 111 (lane 3) and *T.aestivum* cv. Chinese spring (lane 4) had only one common band of 243bp, which was also present in *T.dicoccoides* (lane 1). Interestingly, this band of 243bp was not observed in any of the other cereals studied. *T.dicoccum* (lane 2) and *Aegilops tauschii* (lane 5) did not show any amplification at this locus, possibly indicating the presence of a null allele in them. On the other hand, *H.bogdanii*, which showed no amplification at the OS1E6 locus, resulted in two bands when amplified with primers flanking the RM122 locus (lane 9). Amplification of the RM122 locus from 15 different cereals thus resulted in identification of 7 alleles of varying sizes (Table 5.1).

5.4 DISCUSSION

5.4.1 Sequence variations in repetitive and non-repetitive region: a cause for allelic polymorphism at microsatellite loci

It is now well established that microsatellite-containing sequences i.e. non-coding sequences evolve at a faster rate as compared to coding sequences. Consequently, a large number of informative nucleotide variations can be detected on studying a shorter stretch of DNA while comparing closely related species. $(GATA)_n$ is the most frequent tetranucleotide reported in rice with about 270 $(GATA)_n$ motifs in the entire genome, while the microsatellite $(AC)_n$ is the second highest in frequency in

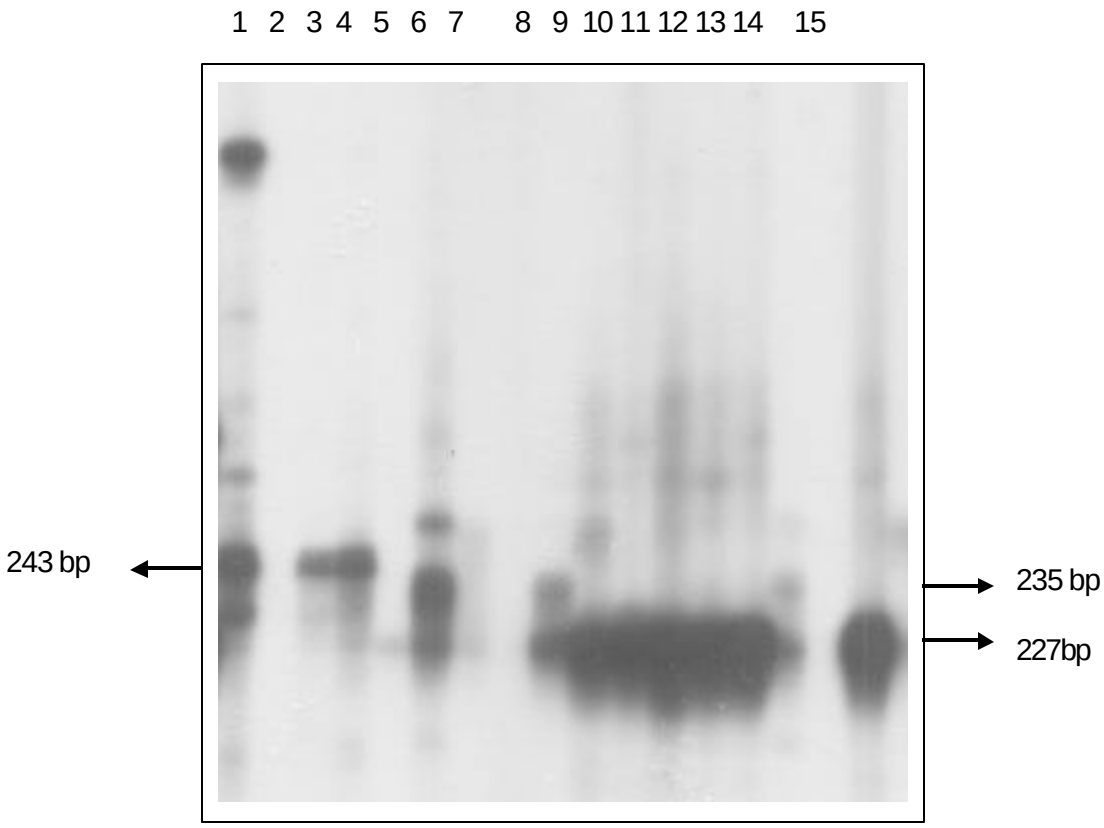


Figure 5.3: Radioactive PCR amplification of the RM122 locus from different cereals. Genomic DNA of several cereals was amplified and loaded on 6% denaturing polyacrylamide gel. Lanes 1 to 15 are as shown in Table 5.1.

the rice genome (Panaud *et al*, 1995; McCouch *et al*, 1997). Since the OS1E6 locus contains both $(GATA)_n$ and $(AC)_n$ repeats, it was thought that it would be interesting to exploit this locus to analyze fine variations in the rice genome. In this context, I have shown that PCR amplification of the OS1E6 locus from different rice wilds and cultivars resulted in 8 alleles varying from 192bp to 268bp in size. Moreover, the *O.sativa* primers show amplification in other *Oryza* species at the same stringency level in PCR indicating that these sequences might be located in the conserved region of the genome. However, the alleles obtained in different *Oryza* species are of different sizes and no two species have minimum two alleles of the same size.

There is very little information available in plants regarding sequences of different alleles of a microsatellite locus within the same species and between different species. On this background, the sequence variation of different alleles at the OS1E6 locus observed in my analysis has revealed that increase in the number of AC repeats is responsible for the increase in length of the alleles of *O.minuta* and *O.granulata*. The increased size of Taichung-65 allele, on the other hand, is due to an insertion of a $(TGTAG)_2$ motif followed by a stretch of GATA repeat containing 6 imperfect motifs. However, for the other rice wilds and cultivars studied, mere variations in repeat region could not account for the varying sizes of the alleles. In these rice genotypes, differences in allele sizes are due to variations in the adjoining non-repetitive region as well. In case of Malkolam, Basmati-370 and *O.latifolia*, there is an increase in repeat number with a subsequent deletion in the sequence immediately following the $(GATA)_n$ repeat. In addition, in *O.latifolia* there is also an expansion of the AC repeat, followed by deletion of sequences just preceding this repeat, thus resulting in an allele of the observed size of 234bp. In case of Tambdarambhog, the decreased size of the allele is due to a deletion in the region between the GATA and AC repeats. In addition, in Malkolam, Basmati-370 and *O.latifolia* the GATA repeat has been interrupted by the insertion of a CAGG motif.

Several examples are available in different systems regarding variations at microsatellite loci. Inter-specific variation at a microsatellite-containing intergenic region in the mitochondrial genome of genus *Pinus* was due to expansion and contraction of the repeat motif (Soranzo *et al*, 1999).

Phelps *et al* (1996) have demonstrated microsatellite repeat variation within the *y1* gene of maize and teosinte. Studies of microsatellites in the TNF (tumor necrosis factor) region (in major histocompatibility complex of humans and primates) and in chromosome 4 indicate that the inter-specific differences in allele length are not always due to changes in repeat number but also due to insertions and deletions in the sequences flanking the repeats and interruption of perfect repeats which are correlated to microsatellite stability (Blanquer-Maumont and Crouau-Roy, 1995; Garza *et al*, 1995; Crouau-Roy *et al*, 1996). Deka et al (1995) have also reported that in humans increase or decrease in allele size is not in increments of 2bp but could involve insertions / deletions of single nucleotides or might be due to involvement of some other more complex phenomena. Another school of thought has suggested that the most likely mechanism for evolution of simple sequence repeats is slipped strand mispairing (Levinson and Gutman, 1987, Charlesworth *et al*, 1994). The significant correlation observed between the average number of repeats and the amount of variation obtained in our analysis indicates that replication slippage may be the molecular mechanism responsible for generating variability at the OS1E6 locus in rice.

5.4.2 Microsatellite loci: a tool for comparative genome mapping in cereals

Comparative genetic studies using rice, wheat, maize, oat, sorghum, foxtail millet and sugarcane have demonstrated that gene content and gene order are highly conserved between species within the grass family, both at the map and megabase level (Devos and Gale, 1997). Since microsatellites are non-coding sequences, their conservation between species is less marked. However, since all the plant genomes contain repetitive regions and amplification has been obtained earlier using rice microsatellite flanking primers on four different members of the family Gramineae (section 3.3.4 of chapter 3), I wanted to see if the OS1E6 locus of rice was conserved in different cereals. Interestingly, my studies have indicated that the OS1E6 locus is highly conserved in all the cereals studied except *H.bogdanii*, where no band has been obtained on amplification with primers flanking the OS1E6 locus. The null allele in *H.bogdanii* could be either due to absence of the

OS1E6 locus (Thomas *et al*, 1994) or as a result of mutations (point mutations, insertions or deletions) at the primer-binding site (Callen *et al*, 1993). It has been further shown that the conservation of the microsatellite-containing OS1E6 locus is not an isolated phenomenon, but the $(GA)_n$-containing locus has also been conserved in almost all the cereals studied with the exception of *T.dicoccum* and *A.tauchii*. The reasons for obtaining a null allele in these two genotypes could be as stated above for *H.bogdanii*. The other wheat species viz. *T.dicoccoides*, *T.durum* landrace Narsimgarh 111 and *T.aestivum* cv. Chinese spring have shown amplification of this locus but the size of alleles obtained is higher than expected suggesting that the RM122 locus has been selectively conserved in wheat. Zhao and Kochert (1993b) have also shown that the $(GGC)_n$ containing microsatellite locus, RTL011, from rice could be amplified from maize and bamboo indicating that some microsatellite loci are present in conserved regions of the genome. Further examples of cross-species amplification and conservation of microsatellite loci across plant and animal systems have been sited in the discussion in chapter 3 (section 3.4.3).

Microsatellites, being non-coding sequences, additions and deletions occur on a rapid time scale and hence, their conservation across different genera is thought to be low. However, the data in this chapter reveal that these microsatellite-containing loci are conserved in different cereals and, hence, they could be used as tools for comparative mapping with a common set of primers. Since microsatellite loci have been shown to be conserved across different species and genera, a large proportion of polymorphic microsatellites could be cross amplified and further used to search for economically useful loci.

CHAPTER 6

General Discussion

Microsatellites: Universal Genetic Mapping Reagents

6.1 Introduction

Microsatellites or simple sequence repeats (SSRs) are a group of repetitive DNA sequences that represent a significant portion of eukaryotic genomes and are responsible for the large amount of genetic diversity present in different genomes. The advent of PCR-based markers has made it possible to quantify the diversity and to study the mechanisms responsible for generation of variation at these microsatellite loci. One more feature of microsatellites is that they provide a common language for collaborative research and act as universal genetic mapping reagents (Silver, 1992) since they can be distributed all over the world. Besides, they are amenable to automation (Thomas and Scott, 1993) which is not possible for RFLP markers and is rather difficult for RAPD markers since they are not easily scorable. Thus, STMS markers are the preferred markers for wide and varied applications like genetic diversity analysis (Yang *et al*, 1994), cultivar identification (Thomas and Scott, 1993), determination of hybridity (Provan *et al*, 1996) and diagnosis of important traits in plant breeding programs (Yu *et al*, 1994).

In rice genome, although extensive molecular data is available including its DNA sequence, comparatively less work has been carried out involving microsatellites. Since microsatellites have a great promise in genome research, I was interested in the following specific aspects in rice:

(1) To exploit the advantages of microsatellites for inter- and intra-specific studies

(2) To analyze such regions at the sequence level in order to determine the probable causes / mechanisms responsible for generating variability at these loci.

(3) To examine if these loci are conserved across different cereal genera.

6.2 High discriminatory potential of STMS markers

Since microsatellites exhibit a high degree of polymorphism due to variations in the repeat number at a particular locus (Schlotterer *et al*, 1997), I analyzed 42 closely related elite cultivars with primers flanking the microsatellite motifs. Using STMS markers it has been observed that the values of genetic similarity show a sharp decline (0.37), while the values of

average heterozygosity (0.88) and marker index (5.49) increase. Using nine STMS markers, the probability of obtaining an identical match by chance is $3.46X10^{-5}$, indicating that 3,46,000 elite cultivars could be differentiated with these markers. This reflects the high discriminatory potential of STMS markers and ratifies their use for analysis of genotypes with a narrow genetic base. Thus, out of the three marker systems viz. RAPD, ISSR and STMS used for analysis, STMS markers have been found to be the most informative.

6.3 Diagnostic potential of $(GATA)_n$-containing microsatellite loci

When three $(GATA)_n$-containing microsatellite loci were analyzed for their ability to determine variability between cultivated and wild genotypes of rice, it has been observed that the bacterial blight resistant genotypes cluster separately from the other rice genotypes. Also, the bacterial blight resistant genotypes form different groups depending on the resistance genes present in them. Thus, even the small amount of data collected at three polymorphic SSR loci exhibiting length polymorphism could reveal the diagnostic potential of these microsatellite markers.

6.4 High mutation rate at microsatellite loci

Henderson and Petes (1992) and Weber and Wong (1993) have concluded that microsatellite loci are extremely unstable and exhibit an exceptionally high rate of mutations that add or subtract a small number of perfect repeats. Recently, Udupa *et al* (1999) have proposed that the main reason for the high level of polymorphism at the intra-specific level is the high mutation rate at microsatellite loci. However, mutations in both the SSR region as well as the flanking region can contribute to the allelic variation among species making it impossible to infer SSR number from the allele size leading to serious underestimation of the genetic relationship between taxa (Garza *et al*, 1995; Garza and Freimer, 1996; Nauta and Weising, 1996). In order to minimize the risk of misinterpretation and maximize the genetic information that can be obtained, DNA sequencing of SSR alleles is essential (Goldstein and Pollock, 1997; Orti *et al*, 1997) as this information can then be used to meaningfully address applied and evolutionary questions.

In my work, I have sequenced the alleles obtained at two microsatellite loci, one containing simple repeats and the other containing compound repeats. Sequencing the alleles at the microsatellite locus, RM122, containing perfect $(GA)_n$ repeats, has revealed that the sequences flanking the SSR region are conserved in different cultivated and wild genotypes of rice and variation in repeat number is the only cause of polymorphism at this locus. An interesting phenomenon observed at this locus, is the expansion of the repeat in cultivated rice, possibly indicating that SSRs are created during evolution. At the second locus, OS1E6, containing compound repeats, difference in allele sizes is due to the repeat region as well as the flanking repeat motifs. Several workers have suggested that pure repeats are unstable with stepwise mutations and deletions increasing with the size of the pure repeat (Amos *et al*, 1996; Primmer *et al*, 1996b; Wierdl *et al*, 1997). However, interruptions within pure repeats appear to stabilize them. Petes *et al* (1997) have observed that interruptions in GT repeats have decreased the mutation rates fivefold. Pure SSRs may, therefore, be in minority and may actually represent a transient evolutionary state.

6.5 Mutational models for microsatellite evolution

Two mutational mechanisms, which have been proposed to explain the high degree of variation at microsatellite loci are replication slippage and unequal recombination. Tautz and Schlotterer (1994) have suggested that replication slippage is the major mechanism and is more efficient with shorter repeat motifs such as mono- or di-nucleotides (Schlotterer and Tautz, 1992). However, the DNA mismatch repair system responsible for repairing most of the slippage events occurring at microsatellite loci (Strand *et al*, 1993) is more efficient on shorter repeats rather than longer ones and thus counterbalances the effects of primary slippage. As a result, loci containing tri- or tetra-nucleotide repeats may be as polymorphic if not more than those containing dinucleotides. A survey of microsatellites in *Drosophila melanogaster* however, has revealed a lack of correlation between the degree of polymorphism and the average length of repeats (Goldstein and Clark, 1995; Schlotterer *et al*, 1997). Analysis of the sequencing data obtained by me at

the OS1E6 locus suggests that replication slippage is the major mechanism for generating variability in rice.

In population genetics, two main mutational models are used viz. infinite allele model (Kimura and Crow; 1964) and the step-wise mutation model (Ohta and Kimura, 1973; Kimura and Ohta, 1978; Valdes *et al*, 1993; Charlesworth *et al*, 1994). The infinite mutation model assumes that all new mutations generate alleles not previously represented in the population while the stepwise mutation model assumes that mutation changes allelic states back and forth and does not necessarily create a new allele. However, studies on yeast artificial constructs (Henderson and Petes, 1992) and human pedigrees (Weber and Wong, 1993) have demonstrated that most mutations involve the addition or subtraction of a small number of repeat motifs, which contradicts the basic assumption of the infinite allele model. My study of the RM122 locus has shown that the repeat expands in steps of two motifs, indicating that stepwise mutation model may be responsible for the observed expansion in the repeat region.

In general, caution needs to be exercised when using microsatellite sequences to estimate the divergence time between different species in evolution. It may not be possible to infer whether the increase in repeat number has occurred one step at a time or through larger additions once a certain threshold number of repeats is reached (Valdes *et al*, 1993; Armour *et al*, 1994; Campuzano *et al*, 1996; Meisser *et al*, 1996). If the repeat expansions have occurred due to larger additions, analysis of these loci would cause excessive overestimation of the genetic distances based on stepwise mutation model (Slatkin, 1995; Goldstein *et al*, 1995). Further, the evolutionary rates at different microsatellites, or even in different parts of the same complex microsatellite locus are highly variable. The molecular mechanisms for these different rates are as yet not understood.

Sequencing several different types of microsatellite loci would help in understanding the mechanisms responsible for the birth and early evolution of microsatellites and would help to elucidate the evolutionary pathway and estimate the approximate divergence time of different species and subsequently different genera.

6.6 Demonstration of usefulness of microsatellites at inter-specific and inter-genus level

In my research on microsatellites in rice, I have observed that STMS markers are not only useful for determining intra-specific variability but are also useful at the inter-specific level. For example, STMS markers developed in *O.sativa* could amplify DNA from different species belonging to the genus *Oryza*. In a similar study, Udupa *et al* (1999) have been able to amplify DNA of *C.reticulatum* and *C.echinospermum* using primers designed for *C.arietinum*. I have also shown successful amplification using primers from *O.sativa* in different genera of the family Gramineae. This is particularly exciting because SSRs tend to be less frequent in plants (Langercrantz *et al*, 1993) and earlier studies have revealed that microsatellite loci are not conserved across Asteraceae (Whitton *et al*, 1997). The process leading to the conservation of microsatellites thus seems to be unrelated to those responsible for any existing polymorphism. However, for studies requiring more extensive cross-species amplification in plants, it may be necessary to target SSRs anchored in genes since the degree of conservation is higher. Although I have demonstrated successful cross-species amplification of several rice microsatellite loci in cereal genera, sequencing results have shown that the maintenance of variation at SSR loci is a complex phenomenon involving mutational processes at the repeat as well as flanking regions.

Thus, in conclusion, my work demonstrates the usefulness of microsatellites not only in rice genomics for the analysis of inter- and intra-specific variability and for studying a complex phenomenon like domestication, but also as a tool to study other cereal genera. In addition to the applied aspects mentioned above, I have also tried to explain the mechanisms by which variation is generated at microsatellite loci. Further studies, of course, need to be carried out at more SSR loci representing genes and other regions in the genome to have concrete evidence and proof to put forth a specific mechanism underlying the generation of microsatellites and their variation leading to phylogenetic evolution in the genus *Oryza*.

REFERENCES

Abenes MLP, Angeles ER, Khush GS, Huang N (1993) Selection of bacterial blight resistant rice plants in the F2 generation via their linkage to molecular markers. Rice Genet Newsl 10: 120-123

Aggarwal RK, Brar DS, Khush GS (1997) Two new genomes in the *Oryza* complex identified on the basis of molecular divergence analysis using total genomic DNA hybridization. Mol Gen Genet 254: 1-12

Aggarwal RK, Brar DS, Nandi S, Huang N, Khush GS (1999) Phylogenetic relationships among *Oryza* species revealed by AFLP markers. Theor Appl Genet 98: 120-1328

Ahn S, Tanksley SD (1993) Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci, USA 90: 7980-7984

Ahn SN, Bollich CN, Tanksley SD (1992) RFLP tagging of a gene for aroma in rice. Theor Appl Genet 84: 825-828

Ahn S, Anderson JA, Sorrells ME, Tanksley SD (1993) Homeologous relationships of rice, wheat and maize chromosomes. Mol Gen Genet 241: 483-490

Akagi H, Yohozeki Y, Inagaki A, Fujimura T (1996a) Microsatellite DNA markers for rice chromosomes. Theor Appl Genet 7: 1071-1077

Akagi H, Yohozeki Y, Inagaki A, Nakamura A, Fujimura T (1996b) A co-dominant DNA marker closely linked to the rice nuclear restorer gene Rf-1, identified with inter-SSR fingerprinting. Genome 39: 1205-1209

Akagi H, Yohozeki Y, Inagaki A, Fujimura T (1997) Highly polymorphic microsatellites of rice consist of AT repeats, and a classification of closely related cultivars with these microsatellite loci. Theor Appl Genet 94: 61-67

Ali ML, Pathan MS, Zhang J, Bai G, Sarkarung S, Nguyen HT (2000) Mapping of QTLs for root traits in a recombinant inbred population from two indica ecotypes in rice. Theor Appl Genet 101: 756-76

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410

Amos W, Sawcer SJ, Feakes RW, Rubinsztein DC (1996) Microsatellites show mutational bias and heterozygote instability. Nat Genet 13: 390-391

Antonio BA, Inoue T, Kajiya H, Nagamura Y, Kurata N, Minobe Y, Yano M, Nakagahra M, Sasaki T (1996) Comparison of genetic distance and order of DNA markers in five populations of rice. Genome 39: 946-956

Armour JAL, Neumann R, Gobert S, Jeffreys AJ (1994) Isolation of human simple repeat loci by hybridization selection. Hum Mol Genet 3: 599-605

Arumunagathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol 9: 208-218

Arunachalam V (2000) Paradigms of participatory enhancement of rice productivity: suggestions based on a case study in India. Current Sci 79: 369-372

Ashikawa I, Kurata N, Nagamura Y, Minobe Y (1994) Cloning and mapping of telomere-associated sequences from rice. DNA Res 1: 67-76

Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD (1997) Microsatellites and a single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theor Appl Genet 94: 773-781

Babujee L, Gnanamanickam SS (2000) Molecular tools for characterization of rice blast pathogen (*Magnaporthe grisea*) population and molecular marker-assisted breeding for disease resistance. Curr Sci 78: 248-257

Bao JS, Zheng XW, Xia YW, He P, Shu QY, Lu X, Chen Y, Zhu LH (2000) QTL mapping for the paste viscosity characteristics in rice (*Oryza sativa* L.). Theor Appl Genet 100: 280-284

Barbier P, Ishiyama A (1990) Variation in the nucleotide sequences of a prolamin gene family in wild rice. Plant Mol Biol 15: 191-195

Barbier P, Morishima H, Ishiyama A (1991) Phylogenetic relationships of annual and perennial wild rice: probing by direct DNA sequencing. Theor Appl Genet 81: 693-702

Barnes SR (1991) RFLP analysis of complex traits in crop plants. Symp Soc Exp Biol 45: 219-228

Beckmann JS, Soller M (1986) Restriction fragment length polymorphisms in genetic improvement of agricultural species. Euphytica 35: 111-124

Beckman JS, Soller M (1990) Towards a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellites. Bio/technol 8: 930-932

Beer SC, Goffreda J, Phillips TD, Murphy JP, Sorrells ME (1993) Assessment of genetic variation in *Avena sterilis* using morphological traits, isozymes and RFLPs. Crop Sci 33: 1386-1393

Bell CJ, Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis.* Genomics 19: 137-144

Bennett MD, Smith JB (1991) Nuclear DNA amount in angiosperms. Phil Trans R Soc London (Biol) 344: 309-345

Bennetzen JL, Kellogg EA (1997) Do plants have a one way ticket to genome obesity? Plant Cell 9: 1509-1514

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

Blair MW, McCouch SR (1997) Microsatellite and sequence-tagged site markers diagnostic for the rice bacterial leaf blight resistance gene *xa-5*. Theor Appl Genet 95: 174-184

Blair MW, Panaud O, McCouch SR (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

Blanquer-Maumont A, Crouau-Roy B (1995) Polymorphism, monomorphism, and sequences in conserved microsatellites in Primates species. J Mol Evol 41: 492-497

Bligh HFJ, Till RI, Jones CA (1995) A microsatellite sequence closely linked to the *Waxy* gene of *Oryza sativa*. Euphytica 86: 83-85

Bligh HFJ, Blackhall NW, Edwards KJ, McClung AM (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

Bohn M, Utz HF, Melchinger AE (1999) Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs and SSRs and their use for predicting progeny variance. Crop Sci 39: 228-237

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution human evolutionary trees with polymorphic microsatellites. Nature 368: 455-457

Brar DS, Khush GS (1997) Alien introgression in rice. Plant Mol Biol 35: 35-47

Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, Sutherland GR (1993) Incidence and origin of null alleles in the $(AC)_n$ microsatellite markers. Am J Hum Genet 52: 922-927

Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A, Zara F, Canizares J, Koutnikova H, Bidichandani S, Gellera C, Brice A, Trouillas P, DeMichele G, Filla A, De Frutos R, Palau F, Patel PI, DiDonato S, Mandel JL, Cocozza S, Koenig M, Pandolfo M (1996) Friedreich Ataxia. The first autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 271: 1423-1427

Carmona PS (1990) Contribution of INGER (International Network for the Genetic Evaluation of Rice) to broaden the genetic base of rice in the state of Rio Grande do Sul, Brazil. In: INGER-Latin America Report. First semester, 1990, Cali, Colombia; Bilingual Edition. pp 153-158

Carter TE Jr, Gizlice Z, Burton JW (1993) Coefficient of parentage and genetic similarity estimates for 258 North American public cultivars released during 1945-88. USDA Tech Bull, 1814, US Gov Print Office, Washington DC

Caskey CT, Pizzuti A, Fu YH, Fenwick RG, Nelson DL (1992) Triplet repeat mutations in human diseases. Science 256: 784-788

Causse MA, Fulton TM, Cho YG, Ahn SN, Chungwongse J, Wu K, Xiao J, Yu Z, Ronald PC, Harrington SE, Second G, McCouch SR, Tanksley SD (1994) Saturated molecular map of the rice genome based on an inter-specific backcross population. Genetics 138: 1251-1274

Chakraborty R, Deka R, Jin L, Ferrell RE (1992) Allele sharing at six VNTR loci and genetic distances among three ethnically defined populations. Am J Hum Biol 4: 387-397

Chang TT (1976a) Rice. In: Evolution of crop plants, NW Simmonds (ed), Longman, London. pp 98-104

Chang TT (1976b) The origin, evolution, cultivation, dissemination and diversification of cultivated rices. Euphytica: 25: 425-441

Chang TT (1984) Conservation of rice genetic resources: luxury or necessity? Science 224: 251-256

Charcosset A, Lefort-Buson M, Gallais A (1991) Relationship between heterosis and heterozygosity at marker loci; a theoretical computation. Theor Appl Genet 81: 571-575

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371: 215-220

Chen ZL, Qu LJ (1997) Plant molecular biology in China. Plant Mol Biol Reporter 15: 273-277

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

Chen M, San Miguel P, de Oliveira AC, Woo SS, Zhang H, Wing RA, Bennetzen JL (1997b) Microcolinearity in *sh2*-homologous regions of the maize, rice and sorghum genomes. Proc Natl Acad Sci 94: 3431-3435

Chen M, San Miguel P, Bennetzen JL (1998) Sequence organization and conservation in *sh2ia1*-homologous regions of sorghum and rice. Genetics 148: 435-443

Chen S, Lin XH, Xu CG, Zhang Q (2000) Improvement of bacterial blight resistance of 'Minghui 63', an elite restorer line of hybrid rice, by molecular marker-assisted selection. Crop Sci 40: 239-244

Cho YG, Eun MY, McCouch SR, Chae YA (1994) The semi-dwarf gene, *sd*-1, of rice (*Oryza sativa* L.). II. Molecular mapping and marker-assisted selection. Theor Appl Genet 89: 54-59

Cho YG, Blair MW, Panaud O, McCouch SR (1996) Cloning and mapping of variety-specific rice genomic DNA sequences: amplified fragment length polymorphisms (AFLP) from silver-stained polyacrylamide gels. Genome 39: 373-378

Cho YG, McCouch SR, Kuiper M, Kang MR, Pot J, Groenen JTM, Eun MY (1998) Integrated map of AFLP, SSLP and RFLP markers using a recombinant inbred population of rice (*Oryza sativa* L.) Theor Appl Genet 97: 370-380

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100: 713-722

Chowdari KV (1998) Genetic diversity analysis in rice and pearl millet using DNA markers. PhD thesis

Chowdari KV, Davierwala AP, Gupta VS, Ranjekar PK, Govila OP (1998a) Genotype identification and assessment of genetic relationships in pearl millet [*Pennisetum glaucum* (L.) R. Br.] using microsatellites and RAPDs. Theor Appl Genet 97: 154-162

Chowdari KV, Ramakrishna W, Tamhankar SA, Hendre RR, Gupta VS, Sahasrabudhe NA, Ranjekar PK (1998b) Identification of minor DNA variations in rice somaclonal variants. Plant Cell Rep 18: 55-58

Chowdari KV, Venkatachalam SR, Davierwala AP, Gupta VS, Ranjekar PK, Govila OP (1998c) Hybrid performance and genetic distance as revealed by the $(GATA)_4$ microsatellite and RAPD markers in pearl millet. Theor Appl Genet 97: 163-169

Chunwongse J, Martin GB, Tanksley SD (1993) Pre-germination genotypic screening using PCR amplification of half seeds. Theor Appl Genet 86: 694-698

Collins FS (1992) Positional cloning, let's not call it reverse anymore. Nat Genet 1: 3-6

Coltman DW, Bowen WD, Wright JM (1996) PCR primers for harbour seal (*Phoca vitulina concolour*) microsatellites amplify polymorphic loci in other pinniped species. Mol Ecol 5: 161-163

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 TS, Lookhart GL, Walker DE, Harrell LG, Albers LD, Rogers DM (1985) Genetic relationships among hard red winter wheat cultivars as evaluated by pedigree analysis and gliadin polyacrylamide gel electrophoretic patterns. Crop Sci 25: 1058-1063

Crouau-Roy B, Service S, Slatkin M, Freimer N (1996) A fine scale comparison of the human and chimpanzee genomes: linkage, linkage disequilibrium and sequence. Hum Mol Genet 5: 1131-1137

Cuevas-Perez, FE, Guimaraes EP, Berrio LE, Gonzalez DI (1992) Genetic base of irrigated rice in Latin America and the Caribbean, 1971 to 1989. Crop Sci 32: 1054-1058

Cummings CJ, Zoghbi HY (2000) Fourteen and counting: unraveling trinucleotide repeat diseases. Hum Mol Genet 9: 909-916

Dallas JF (1988) Detection of DNA "fingerprints" of cultivated rice by hybridization with human minisatellite DNA probe. Proc Natl Acad Sci 85: 6831-6835

Dallas JF, McIntyre CL, Gustafson JP (1993) An RFLP species-specific DNA sequence for the A genome of rice. Genome 36: 445-448

Dally AM, 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

Deka R, Jin L, Shriver MD, Yu LM, Decroo S, Hundrieser J, Bunker CH, Ferrell RE, Chakraborty R (1995) Population genetics of dinucleotide $(dC\text{-}dA)_n$ $(dC\text{-}dT)_n$ polymorphisms in world populations. Am J Hum Genet 56: 461-474

Delannay X, Rodgers DM, Palmers RG (1983) Relative genetic contributions among ancestral lines to North American soybean cultivars. Crop Sci 23: 944-949

Deragon JM, Landry BS (1992) RAPD and other PCR-based analyses of plant genomes using DNA extracted from small leaf discs. PCR Methods Appl 1: 175-180

Deshpande VG, Ranjekar PK (1980) Repetitive DNA in three Gramineae species with low DNA content. Hoppe-Seyler Z Physiol Chem 361: 1223-1233

Deshpande AD, Ramakrishna W, Mulay GP, Gupta VS, Ranjekar PK (1998) Evolutionary and polymorphic organization of the *knotted-1* homeobox in cereals. Theor Appl Genet 97: 135-140

Devos KM, Gale MD (1997) Comparative genetics in the grasses. Plant Mol Biol 35: 3-15

Devos KM, Atkinson MD, Chinoy CN, Francis HA, Harcourt RL, Koebner RMD, Liu CJ, Masojc P, Xie DX, Gale MD (1993) Chromosomal rearrangements in the rye genome relative to that of wheat. Theor Appl Genet 85: 573-680

Dilday RH (1990) Contribution of ancestral lines in the development of new cultivars of rice. Crop Sci 30: 905-911

Dong NV, Subudhi PK, Luong PN, Quang VD, Quy TD, Zheng HG, Wang B, Nguyen HT (2000) Molecular mapping of a rice gene conditioning thermosensitive genic male sterility using AFLP, RFLP and SSR techniques. Theor Appl Genet 100: 727-734

Dudley JW (1993) Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Sci 33: 660-668

Dunford RP, Kurata N, Laurie DA, Money TA, Minobe Y, Moore G (1995) Conservation of fine-scale DNA marker order in the genomes of rice and the Triticeae. Nucl Acids Res 23: 2724-2728

Edwards MD, Stuber CW, Wendel JF (1987) Molecular facilitated investigations of quantitative trait loci in maize. I. Number, genomic distribution and types of gene action. Genetics 116: 113-125

Edwards A, Hammond HA, Jim L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12: 241-253.

Engel SR, Linn RA, Taylor JF, Davis SK (1996) Conservation of microsatellite loci across species of Artiodactyls – implications for population studies. J Mammal 77: 504-508

Epplen JT (1988) On simple repeat $GA^{T}/_{C}A$ sequences in animal genomes: a critical reappraisal. J Hered 79: 409-417

Epplen JT (1992) The methodology of multilocus DNA fingerprinting using radioactive or nonradioactive oligonucleotide probes specific for simple repeat motifs. In: Advances in Electrophoresis, Chrambach A, Dunn MJ, Radola BJ (eds), VCH Publishers Inc, New York, USA. vol 5, pp 62-112

Epplen JT, McCarrey JR, Sutou S, Ohno S (1982) Base sequence of a cloned snake W-chromosome DNA fragment and identification of a male-specific putative mRNA in the mouse. Proc Natl Acad Sci, USA 79: 3798-3802

Erickson RP, Ross CE, Gorski JL, Stalvey JRD, Drumm MM (1988) Bkm sequences from the human X chromosome contains large clusters of GATA / GACA repeats. Ann Hum Genet 52: 167-176

Falconer DS (1960) Introduction to Quantitative Genetics. Ronald Press Co, New York, USA

Field D, Willis C (1996) Long, polymorphic microsatellites in simple organisms. Proc R Soc Lond 263: 209-215

Fitzsimmons NN, Moritz S, Moore SS (1995) Conservation and dynamics of microsatellite loci over 300 Myr of marine turtle evolution. Mol Biol Evol 12: 432-440

Flavell RB (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 retrotransposon probes. Jpn J Genet 68: 195-204

Fukuoka S, Hosaka K, Kamijima O (1992) Use of random amplified polymorphic DNAs (RAPDs) for identification of rice accessions. Japan J Genet 67: 243-252

Garland S, Lewin L, Blakeney A, Reinke R, Henry R (2000) PCR-based molecular markers for the fragrance gene in rice (*Oryza sativa* L.). Theor Appl Genet 101: 364-371

Garza JC, Freimer NB (1996) Homoplasy for size at microsatellite loci in humans and chimpanzees. Genome Res 6: 211-217

Garza JC, Slatkin M, Freimer NB (1995) Microsatellite allele frequencies in humans and chimpanzees with implications for constraints on allele size. Mol Biol Evol 12: 594-603

Ge S, Sang T, Lu BR, Hong DY (1999) Phylogeny of rice genomes with emphasis on origins of allotetraploid. Proc Natl Acad Sci 96: 14400-14405

Gemmell NJ, Allen PJ, Goodman SJ, Reed JZ (1997) Inter-specific microsatellite markers for the study of pinniped populations. Mol Ecol 6: 661-666

Ghareyazie B, Huang N, Second G, Bennett J, Khush GS (1995) Classification of rice germplasm. I. Analysis using ALP and PCR-based RFLP. Theor Appl Genet 91: 218-227

Gilbert DA, Lehman N, O'Brien SJ, Wayne RK (1990) Genetic fingerprinting reflects population differentiation in the Channel Island fox. Nature 344: 764-767

Glaszmann JC (1987) Isozymes and classification of Asian rice varieties. Theor Appl Genet 74: 21-30

Goff SA (1999) Rice as a model for cereal genomics. Curr Opin Plant Biol 2: 86-89

Goldberg RB, Hoschek G, Kamaley JC (1978) Sequence complexity of nuclear and polysomal RNA in leaves of the tobacco plant. Cell 14: 123-131

Goldstein DB, Clark AG (1995) Microsatellite variation in North American populations of *Drosophila melanogaster*. Nucl Acids Res 23: 3882-3886

Goldstein DB, Pollock DD (1997) Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. J Hered 88: 335-342

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139: 463-471

Goodfellow PN (1992) Variation is now the theme. Nature 359: 777-778

Goodwin ID, Aitken EAB, and Smith LW (1997) Application of inter-simple sequence repeat (ISSR) markers to plant genetics. Electrophoresis 18: 1524-1528.

Graner A, Ludwig WF, Melchinger AE (1994) Relationships among European barley germplasm: II. Comparison of RFLP and pedigree data. Crop Sci. 34: 1199-1205

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

Gupta VS, Ramakrishna W, Rawat SR, Ranjekar PK (1994) $(CAC)_5$ detects DNA fingerprints and sequences homologous to gene transcripts in rice. Biochem Genet 32: 1-8

Gupta PK, Balyan IS, Sharma PC, Ramesh B (1996) Microsatellite in plants - a new class of molecular markers. Curr Sci 70: 45-54

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

Hallauer AR and Miranda JB (1988) Quantitative genetics in maize breeding. Iowa State University Press, Ames, USA.

Han F, Kilian A, Chen JP, Kudrna D, Steffenson B, Yamamoto K, Matsumoto T, Sasaki T, Kleinhofs A (1999) Sequence analysis of a rice BAC covering the syntenous barley *Rpg*1 region. Genome 42: 1071-1076

Hancock JM (1996) Simple sequences and the expanding genome. Bioessays 18: 421

Harrington SE, Bligh HFJ, Park WD, Jones CS, McCouch SR (1997) Linkage mapping of starch branching enzyme III in rice (*Oryza sativa* L.) and prediction of location of orthologous genes in other grasses. Theor Appl Genet 94: 564-568

Henderson ST, Petes TD (1992) Instability of simple sequence DNA in *Saccharomyces cerevisiae*. Mol Cell Biol 12: 2749-2757

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, Mew T, Rodriguez R, Huang N (1994) Identification of blast resistance gene, *Pi-2(t)* in rice plants by flanking DNA markers. Rice Genet Newsl 11: 144-146

Hittalmani S, Foolad MR, Mew T, Rodriguez RL, Huang N (1995) Development of a PCR-based marker to identify rice blast resistance gene, *Pi-2(t)* in a segregating population. Theor Appl Genet 91: 9-14

Hittalmani S, Kumar GK, Kulkarni N, Sashidhar HE (1999) DNA markers assist in reducing the number of generations of backcrosses in breeding for blast resistance in rice. In: Proceedings of the International Program on Rice Biotechnology, Sept 20-24, Phuket, Thailand. pp 21

Hittalmani S, Parco A, Mew TV, Zeigler RS, Huang N (2000) Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. Theor Appl Genet 100: 1121-1128

Hu CH (1970) Cytogenetic studies of *Oryza officinalis* complex. III. The genomic constitution of *O.punctata* and *O.eichengeri*. Cytologia 35: 304-318

Hu J, Quiros CF (1991) Identification of broccoli and cauliflower cultivars with RAPD markers. Plant Cell Rep 10: 505-511

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

Huang N, McCouch S, Mew T, Parco A, Guideroni E (1994) Development of an RFLP map from a doubled haploid population in rice. Rice Genet Newsl 11: 134-137

Huang L, Millet E, Feldman M (1996) Genetic relationships among Mediterranean common wheat cultivars as determined by RFLP and parentage analyses. J Genet Breed 50: 377-382

Huang N, Angeles ER, Domingo J, Magpantay G, Singh S, Zhang G, Kumaravadivel N, Bennett J, Khush GS (1997) Pyramiding bacterial blight resistance genes in rice: marker assisted selection using RFLP and PCR. Theor Appl Genet 95: 313-320

Ikeda R, Wakasa K (1997) Frontier of rice breeding by utilization of genetic resources and biotechnology. In: Plant biotechnology and plant genetic resources for sustainability and productivity, Watanabe KN, Pehu E (eds), RG Landes Company, Austin. pp 103-115

Innan H, Terauchi R, Miyashita T (1997) Microsatellite polymorphism in natural populations of the wild plant *Arabidoposis thaliana*. Genetics 146: 1441-1452

Inukai T, Mackill DJ, Bonman JM, Sarkarung S, Zeigler RS, Nelson RJ, Takamure I, Kinoshita T (1992) A blast resistance gene *Pi-3 (t)* in near-isogenic line C104PKT. Rice Genet Newsl 9: 94-95

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

Ishii T, McCouch SR (2000) Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Gramineae species members. Theor Appl Genet 100: 1257-1266

Ishii T, Terachi T, Tsunewaki K (1988) Restriction endonuclease analysis of chloroplast DNA from A-genome diploid species of rice. Jpn J Genet 63: 523-536

Ishii T, Brar DS, Multani DS, Khush GS (1994) Molecular tagging of genes for brown planthopper resistance and earliness introgressed from *Oryza australiensis* into cultivated rice, *O.sativa*. Genome 37: 217-221

Ishii T, Brar DS, Second G, Tsunewaki K, Khush GS (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

Jackson MT (1997) Conservation of rice genetic resources: the role of the International Rice Genebank at IRRI. Plant Mol Biol 35: 61-67

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. Trends Ecol Evol 11: 424-429

Jena KK, Khush GS, Kochert G (1992) RFLP analysis of rice (*Oryza sativa* L.) introgression lines. Theor Appl Genet 84: 608-616

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

Joshi SP, Ranjekar PK, Gupta VS (1999) Molecular markers in plant genome analysis. Curr Sci 77: 230-240

Joshi SP, Gupta VS, Aggarwal RK, Ranjekar PK, Brar DS (2000) Genetic diversity and phylogenetic relationship as revealed by inter-simple sequence repeat (ISSR) polymorphisms in the genus *Oryza*. Theor Appl Genet 100: 1311-1320

Kajiya H, Shimano T, Monna L, Yano M, Sasaki T (1996) Effective production of RFLP markers between closely related varieties of rice by genomic subtraction. Breed Sci 46: 387-391

Kaneda C (1985) Development of very high-yield rice varieties. Farming-Japan 19: 25-29

Karhu A, Dieterich JH, Savolainen O (2000) Rapid expansion of microsatellite sequences in pines. Mol Biol Evol 17: 259-265

Katayama T (1982) Cytological studies on the genus *Oryza*, XIII. Relationship between the genomes E and D. Jpn J Genet 57: 613-621

Katiyar SK, Chandel G, Verulkar SB, Huang B, Bennett J (1999) Molecular approaches for the enhancement of gall midge resistance in rice: present status. In: Proceedings of the International Program on Rice Biotechnology, Sept 20-24, Phuket, Thailand. pp 87

Kawase M (1994) Application of the restriction landmark genomic scanning (RLGS) method to rice cultivars as a new fingerprinting technique. Theor Appl Genet 89: 861-864

Kay SA, Keith B, Shinozaki K, Chye ML, Chua NH (1989) The sequence of the rice phytochrome gene. Nucl Acids Res 17: 2865-2866

Kempthorne O (1969) An introduction to genetic statistics. Iowa State University Press, Ames, Iowa

Kennard WC, Phillips RA, Porter RA, Grombacher AW (2000) A comparative map of wild rice (*Zizania palustris* L. 2n=2x=30). Theor Appl Genet 101: 677-684

Khush GS (1997) Origin, dispersal, cultivation and variation of rice. Plant Mol Biol 35: 25-34

Khush GS (1995) Modern varieties – their real contribution to food supply and equity. GeoJournal 35.3: 275-284

Khush GS, Brar DS, Zapata FJ, Nelson R, McCouch S, Bottrell DG (1993) Biotechnology for rice improvement. In: Proceedings of the Tenth Australian Plant Breeding Conference, April 18-23, Vol. 1. Focused Plant Improvement: Towards Responsible and Sustainable Agriculture.

Kiefer-Meyer MC, Reddy AS, Delsney M (1995) Characterization of a dispersed repetitive DNA sequence associated with the CCDD genome of wild rice. Genome 38: 681-688

Kijas JM, Fowler JC, Thomas MR (1995) An evaluation of sequence tagged microsatellite site markers for genetic analysis within Citrus and related species. Genome 38: 349-355

Kilian A, Kudrna DA, Kleinhofs A, Yano M, Kurata N, Steffenson B, Sasaki T (1995) Rice-barley synteny and its applications to saturation mapping of the barley *Rpg*1 region. Nucl Acids Res 23: 2729-2733

Kilian A, Chen J, Han F, Steffenson B, Kleinhofs A (1997) Towards map-based cloning of the barley stem rust resistance genes *Rpg*1 and *rpg*4 using rice as an intergenomic cloning vehicle. Plant Mol Biol 35: 187-195

Kim HS, Ward RW (1997) Genetic diversity in eastern US soft winter wheat (*Triticum aestivum* L. em. Thell.) based on RFLPs and coefficients of parentage. Theor Appl Genet 94: 472-479

Kim NS, Shim KM, Yoon KM, Shin YB, Cho YS, Eun MY, Cho YG (1996) Polymorphism and phylogenetic analyses in rice cultivated in Korea by RAPDs and RFLPs. Korean J Genetics 18: 211-224

Kimura M, Crow J (1964) The number of alleles that can be maintained in a finite population. Genetics 49: 725-738

Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. Proc Natl Acad Sci, USA 75: 2868-2872

Kinoshita T (1995) Report of committee on gene symbolization. Rice Genet Newsl 12: 9-153

Knauft DA, Gorbet DW (1989) Genetic diversity among peanut cultivars. Crop Sci 29: 1417-1422

Ko HL, Cowan DC, Henry RJ, Graham GC, Blakeney AB, Lewin LG (1994) Random amplified polymoprhic DNA analysis of Australian rice (*Oryza sativa* L.) varieties. Euphytica 80: 179-189

Koh HJ, Heu MH, McCouch SR (1996) Molecular mapping of the gene controlling the super-giant embryo character in rice (*Oryza sativa* L.). Theor Appl Genet 93: 257-261

Kojima T, Nagaoka T, Noda K, Ogihara Y (1998) Genetic linkage map of ISSR and RAPD markers in Eikorn wheat in relation to that of RFLP markers. Theor Appl Genet 96: 37-45

Kondo Y, Mori M, Kuramoto T, Yamada J, Beckmann JS, Simonchazottes D, Montagutelli X, Guenet JL, Serikawa T (1993) DNA fragments mapped by reciprocal use of microsatellite primers between mouse and rat. Mamm Genome 4: 571-576

Kuhl DP, Caskey CT (1993) Trinucleotide repeats and genome variation. Curr Opin Genet Dev 3: 404-407

Kumar LS (1999) DNA markers in plant improvement: an overview. Biotechnol Adv 17: 143-182

Kurata N, Moore G, Nagamura Y, Foote T, Yano M, Minobe Y, Gale M (1994a) Conservation of genome structure between rice and wheat. Bio/technol 12: 276-278

Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shomura A, Shimizu T, Lin SY, Inoue T, Fukuda A, Shimano T, Kuboki Y, Toyama T, Miyamoto Y, Kirihara T, Hayasaka K, Miyao A, Monna L, Zhong HS, Tamura Y, Wang Z-X, Momma T, Umehara Y, Yano M, Sasaki T, Minobe Y (1994b) A 300 kilobase interval genetic map of rice including 883 expressed sequences. Nature Genet 8: 365-372

Kurata N, Umehara Y, Tanoue H, Sasaki T (1997) Physical mapping of the rice genome with YAC clones. Plant Mol Biol 35: 101-113

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185-199

Lang NT, Subudhi PK, Viramani 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

Langercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucl Acids Res 21: 1111

Lee M (1995) DNA markers and plant breeding programs. Adv Agron 55: 265-344

Levinson G, Gutman GA (1987) Slipped strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4: 203-221

Li Z, Pinson SRM, Stansel JW, Park WD (1995a) Identification of quantitative trait loci (QTLs) for heading date and plant height in cultivated rice (*Oryza sativa* L.). Theor Appl Genet 91: 374-381

Li Z, Pinson SRM, Marchetti MA, Stansel JW, Park WD (1995b) Characterization of quantitative trait loci (QTLs) in cultivated rice to field resistance to sheath blight (*Rizoctonia solani*). Theor Appl Genet 91: 382-388

Li Z, Pinson SRM, Park WD, Paterson AH, Stansel JW (1997) Epistasis for three grain yield components in rice (*Oryza sativa* L.). Genetics 145: 453-465

Li ZK, Luo LJ, Mei HW, Paterson AH, Zhao XH, Zhong DB, Wang YP, Yu XQ, Zhu L, Tabien R, Stansel JW, Ying CS (1999) A "defeated" rice resistance gene acts as a QTL against a virulent strain of *Xanthomonas oryzae* pv *oryzae*. Mol Gen Genet 26: 58-63

Li SG, Ma YQ, Wang WM, Liu GQ, Zhou KD, Zhu LH (2000) Molecular tagging of a new recessive gene for late heading in a rice cultivar 8987. Acta Genet Sin 27: 133-138

Liang CZ, Gu MH, Pan XB, Liang GH, Zhu LH (1994) RFLP tagging of a new semi-dwarfing gene in rice. Theor Appl Genet 88: 898-900

Lin MS (1991) Genetic base of japonica rice varieties released in Taiwan. Euphytica 56: 43-46.

Lin SC, Yuan LP (1980) Hybrid rice breeding in China. In: Innovative approaches to rice breeding, International Rice Research Institute (ed), IRRI, Manila, Philippines. pp 35-51

Lin XH, Zhang DP, Xie YF, Gao HP, Zhang Q (1996) Identifying and mapping a new gene for bacterial blight resistance in rice based on RFLP markers. Phytopathol 86: 1156-1159

Lin HX, Yamamoto T, Sasaki T, Yano M (2000) Characterization and detection of epistatic interactions of 3 QTLs, Hd1, Hd2, and Hd3, controlling heading date in rice using nearly isogenic lines. Theor Appl Genet 101: 1021-1028

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

Liu A, Zhang Q, Li H (1992) Location of a gene for wide compatibility in the RFLP linkage map. Rice Genet Newsl 9: 134-136

Liu KD, Yang GP, Zhu SH, Zhang Q, Wang XM, Saghai Maroof MA (1996) Extraordinarily polymorphic ribosomal DNA in wild and cultivated rice. Genome 39: 1109-1116

Liu KD, Wang J, Li HB, Xu CG, Liu AM, Li XH, Zhang Q (1997) A genome-wide analysis of wide compatibility in rice and the precise location of the S_5 locus on the molecular map. Theor Appl Genet 95: 809-814

Lorieux M, Ndjiondjop MN, Ghesquiere A (2000) A first inter-specific *Oryza sativa* X *O.glaberrima* microsatellite-based genetic linkage map. Theor Appl Genet 100: 593-601

Love JM, Knight AM, McAleer MA, Todd JA (1990) Towards construction of a high resolution map of the mouse genome using PCR-analyzed microsatellites. Nucl Acids Res 18: 4123-4130

Lu C, Li X, Zhu L, Zhang Q, Yang W, Zhao B, Wang C (1996) A PCR marker-based selection for resistance to bacterial blight in rice. I Chuan Hseuh Pao 23: 110-116

Mackill DJ (1995) Classifying japonica rice cultivars with RAPD markers. Crop Sci 35: 889-894

Mackill DJ, Salam MA, Wang ZY, Tanksley SD (1993) A major photoperiod-sensitivity gene tagged with RFLP and isozyme markers in rice. Theor Appl Genet 83: 536-540

Mackill DJ, Zhang Z, Redona ED, Colowit PM (1996) Level of polymorphism and genetic mapping of AFLP markers in rice. Genome 39: 969-977

Maheshwaran M, Subudhi PK, Nandi S, Xu JC, Parco A, Yang DC, Huang N (1997) Polymorphism, distribution and segregation 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

Martin C, Juliano A, Newbury HJ, Lu BR, Jackson MT, Ford-Lloyd BV (1997) The use of RAPD markers to facilitate the identification of *Oryza* species. Genetic Resources Crop Evol 44: 175-183

Maughan PJ, Saghai Maroof MA, Buss GR (1995) Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. Genome 38: 715-723

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD (1988) Molecular mapping of rice chromosomes. Theor Appl Genet 76: 815-829

McCouch SR, Khush GS, Tanksley SD (1991) Tagging new genes for disease and insect resistance via linkage to RFLP markers. In: Rice genetics II. International Rice Research Institute, Los Banos, Philippines. pp 443-449

McCouch SR, Abenes ML, Angeles ER, Khush GS, Tanksley SD (1992) Molecular tagging of a recessive gene, *xa-5*, for resistance to bacterial blight of rice. Rice Genet Newsl 8: 143-145

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

Meisser W, Li HS, Stewart CB (1996) The birth of microsatellites. Nature 381: 483

Melchinger AE, Graner A, Singh M, Messmer MM (1994) Relationships among winter and spring cultivars revealed by RFLPs. Crop Sci 34: 1191-1199

Messmer MM, Melchinger AE, Hermann RG, Boppermaier J (1993) Relationships among early European maize inbreds: II. Comparison of pedigree and RFLP data. Crop Sci 33: 944-950

Miyamoto M, Ando I, Rybka K, Kodama O, Kawasaki S (1996) High resolution mapping of the *indica*-derived rice blast resistance genes. I. *Pi*-b. Mol Plant Microbe Interact 9: 6-13

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

Mochizuki K, Ohtsubo H, Hirano HY, Sano Y, Ohtsubo E (1993) Classification and relationships of rice strains with A genome by identification of transposable elements at nine loci. Jpn J Genet 68: 205-217

Mohan M, Nair S, Bentur JS, Rao UP, Bennett J (1994) RFLP and RAPD mapping of the rice *Gm*2 gene that confers resistance to biotype 1 of gall midge (*Orseolia oryzae*). Theor Appl Genet 87: 782-788

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

Mohan M, Sathyanarayanan PV, Kumar A, Srivastava MN, 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

Monna L, Miyao A, Inoue T, Fukuoka S, Yamazaki M, Sun Zhong H, Sasaki T, Minobe Y (1994) Determination of RAPD markers in rice and their conversion into sequence-tagged sites (STSs) and STS-specific primers. DNA Res 1: 139-148

Monna L, Miyao A, Zhong HS, Sasaki T, Minobe Y (1995) Screening of RAPD markers linked to the photoperiod-sensitivity gene in rice chromosome 6 using bulked segregant analysis. DNA Research 2: 101-106

Moore G (1995) Cereal genome evolution; pastoral pursuits with 'Lego' genomes. Curr Opinion in Genet and Development 5: 717-724

Moore SS, Sargeant LL, King TJ, Mattick JS, Georges M, Hetzel DJS (1991) The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. Genomics 10: 654-660

Moore G, Gale MD, Kurata N, Flavell RB (1993) Molecular analysis of small grain cereal genomes: current status and prospects. Bio/technol 11: 584-589

Morchen M, Cuguen J, Michaelis G, Hanni C, Saumitou-Laprade P (1996) Abundance and length polymorphism of microsatellite repeats in *Beta vulgaris* l. Theor Appl Genet 92: 326-333

Morgante M, Olivieri AM (1993) PCR-amplified microsatellite markers in plant genetics. Plant J 3: 175-182

Morgante M, Rafalski A, Biddle P, Tingey S, Olivieri AM (1994) Genetic mapping and variability of seven soybean simple sequence repeat loci. Genome 37: 763-769.

Morin PA, Moore JJ, Chakraborty R, Jin L, Goodall J, Woodruff DS (1994) Kin selection, social structure, gene flow and evolution of chimpanzees. Science 265: 1193-1201

Morinaga T (1964) Cytological investigations on *Oryza* species. In: Rice genetics and cytogenetics, IRRI (ed), Elsevier, Amsterdam. pp 91-103

Morishima H, Hinata K, Oka HI (1963) Comparison of modes of evolution of cultivated forms from two wild rice species, *Oryza breviligulata* and *O.perennis*. Evolution 17: 170-181

Mundt CC (1990) Probability of mutation to multiple virulence and durability of resistance gene pyramids. Phytopathol 80: 221-223

Murphy JP, Cox TS, Rodgers DM (1986) Cluster analysis of red winter wheat cultivars based upon coefficients of parentage. Crop Sci 26: 672-676

Nagamura Y, Antonio BA, Sasaki T (1997) Rice molecular genetic map using RFLPs and its applications. Plant Mol Biol 35: 79-87

Nagano H, Kawasaki S, Kishima Y, Sano Y (2000) Structural differences in the vicinity of the *waxy* locus among the *Oryza* species with the AA genome: identification of variable regions. Theor Appl Genet 100: 376-383

Nair S, Bentur JS, Prasada Rao U, Mohan M (1995a) DNA markers tightly linked to a gall midge resistance gene (*Gm2*) are potentially useful for marker-aided selection in rice breeding. Theor Appl Genet 91: 68-73

Nair S, Kumar A, Srivastava MN, Mohan M (1995b) PCR-based DNA markers linked to a gall midge resistance gene, *Gm*4t has potential for marker-aided selection in rice. Theor Appl Genet 92: 660-665

Nair S, Srivastava MN, Kumar A, Mohan M (1996) PCR-based DNA markers linked to a gall midge resistance gene, *Gm*4t has potential for marker-aided selection in rice. Theor Appl Genet 92: 660-665

Naqvi NI, Bonman JM, Mackill DJ, Nelson RJ, Chattoo BB (1995) Identification of RAPD markers linked to a major blast resistance gene in rice. Mol Breed 1: 341-348

Naqvi NI, Chattoo BB (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

Nauta MJ, Weissing FJ (1996) Constraints on allele size at microsatellite loci – implications for genetic differentiation. Genetics 143: 1021-1032

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci, USA 70: 3321-3323

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

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

Ohta T, Kimura M (1973) The model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population. Genet Res: 22: 201-204

Oka HI (1988) Origin of cultivated rice. In: Developments in crop science, 14, Elsevier / Japan Scientific Societies Press, Amsterdam / Tokyo. pp 254

Oka HI (1991) Genetic diversity of wild and cultivated rice. In: Rice Biotechnology. Biotechnology in Agriculture No 6, Khush GS, Toenniessen GH (eds), CAB International, Wallingford, UK and IRRI, Manila, Philippines. pp 55-81

Oka HI, Chang WT (1962) Rice varieties intermediate between wild and cultivated forms and the origin of the japonica type. Bot Bull Acad Sinica 3: 109-131

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

Olufowote JO, Xu Y, Chen X, Park WD, Beachell HM, Dilday RH, Goto M, McCouch SR (1997) Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. Genome 40: 370-378

Orti G, Pearse DE, Avise JC (1997) Phylogenetic assessment of length variation at a microsatellite locus. Proc Natl Acad Sci, USA 94: 10745-10749

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

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

Parasnis AS, Ramakrishna W, Chowdari KV, Gupta VS, Ranjekar PK (1999) Microsatellite $(GATA)_n$ reveals sex-specific differences in Papaya. Theor Appl Genet 99: 1047-1052

Parsons BJ, Newbury JH, Jackson MT, Ford-Lloyd BV (1997) Contrasting genetic diversity relationships are revealed in rice (*Oryza sativa* L.) using different marker types. Mol Breed 3: 115-125

Paterson AH (1995) Molecular dissection of quantitative traits: Progress and Prospects. Genome Research 5: 321-333

Paterson AH, Tanksley SA, Sorrells ME (1991) DNA markers in plant improvement. Adv Agron 46: 39-90

Paterson AH, Lin YR, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269: 1714-1718

Peakall R, Gilmore S, Keys W, Morgante M, Rafalski A (1998) Cross species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: Implications for the transferability of SSRs in plants. Mol Biol Evol 15: 1275-1287

Petes TD, Greenwell PW, Dominska M (1997) Stabilization of microsatellite sequences by variant repeats in the yeast *Saccharomyces cerevisiae*. Genetics 146: 491-498

Phelps TL, Hall AE, Buckner B (1996) Microsatellite repeat variation within the *y1* gene of maize and teosinte. J Hered 87: 396-399

Phillips WJ, Chapman CGD, Jack PL (1994) Molecular cloning and analysis of one member of a polymoprhic family of GACA-hybridizing DNA repeats in tomato. Theor Appl Genet 88: 845-851

Plaschke J, Ganal MW, Roder MS (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor Appl Genet 91: 1001-1007

Powell W, Morgante M, Andre C, McNicol JW, Machray GC, Doyle JJ, Tingey SV, Rafalski JA (1995) Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. Curr Biol 5: 1023-1029

Powell W, Machray GC, Provan J (1996a) Polymorphism revealed by simple sequence repeats. Trends in Plant Sci 1: 215-221

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996b) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2: 225-238

Prabhu RR, Webb D, Jessen H, Luk S, Smith S, Greshoff PM (1997) Genetic relatedness among soybean genotypes using DNA amplification fingerprinting (DAF), RFLP and pedigree. Crop Sci 37: 1590-1595

Prevost A, Wilkinson MJ (1999) A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. Theor Appl Genet 98: 107-112

Price AH, Steele KA, Moore BJ, Barraclough PP, Clark LJ (2000) A combined RFLP and AFLP linkage map of upland rice (*Oryza sativa* L.) used to identify QTLs for root-penetration ability. Theor Appl Genet 100: 49-56

Primmer CR, Moller AP, Ellegren H (1996a) A wide-range survey of cross-species microsatellite amplification in birds. Mol Ecol 5: 365-378

Primmer CR, Ellegren H, Saino N, Moller AP (1996b) Directional evolution in germline microsatellite mutations. Nat Genet 13: 391-393

Provan J, Shepherd L, Powell W, Waugh R (1996) Analysis of intra-specific somatic hybrid of potato (*Solanum tuberosum*) using simple sequence repeats. Plant Cell Rep 16: 196-199

Provan J, Corbett G, McNicol JW, Powell W (1997) Chloroplast DNA variability in wild and cultivated rice (*Oryza* ssp.) revealed by polymorphic chloroplast simple sequence repeats. Genome 40: 104-110

Qian HR, Zhuang JY, Lin HX, Lu J, Zheng KL (1995) Identification of a set of RFLP probes for subspecies differentiation in *Oryza sativa* L. Theor Appl Genet 90: 878-884

Rafalski JA, Tingey SV (1993) Genetic diagnostics in plant breeding: RAPDs, microsatellites, and machines. Trends Genet 9: 275-280

Rajyashri KR, Nair S, Ohmido N, Fukui K, Kurata N, Sasaki T, Mohan M (1998) Isolation and FISH mapping of Yeast Artificial Chromosomes (YACs) encompassing an allele of the *Gm*2 gene for gall midge resistance in rice. Theor Appl Genet 97: 507-514

Ramakrishna W, Lagu MD, Gupta VS, Ranjekar PK (1994) DNA fingerprinting in rice using oligonucleotide probes specific for simple repetitive DNA sequences. Theor Appl Genet 88: 402-406

Ramakrishna W, Chowdari KV, Lagu MD, Gupta VS, Ranjekar PK (1995) DNA fingerprinting to detect genetic variation in rice using hypervariable DNA sequences. Theor Appl Genet 90: 1000-1006

Reddy AS, Kiefer-Meyer MC, Delsney M (1993) Characterization of new variants of a satellite DNA from *Oryza officinalis*, specific for the CC genome of wild rice. Genome 36: 750-761

Reddy OUK, Siddiq EA, Sarma NP, Ali J, Hussain AJ, Nimmakayala P, Ramasamy P, Pammi S, Reddy AS (2000) Genetic analysis of temperature-sensitive male sterility in rice. Theor Appl Genet 100: 794-801

Redona ED, Mackill DJ (1996a) Mapping quantitative trait loci for seedling vigor in rice using RFLPs. Theor Appl Genet 92: 395-402

Redona ED, Mackill DJ (1996b) Molecular mapping of quantitative trait loci in *japonica* rice. Genome 39: 395-403

Rico C, Rico I, Hewitt G (1996) 470 million years of conservation of microsatellite loci among fish species. Proc R Soc Lond B Biol Sci 263: 549-557

Roder MS, Plaschke J, Konig SU, Borner A, Sorrells ME, Tanksley SD, Ganal MW (1995) Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 246: 327-333

Rogers SO, Bendich AJ (1988) Extraction of DNA from plant tissues. In: Plant Molecular Biology Manual A6.1, Gelvin SB, Schilperoort RA (eds), Kluwer Academic Publishers, Dordrecht

Rohlf FJ (1993) NTSYS-pc Numerical Taxonomy and Multivariate Analysis System Version 1.80. State University of New York, Stony Brook, NY, (Owner's manual) Exeter Software, Setauket, New York

Ronald PC (1997) The molecular basis of disease resistance in rice. Plant Mol Biol 35: 179-186

Ronald PC, Albano B, Tabien R, Abenes L, Wu K, McCouch SR, Tanksley SD (1992) Genetic and physical analysis of the rice bacterial blight disease resistance locus, *Xa*-21. Mol Gen Genet 236: 113-120

Rost TL (1997) Rice anatomy - Introduction. From course material of Section of Plant Biology, Division of Biological Sciences, University of California, Davis [Online], Available http://www-plb.ucdavis.edu/courses/plb105/students/rice/introduction/intro.html

Rubinstein DC, Amos W, Leggo J, Goodburn S, Jain S, Li SH, Margolis RL, Ross CA, Ferguson-Smith MA (1995) Microsatellite evolution - evidence for directionality and variation in rate between species. Nature Genet 10: 337-343

Rubinsztein DC, Amos B, Cooper G (1999) Microsatellite and trinucleotide-repeat evolution: evidence for mutational bias and different rates of evolution in different lineages. Philos Trans R Soc Lond B Biol Sci 354: 1095-1099

Rychlik W, Rhoads RE (1989) A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. Nucl Acids Res 17: 8543-8551

Saghai Maroof MA, Yang GP, Biyashev RM, Maughan PJ, Zhang Q (1996) Analysis of the barley and rice genomes by comparative RFLP linkage mapping. Theor Appl Genet 92: 541-551

Saghai Maroof MA, Yang GP, Zhang Q, Gravois KA (1997) Correlation between molecular marker distance and hybrid performance in US southern long grain rice. Crop Sci 37: 145-150

Saito A, Yano M, Kishimoto N, Nakagahra M, Yoshimura A, Saito K, Kuhara S, Ukai Y, Kawase M, Nagamine T, Yoshimura S, Ideta O, Ohsawa R, Hayano Y, Iwata N, Sugiura M (1991) Linkage map of restriction fragment length polymorphism loci in rice. Jpn J Breed 41: 665-670

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

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring harbor, NY, 2nd ed

Sanchez AC, Llag LL, Yang D, Brar DS, Ausubel F, Khush GS, Yano M, Sasaki T, Li Z, Huang N (1999) Genetic and physical mapping of *xa*-13, a recessive bacterial blight resistance gene in rice. Theor Appl Genet 98: 1022-1028

Sanchez AC, Yang D, Khush GS, Zhu Y, Huang N (1997) Construction of overlapping BAC clones for *xa*-5 region of chromosome 5. Rice Genet Newsl 14: 118-119

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

Sasaki T, Burr B (2000) International rice genome sequencing project: the effort to completely sequence the rice genome. Curr Opin Plant Biol 3: 138-141

Sastry JG, Ramakrishna W, Sivaramakrishnan S, Thakur RP, Gupta VS, Ranjekar PK (1995) DNA fingerprinting detects genetic variability in the pearl millet downy mildew pathogen *(Sclerospora gramicola*). Theor Appl Genet 91: 856-861

Sato YI, Tang LH, Nakamura I (1995) Amplification of DNA fragments from charred rice grains by polymerase chain reaction. Rice Genet Newsl 12: 260-261

Schaffer HE, Sederoff RR (1981) Improved estimation of DNA fragment lengths from agarose gels. Anal Biochem 115: 113-122

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

Schlotterer C, Amos B, Tautz D (1991) Conservation of polymorphic simple sequence loci in cetacean species. Nature 354: 63-65

Schlotterer C, Vogel C, Tautz D (1997) Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural *Drosophila melanogaster* populations. Genetics 146: 309-320

Schut JW, Qi X, Stam P (1997) Association between relationship measures based on AFLP markers, pedigree data and morphological traits in barley. Theor Appl Genet 95: 1161-1168

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

Second G (1985a) Evolutionary relationships in the *Sativa* group of *Oryza* based on isozyme data. Genet Sel Evol 17: 89-114

Second G (1985b) A new insight into the genome differentiation in *Oryza* L. through isozyme studies. In: Advances in chromosome and cell genetics, Sharma AR, Sharma A (eds), Oxford and IBH Pub, New Delhi Bombay Calcutta. pp 75

Second G (1991) Molecular markers in rice systematics and evaluation of genetic resources. In: Biotechnology in Agriculture and Forestry, Rice, Springer Verlag, Berlin, Heidelberger. vol 14 pp168-194

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

Sharp PI, Kries M, Shewry PR, Gale MD (1988) Location of β-amylase sequences in wheat and it relatives. Theor Appl Genet 75: 286-290

Shimamoto K (1995) The molecular biology of rice. Science 270: 1772-1773

Shizuya H, Birren B, Kim UJ, Macino V, Slepak T, Tachiiri Y, Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc Natl Acad Sci, USA 85: 8794-8797

Simpson PR (1990) Variation among dispersed $(GATA)_n$ sequences in *Drosophila melanogaster*. Genome 33: 750-754

Singh L, Phillips C, Jones KW (1984) The conserved nucleotide sequences of *Bkm*, which defines *Sxr* in the mouse, are transcribed. Cell 36: 111-120

Singh K, Ishii T, Parco A, Huang N, Brar DS, Khush GS (1996) Centromere mapping and orientation of the molecular linkage map of rice. Proc Natl Acad Sci, USA; 93: 6163-6168

Skorupska HT, Choi IS, Rao-Arelli AP, Bridges WC (1994) Resistance to soybean cyst nematode and molecular polymorphism in various sources of Peking soybean. Euphytica 75: 63-70

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139: 457-462

Sliver LM (1992) Bouncing off microsatellites. Nat Genet 2: 8-9

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel tests of matrix correspondence. Syst Zool 35: 627-632

Smulders MJM, Bredemeijer G, Ruskortekaas W, Arens P, Vosman B (1997) Use of short microsatellites from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. Theor Appl Genet 84: 264-272

Sneath PHA, Sokal RR (1973) Numerical Taxonomy, Freeman, San Francisco. pp 573

Sokal RR, Sneath PHA (1963) Principles of Numerical Taxonomy, Freeman, San Francisco. pp 359

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald P (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa*21. Science 270: 1804-1806

Soranzo N, Provan J, Powell W (1999) An example of microsatellite length variation in the mitochondrial genome of conifers. Genome 42: 158-161

Souza E, Sorrells ME (1989) Pedigree analysis of North American oat cultivars released from 1951 to 1985. Crop Sci 29: 595-601

St Martin SK (1982) Effective population size for the soybean improvement program in maturity groups 00 to IV. Crop Sci 22: 151-152

Stallings RL (1995) Conservation and evolution of $(CT)_n$ / $(GA)_n$ microsatellite sequences at orthologous positions in diverse mammalian genomes. Genomics 25: 107-113

Staub JE, Serquen FC, Gupta M (1996) Genetic markers, map construction and their application in plant breeding. Hort Sci 31: 729-741

Strand M, Prolla TA, Liskay RM, Petes TD (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365: 274-276

Stuber CW (1995) Mapping and manipulating quantitative traits in maize. Trends Genet 11: 477-481

Sun GL, Diaz O, Salomon B, von Bothmer R (1999) Genetic diversity in *Elymus caninus* as revealed by isozyme, RAPD and microsatellite markers. Genome 42: 420-431

Takeuchi Y, Abe T, Sasahara T (2000) RFLP mapping of QTLs influencing shoot regeneration from mature seed-derived calli in rice. Crop Sci 40: 245-247

Tan YF, Li JX, Yu SB, Xing YZ, Xu CG, Zhang Q (1999) The three important traits for cooking and eating quality of rice grains are controlled by a single locus in an elite rice hybrid, Shanyou 63. Theor Appl Genet 99: 642-648

Tan YF, Xing YZ, Li JX, Yu SB, Xu CG, Zhang Q (2000) Genetic bases of appearance of quality of rice grains in Shanyou 63, an elite rice hybrid. Theor Appl Genet 101: 823-829

Tang D, Wu W, Li W, Lu H, Worland AJ (2000) Mapping of QTLs conferring resistance to bacterial leaf streak in rice. Theor Appl Genet 101: 286-291

Tanksley SD (1993) Mapping polygenes. Annu Rev Genet 27: 205-233

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92: 191-203

Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: New tools for an old science. Bio/technol 7: 257-264

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

Tautz D (1989) Hypervariability of simple sequence as a general source for polymorphic DNA markers. Nucl Acids Res 17: 6463-6471

Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucl Acids Res 12: 4127-4138

Tautz D, Schlotterer C (1994) Simple sequences. Curr Opin Gen Dev 4: 832-837

Tautz D, Trick M, Dover GA (1986) Cryptic simplicity in DNA is a major source of genetic variation. Nature 322: 652-656

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100: 697-712

Thanh ND, Zheng HG, Dong NV, Trinh LN, Ali ML, Nguyen HT (1999) Genetic variation in root morphology and microsatellite DNA loci in upland rice (*Oryza sativa* L.) from Vietnam. Euphytica 105: 43-51

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408: 796 – 815

Thomas MR, Scott NS (1993) Microsatellite repeats in grapevine reveal DNA polymorphisms when analyzed as sequence-tagged sites (STSs). Theor Appl Genet 86: 985-990

Thomas MR, Cain P, Scott NS (1994) DNA typing of grapevines: a universal methodology and database for describing cultivars and evaluating relatedness. Plant Mol Biol 25: 939-949

Thompson JD, Higgins GD, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl Acids Res 22: 4673-4680

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

Tripathy JN, Zhang J, Robin S, Nguyen TT, Nguyen HT (2000) QTLs for cell-membrane stability mapped in rice (*Oryza sativa* L.) under drought stress. Theor Appl Genet 100: 1197-1202

Tsunematsu H, Yoshimura A, Harushima Y, Nagamura Y, Kurata N, Yano M, Sasaki T, Iwata N (1996) RFLP framework map using recombinant inbred lines in rice. Breed Sci 46: 279-284

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

Umehara Y, Inagaki A, Tanoue H, Yasukochi Y, Nagamura Y, Saji S, Otsuki Y, Fujimura T, Kurata N, Minobe Y (1995) Construction and characterization of a rice YAC library for physical mapping. Mol Breed 1: 79-89

USDA, ARS, National Genetic Resources Program (2001) Germplasm Resources Information Network - (GRIN). [Online Database] National Germplasm Resources Laboratory, Beltsville, Maryland. Available: http://www.ars-grin.gov/cgi-bin/npgs/html/splist.pl?8617 (04 January 2001)

Valdes AM, Slatkin M, Freimer NB (1993) Allelic frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics 133: 737-749

Van Deynze AE, Dubcovsky J, Gill KS, Nelson JC, Sorrells ME, Dvorak J, Gill BS, Lagudah ES, McCouch SR, Appels R (1995) Molecular-genetic maps for group 1 chromosomes of Triticeae species and their relation to chromosomes in rice and oat. Genome 38: 45-59

van Treuren R, Kuittinen H, Karkkainen K, Baena-Gonzalez K, Savolainen O (1997) Evolution of microsatellites in *Arabis petraea* and *Arabis lyrata*, outcrossing relatives of *Arabidopsis thaliana*. Mol Biol Evol 14: 220-229

Vaughan DA (1989) The genus *Oryza* L. Current status of taxonomy. IRRI res paper ser no 138

Viramani SS (1994) Hybrid rice technology: new developments and future prospects, IRRI publication. pp 157-171

Virk PS, Ford-Lloyd BV, Jackson MT, Newbury HJ (1995a) Use of RAPD for the study of diversity within plant germplasm collections. Heredity 74: 170-179

Virk PS, Newbury HJ, Jackson MT, Ford-Lloyd BV (1995b) The identification of duplicate accessions within a rice germplasm collection using RAPD analysis. Theor Appl Genet 90: 1049-1055

Virk PS, Newbury HJ, Jackson MT, Ford-Lloyd BV (2000) Are mapped DNA markers more useful for assessing genetic diversity? Theor Appl Genet 100: 607-613

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Fritjers A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new concept for DNA fingerprinting. Nucl Acids Res 23: 4407-4414

Vosman B, Arens P (1997) Molecular characterization of GATA / GACA microsatellite repeats in tomato. Genome 40: 25-33

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

Wang XK, Cheng KS, Lu YX, Luo J, Huang NW, Liu GR (1984) Studies on indigenous rice in Yunnan and their utilization, III. Glabrous rice in Yunnan. Acta Univ Agric Pekinensis 10: 333-344

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

Wang GL, Wing RA, Paterson AH (1993) PCR amplification from single seeds, facilitating DNA marker-assisted breeding. Nucl Acids Res 21: 2527

Wang G, Castiglione S, Zhang J, Fu R, Ma J, Li W, Sun Y, Sala F (1994a) Hybrid rice (*Oryza sativa* L.): identification and parentage determination by RAPD fingerprinting. Plant Cell Rep 14: 112-115

Wang GL, Mackill DJ, Bonman M, McCouch SR (1994b) RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. Genetics 136: 1421-1434

Wang Z, Weber JL, Zhong G, Tanksley SD (1994c) Survey of plant short tandem DNA repeats. Theor Appl Genet 88: 1-6

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

Wang GL, Holsten TE, Song WY, Wang HP, Ronald PC (1995b) Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa*-21 disease resistance locus. Plant J 7: 525-533

Wang ZY, Zheng FQ, Shen GZ, Gao JP, Snustad DP, Li MG, Zhang JL, Hong MM (1995c) The amylose content in rice endosperm is related to the post-transcriptional regulation of the waxy gene. Plant J 7: 613-622

Wang GL, Song WY, Ruan DL, Sideris S, Ronald PC (1996) The cloned gene *Xa*-21 confers resistance to multiple *Xanthomonas oryzae* pv *oryzae* isolates in transgenic plants. Mol Plant Microbe Interact 9: 850-855

Wang ZX, Yano M, Yamanouchi U, Iwamoto M, Monna L, Hayasaka H, Katayose Y, Sasaki T (1999) The *Pib* gene for blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. Plant J 19: 55-64

Wang S, Wang J, Jiang J, Zhang Q (2000) Mapping of centromeric regions on the molecular linkage map of rice (*Oryza sativa* L.) using centromere-associated sequences. Mol Gen Genet 263: 165-172

Watt G (1891) Oryza. In: Dictionary of the economic products of India, New Delhi. 5: 498-654

Weber JL (1990a) Human DNA polymorphisms and methods of analysis. Curr Opin Biotechnol 1: 166-171

Weber JL (1990b) Informativeness of human $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ polymorphisms. Genomics 7: 524-530

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44: 388-396

Weber J, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2: 1123-1128

Weising K, Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genome of dicotyledonous angiosperms. Genome 42: 9-19

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucl Acids Res 18: 7213-7218

Wetton JH, Carter RE, Parkin DT, Walters DC (1987) Demographic study of a wild house sparrow population by DNA fingerprinting. Nature 327: 147-149

Whitton J, Rieseberg LH, Ungerer MC (1997) Microsatellite loci are not conserved across the Asteraceae. Mol Biol Evol 14: 204-209

Wicking E, Williamson B (1991) From linked marker to gene. Trends Genet 7: 288-293

Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics 146: 769-779

Williams JGK, Kubelik AR, Livak KL, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl Acids Res 18: 6531-6535

Williams MNV, Pande N, Nair S, Mohan M, Bennett J (1991) Restriction length polymorphism analysis of polymerase chain reaction products amplified from mapped loci of rice (*Oryza sativa* L.) genomic DNA. Theor Appl Genet 82: 489-498

Williams CE, Wang B, Holsten TE, Scambray J, de Assis Goes da Silva F, Ronald PC (1996) Markers for selection of the rice *Xa*-21 disease resistance gene. Theor Appl Genet 93: 1119-1122

Winberg BC, Zhou Z, Dallas JF, McIntyre CL, Gustafson JP (1993) Characterization of minisatellite sequences from *Oryza sativa*. Genome 36: 978-983

Wolff RK, Plaeke KR, Jeffreys AJ, White R (1991) Unequal crossing over between homologous chromosomes is not the major mechanism involved in generation of new alleles at VNTR loci. Genomics 5: 382-384

Wu KS, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 24: 225-235

Wu J, Matsui E, Yamamoto K, Nagamura Y, Kurata N, Sasaki T, Minobe Y (1995) Genomic organization of 57 ribosomal protein genes in rice (*Oryza sativa* L.) through RFLP mapping. Genome 38: 1189-1200

Wu P, Liao CY, Hu B, Yi KK, Jin WZ, Ni JJ, He C (2000) QTLs and epistasis for aluminium tolerance in rice (*Oryza sativa* L.) at different seedling stages. Theor Appl Genet 100: 1295-1303

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

Xiao J, Li J, Grandillo S, Ahn SN, McCouch SR, Tanksley SD, Yuan L (1996a) A wild species contains genes that may significantly increase the yield of rice. Nature 384: 223-224

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

Xiao J, Li J, Grandillo S, Ahn SN, Yuan L, Tanksley SD, McCouch SR (1998) Identification of trait improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics: 150: 899-909

Xie Y, Wu R (1989) Rice alcohol dehydrogenase genes: anaerobic induction, organ specific expression and characterization of cDNA clones. Plant Mol Biol 13: 53-68

Xiong L, Liu KD, Dai XK, Wang SW, Xu CG, Zhang DP, Maroof MAS, Sasaki T, Zhang Q (1997) A high density RFLP map based on the F2 population of a cross between *Oryza sativa* and *O.rufipogon* using Cornell and RGP markers. Rice Genet Newsl 14: 110-116

Xu K, Mackill DJ (1996) A major locus for submergence tolerance mapped on rice chromosome 9. Mol Breed 2: 219-224

Xu J, Constantino SV, Magpantay G, Bennett J, Sarkarung S, Huang N (1998) Classification of rice germplasm. II. Discrimination of indica from japonica via analysis of amplicon length polymorphisms. Plant Cell Rep 17: 640-645

Xu X, Peng M, Fang Z (2000) The direction of microsatellite mutations is dependent upon allele length. Nat Genet 24: 396-399

Yamagishi M, Yano M, Fukuta Y, Fukui K, Otani M, Shimada T (1996) Distorted segregation of RFLP markers in regenerated plants derived from anther culture of an F1 hybrid of rice. Genes Genet Syst 71: 37-41

Yanagihara S, McCouch SR, Ishikawa K, Ogi Y, Maruyama K, Ikehashi H (1995) Molecular analysis of the inheritance of the *S*-5 locus conferring wide compatibility in indica / japonica hybrids of rice (*Oryza sativa* L.). Theor Appl Genet 90: 182-188

Yang GP, Saghai Maroof MA, Xu CG, Zhang Q, Biyashev RM (1994) Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol Gen Genet 245: 187-194

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18, and pUC19 vectors. Gene 33: 103-119

Yap IV, Nelson RJ (1996) Winboot: A program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA-based dendrograms. IRRI discussion paper series no 14, Manila, Philippines

Yi QM, Deng WG, Xia ZP, Pang HH (1995) Polymorphism and genetic relatedness among wild and cultivated rice species determined by AP-PCR analysis. Hereditas 122: 135-141

Yoshimura S, Yoshimura A, Saito A, Kinoshimoto N, Kawase M, Yano M, Nakagahra M, Ogawa T, Iwata N (1992) RFLP analysis of introgressed chromosomal segments in three near-isogenic lines of rice for bacterial blight resistance genes *Xa*-1, *Xa*-3 and *Xa*-4. Jpn J Genet 67: 29-37

Yoshimura S, Yoshimura A, Iwata N, McCouch SR, Abenes ML, Baraoidan MR, Mew TW, Nelson RJ (1995a) Tagging and combining bacterial blight resistance genes in rice using RAPD and RFLP markers. Mol Breed 1: 375-387

Yoshimura S, Yoshimura A, Nelson RJ, Mew TW, Iwata N (1995b) Tagging *Xa*-1, the bacterial blight resistance gene in rice, using RAPD markers. Breed Sci 45: 81-85

Yoshimura S, Umehara Y, Kurata N, Nagamura Y, Sasaki T, Minobe Y, Iwata N (1996) Identification of a YAC clone carrying the *Xa*-1 allele, a bacterial blight resistance gene in rice. Theor Appl Genet 93: 117-122

Yoshimura S, Yamanouchi U, Katayose Y, Toki S, Wang ZX, Kono I, Kurata N, Yano M, Iwata N, Sasaki T (1998) Expression of *Xa*1, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. Proc Natl Acad Sci, USA 95: 1663-1668

Young ND (1990) Potential applications of map-based cloning to plant pathology. Physiol Mol Plant Pathol 37: 81-94

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

Yu ZH, Mackill DJ, Bonman JM, Tanksley SD (1991) Tagging genes for blast resistance in rice via linkage to RFLP markers. Theor Appl Genet 81: 471-476

Yu YG, Saghai Maroof MA, Buss GR, Tolin SA (1994) RFLP and microsatellite mapping of a gene for soybean mosaic virus resistance. Phytopathol 84: 60-64

Yuan LP (1985) Exploration of super high yielding hybrid rice. Hybrid Rice 3: 1-3

Yuan LP, Cheng HX (1986) Hybrid rice breeding and cultivation. Hunan Science and Technology Press, Hunan, China

Zhai W, Lu C, Zhu L, Yang W, Zhang Q (1996) PCR analysis of half-seeds of cereal crops and its application in marker-assisted selection and breeding. Chin J Biotechnol 12: 249-255

Zhang HB, Wing RA (1997) Physical mapping of the rice genome with BACs. Plant Mol Biol 35: 115-127

Zhang Q, Saghai Maroof MA, Lu TY, Shen BZ (1992a) Genetic diversity and differentiation of indica and japonica rice detected by RFLP analysis. Theor Appl Genet 83: 495-499

Zhang K, Shen P, Qian H, Wang J (1992b) Tagging genes for wide compatibility in rice via linkage to RFLP markers. Chinese J Rice Sci 6: 145-150

Zhang QF, Gao YJ, Yang SH, Ragab RA, Saghai Maroof MA, Li ZB (1994a) A diallel analysis of heterosis in elite hybrid rice based on RFLPs and microsatellites. Theor Appl Genet 89: 185-192

Zhang Q, Shen BZ, Dai XK, Mei MH, Saghai Maroof MA, Li ZB (1994b) Using bulked extremes and recessive class to map genes for photoperiod-sensitive genic male sterility in rice. Proc Natl Acad Sci, USA 91: 8675-8679

Zhang Q, Gao YJ, Saghai Maroof MA, Yang SH, Li JX (1995) Molecular divergence and hybrid performance in rice. Mol Breed 1: 133-142

Zhang G, Angeles ER, Abenes MLP, Khush GS, Huang N (1996a) RAPD and RFLP mapping of the bacterial blight resistance gene *xa*-13 in rice. Theor Appl Genet 93: 65-70

Zhang HB, Choi S, Woo SS, Li Z, Wing RA (1996b) Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. Mol Breed 2: 11-24

Zhang Q, Zhou ZQ, Yang GP, Xu CG, Lin KD, Saghai Maroof MA (1996c) Molecular marker heterozygosity and hybrid performance in indica and japonica rice. Theor Appl Genet 93: 1218-1224

Zhao X, Kochert G (1992) Characterization and genetic mapping of a short, highly repeated, interspersed DNA sequence from rice (*Oryza sativa* L.). Mol Gen Genet 231: 353-359

Zhao X, Kochert G (1993a) Clusters of interspersed repeated DNA sequences in the rice genome (*Oryza*). Genome 36: 944-953

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

Zhao X, Wu T, Xie Y, Wu R (1989) Genome-specific repetitive sequences in the genus *Oryza*. Theor Appl Genet 78: 201-209

Zheng K, Subudhi PK, Domingo J, Maopanty G, Huang N (1995) Rapid DNA isolation for marker assisted selection in rice breeding. Rice Genet Newsl 12: 255-258

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

Zhou Z, Bebeli PJ, Somers DJ, Gustafson JP (1997) Direct amplification of minisatellite-region DNA with VNTR core sequences. Theor Appl Genet 95: 942-949

Zhu LH, Chen Y, Xu YB, Xu JC, Cai HW, Ling ZZ (1993) Construction of a molecular map of rice and gene mapping using a double haploid population of a cross between indica and japonica varieties. Rice Genet Newsl 10: 132-135

Zhu J, Gale MD, Quarrie S, Jackson MT, Bryan GJ (1998) AFLP markers for the study of rice biodiversity. Theor Appl Genet 96: 602-611

Zhu JH, Stephenson P, Laurie DA, Li W, Tang D, Gale MD, Jackson MT (1999) Towards rice genome scanning by map-based AFLP fingerprinting. Mol Gen Genet: 184-195

Zhuang CX, Zhang GQ, Mei MT, Lu YG (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 repeats (SSR)-anchored PCR amplification. Genomics 20: 176-183

Zou JH, Pan XB, Chen ZX, Xu JY, Lu JF, Zhai WX, Zhu LH (2000) Mapping quantitative trait loci controlling sheath blight resistance in two rice cultivars (*Oryza sativa* L.). Theor Appl Genet 101: 569-573

CURRICULUM VITAE

Personal Information

Name: Armaity Piroze Davierwala

Date of birth: May 12, 1972

Address: 27, Sindh Hindu Housing Society,

593-C Lullanagar,

Pune 411040 (INDIA)

Tel: 91-20-6833307 / 6831549

Email: adavierwala@yahoo.com

Educational Qualifications

Degree	**Place of study**	**Year of Passing**	Grade / Class	**Marks secured (%)**
B.Sc. (Chemistry)	Fergusson College, Pune (University of Pune)	1992	Ist class with distinction	80.63
M.Sc. (Biochemistry)	Department of Chemistry, University of Pune	1994	Ist class with distinction	68.05

Awards and Scholarships

- Junior and Senior Research Fellowship from Council of Scientific and Industrial Research, New Delhi, India.
- Government National Merit Scholarship for post-graduation studies.
- Secured second rank in the University of Pune at the B.Sc. examination.
- Several scholarships and prizes during graduation.

List of Publications

(1) Ramakrishna W., **Davierwala A.P.**, Gupta V.S. and Ranjekar P.K. (1998) Expansion of $(GA)_n$ dinucleotide at a microsatellite locus associated with domestication in rice. Biochemical Genetics 36: 323-327.

(2) **Davierwala A.P.**, Ramakrishna W., Ranjekar P.K. and Gupta V.S. (2000) Sequence variations at a complex microsatellite locus in rice and its conservation in cereals. Theoretical and Applied Genetics 101: 1291-1298.

(3) **Davierwala A.P.**, Chowdari K.V., Shivkumar A., Reddy A.P.K., Ranjekar P.K. and Gupta V.S. (2001) Genetic diversity evaluation of Indian elite rice varieties using molecular markers. Genetica (In Press).

(4) **Davierwala A.P.**, Ramakrishna W., Chowdari K.V., Ranjekar P.K. and Gupta V. S. Potential of $(GATA)_n$ microsatellites from rice for inter- and intra-specific variability studies. Communicated to Theoretical and Applied Genetics.

(5) **Davierwala A.P.**, Reddy A.P.K., Lagu M.D., Ranjekar P.K. and Gupta V.S. Marker Assisted Selection Of Bacterial Blight Resistance Genes In Rice. Communicated to Biochemical Genetics.

Posters Presented at National and International Conferences / Symposia

(1) **Wadia (Davierwala) A.M.,** Chowdari K.V., Gupta V.S. and Ranjekar. P.K. Genetic fingerprinting of elite rice varieties using RAPDs. 5^{th} annual meeting of National Rice Biotechnology Network, New Delhi, November 13-16, 1996.

(2) Gupta V.S., Ranjekar P.K., Ramakrishna W., Chowdari K.V., Joshi S.P. and **Wadia (Davierwala) A.M.** Characterization of $(CAC)_n$ and $(GATA)_n$ microsatellites in rice. 4^{th} International DNA Fingerprinting Conference, Melbourne, Australia, December 2-7, 1996.

(3) **Davierwala A.P.**, Ramakrishna W., Gupta V.S. and Ranjekar P.K. Expansion of $(GA)_n$ dinucleotide at a microsatellite locus associated with domestication in rice. 7^{th} National Rice Biotechnology Network meeting, Bangalore, October 25-29, 1998.

(4) Gupta V.S., Joshi S.P., **Davierwala A.P.** and Ranjekar P.K. Rice Genomics: Exploitation of DNA markers in hybrid rice program, diagnosis of BLB resistant lines and genotype profiling of wild rice. A collaborative research program. General meeting of the Rockefeller Foundation International Program on Rice Biotechnology at Phuket, Thailand, September 20-24, 1999