

2003

Genetic mapping of agronomic traits from the interspecific cross of *Oryza sativa* (L.) and *Oryza glaberrima* (Steud.)

Gabriel Kayode Aluko

Louisiana State University and Agricultural and Mechanical College, galuko1@lsu.edu

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Recommended Citation

Aluko, Gabriel Kayode, "Genetic mapping of agronomic traits from the interspecific cross of *Oryza sativa* (L.) and *Oryza glaberrima* (Steud.)" (2003). *LSU Doctoral Dissertations*. 1077.
https://digitalcommons.lsu.edu/gradschool_dissertations/1077

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.

GENETIC MAPPING OF AGRONOMIC TRAITS
FROM THE INTERSPECIFIC CROSS
ORYZA SATIVA (L.) AND *ORYZA GLABERRIMA* (STEUD.)

A Dissertation

Submitted to the Graduate
Faculty of the Louisiana State University
and Agricultural and Mechanical College
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in

The Department of Agronomy

By

Gabriel Kayode Aluko
Bsc University of Ibadan, Nigeria 1978
Msc University of Ibadan, Nigeria 1982
M. Agric Studies, Queensland University
Brisbane Australia 1986

December 2003

This work is dedicated to the memory of my late father Chief Michael Owolabi Aluko who gave me a solid foundation in life.

ACKNOWLEDGEMENTS

I wish to express my profound gratitude to Dr. James Oard, my major professor, for his guidance and useful suggestions throughout the period of this research. I am grateful for his constructive criticisms and the correction of this dissertation.

I wish to also thank the members of my committee, Drs. Gerald Myers, Don Labonte, Charlie Johnson and Paul Bell for their contributions to the overall success of my studies and especially the research. I am grateful to the Rockefeller Foundation who provided the fellowship for this degree and funding for the research. I appreciate the contributions from our collaborators Drs. Christine Bergman and Fernando Goffman, Mrs. Naomi Gipson and Janis Delgado of USDA Beaumont, Drs. Cesar Martinez, Joe Tohme and Miss. Carolina Castano from CIAT Colombia, Dr. Monty Jones and the West Africa Rice Development Association (WARDA) for facilitating the fellowship award.

Grateful acknowledgement is also expressed to the head and staff of the Department of Agronomy for making it possible for me to do my research and studies here.

Mrs. Mary Bowen of the Department of Forestry did a lot to help me with microsatellite assays in her lab. I am grateful for her assistance and making materials available to me at all times. I thank my colleagues, Alicia Ryan, Fabian Capdevielle and others in the rice breeding lab for their assistance.

I would like to thank my dear wife Florence and our children, Barnabas Ayokunle, Blessing Odunayo, Esther Busayo and Grace Molayo for a peaceful and loving atmosphere in the home and for checking through the list of references.

Above all, I give all the glory to God the giver of life and knowledge who sustained me throughout the duration of this study. May the fruit of this endeavor be used to the glory of God.

TABLE OF CONTENTS

DEDICATION.....	ii
ACKNOWLEDGEMENTS.....	iii
LIST OF TABLES	v
LIST OF FIGURES	viii
ABSTRACT	x
CHAPTER 1: INTRODUCTION.....	1
CHAPTER 2: QTL MAPPING OF GRAIN QUALITY TRAITS FROM THE INTERSPECIFIC CROSS <i>ORYZA SATIVA</i> (L.) X <i>ORYZA</i> <i>GLABERRIMA</i> (STEUD.).....	9
2.1 Introduction	9
2.2 Materials and Methods	15
2.3 Results	20
2.4 Discussion	38
2.5 References	44
CHAPTER 3: GENETIC MAPPING OF AGRONOMIC TRAITS FROM THE INTERSPECIFIC CROSS <i>ORYZA SATIVA</i> (L.) AND <i>ORYZA GLABERRIMA</i> (STEUD.) USING TRADITIONAL AND NON-PARAMETRIC METHODS	53
3.1 Introduction	53
3.2 Materials and Methods	64
3.3 Results	70
3.4 Discussion	109
3.5 References	117
CHAPTER 4: SUMMARY AND CONCLUSIONS	124
VITA	127

LIST OF TABLES

2.1	Mean values for 10 grain quality traits of parents and 312 DH lines derived from <i>O. sativa</i> (Caiapo) x <i>O. glaberrima</i> (IRGC103544) cross, Colombia, 2001.....	25
2.2	Correlation coefficients among 10 grain quality characteristics of 312 DH lines derived from <i>O. sativa</i> (Caiapo) x <i>O. glaberrima</i> (IRGC103544) cross, Colombia, 2001.....	25
2.3	Chi square values and chromosome location of microsatellite markers showing segregation distortion among 312 doubled haploid lines derived from the cross <i>O. sativa</i> (Caiapo) x <i>O. glaberrima</i> (IRGC 103544).....	27
2.4	Quantitative Trait Loci for Grain Quality Traits among 312 doubled haploid lines derived from the cross <i>O. sativa</i> (Caiapo) x <i>O. glaberrima</i> (IRGC 103544)	33
2.5	List of significant two-way interaction between different loci covering the genome.....	36
3.1	Mean values for 6 agronomic traits of parents and 312 DH lines derived from <i>O. sativa</i> (Caiapo) x <i>O. glaberrima</i> (IRGC103544) cross, Colombia, 2001	72
3.2	Mean values for 6 agronomic traits of parents and 158 DH lines derived from <i>O. sativa</i> (Caiapo) x <i>O. glaberrima</i> (IRGC103544) cross, Crowley, LA, 2002	72
3.3	Correlation coefficients among grain quality characteristics of 312 DH lines derived from <i>O. sativa</i> x <i>O. glaberrima</i> cross at Colombia, 2001 and Crowley, 2002.....	74
3.4	Quantitative Trait Loci for agronomic traits in 312 DH lines derived from <i>O. sativa</i> (Caiapo) x <i>O. glaberrima</i> (IRGC103544) cross; Colombia, 2001 and Crowley, LA, 2002.....	76
3.5	Percent correct classification for 312 DH lines, Colombia, 2001 assigned to groups of tall and short plants using a <i>k</i> -nearest neighbor (<i>k</i> -NN) algorithm) with and without consideration of population structure, Colombia, 2001.....	82

3.6	Percent correct classification for 312 DH lines, Colombia, 2001 into early and late heading date groups using a <i>k</i> -nearest neighbor (<i>k</i> -NN) algorithm with and without consideration of population structure, Colombia, 2001.....	83
3.7	Percent correct classification for 312 DH lines into low and high tillers per plant groups using a <i>k</i> -nearest neighbor (<i>k</i> -NN) algorithm with and without consideration of population structure, Colombia, 2001.....	84
3.8	Percent correct classification for 312 DH lines into short and long panicle length groups using a <i>k</i> -nearest neighbor (<i>k</i> -NN) algorithm with and without consideration of population structure, Colombia, 2001.....	85
3.9	Percent correct classification for 312 DH lines into low and high grain yield groups using a <i>k</i> -nearest neighbor (<i>k</i> -NN) algorithm with and without consideration of population structure, Colombia, 2001.....	86
3.10	Percent correct classification for 312 DH lines, Colombia, 2001 into low and high 1000-grain weight groups using a <i>k</i> -nearest neighbor (<i>k</i> -NN) algorithm with and without consideration of population structure, Colombia, 2001.....	87
3.11	Percent correct classification for 158 DH lines into tall and short plant height groups using a <i>k</i> -nearest neighbor (<i>k</i> -NN) algorithm with and without consideration of population structure, Crowley, LA. 2002.....	89
3.12	Percent correct classification for 158 DH lines, Crowley, 2002 into early and late heading date groups using a <i>k</i> -nearest neighbor (<i>k</i> -NN) algorithm with and without consideration of population structure, Crowley, LA. 2002.....	90
3.13	Percent correct classification for 158 DH lines, Crowley, 2002 into low and high tillers per plant groups using a <i>k</i> -nearest neighbor (<i>k</i> -NN) algorithm with and without consideration of population structure, Crowley, LA. 2002.....	91
3.14	Percent correct classification for 158 DH lines into long and short panicles groups using a <i>k</i> -nearest neighbor (<i>k</i> -NN) algorithm with and without consideration of population structure, Crowley, LA. 2002.....	92
3.15	Percent correct classification for 158 DH lines, Crowley, 2002 into high	

	and low grain yield groups using a <i>k</i> -nearest neighbor (<i>k</i> -NN) algorithm with and without consideration of population structure, Crowley, LA. 2002.....	93
3.16	Percent correct classification for 158 DH lines, Crowley, 2002 into low and high 1000-grain weight groups using a <i>k</i> -nearest neighbor (<i>k</i> -NN) algorithm with and without consideration of population structure, Crowley, LA. 2002.....	94
3.17	Chromosomal location of SSR markers and % correct classification of DH lines using IM, BSA and DA procedures, Crowley 2002, subpopulation 1.....	96
3.18	Chromosomal location of SSR markers and % correct classification of DH lines using MR, IM, BSA and DA procedures, Crowley 2002, subpopulation 2.....	98
3.19	Chromosomal location of SSR markers and % correct classification of DH lines using MR, IM, BSA and DA procedures, Colombia 2001, subpopulation 1.....	100
3.20	Chromosomal location of SSR markers and % correct classification of DH lines using MR, IM, BSA and DA procedures, Colombia 2001, subpopulation 2.....	102
3.21	Chromosomal location of SSR markers and % correct classification of DH lines using MR, IM, BSA and DA procedures, Colombia 2001, subpopulation 3.....	104

LIST OF FIGURES

2.1	Distribution of milling and grain quality characteristics among 312 DH rice lines evaluated at CIAT, Colombia, 2001. C = mean value for Caiapo parent G = mean value for <i>O. glaberrima</i> parent IRGC accession 103544	22
2.2	Assignment of 28 QTLs for 10 grain and milling traits on the rice linkage map adjusted by MapDisto program among 312 DH lines derived from <i>O. sativa</i> (Caiapo) x <i>O. glaberrima</i> (IRGC 103544) cross. Confidence intervals for each QTL indicated as bar to the right of chromosome. QTLs in bold indicate positive allelic effects from IRGC 103544. H = QTL reported by He et al. (1999), T = QTL reported by Tan et al. (2001), L = QTL reported by Lanceras et al. (2000), Z = QTL reported by Zhou et al. (2003). QTL legend: <i>br</i> = percent brown rice; <i>hr</i> = percent head rice; <i>rb</i> = percent rice bran; <i>mr</i> = percent milled rice; <i>amy</i> = amylose content; <i>alk</i> = alkali spreading score; <i>pro</i> = percent protein content; <i>gl</i> = grain length; <i>gw</i> = grain width; <i>lwr</i> = length/width ratio. * = QTL detected in this study.....	34
2.3	The effect of interaction between the marker loci RM253 and RM148 on head rice percent and marker loci RM333 and RM81B on amylose content.....	36
3.1	Schematic diagram of the development of 312 doubled haploid lines derived from <i>O. sativa</i> (Caiapo) x <i>O. glaberrima</i> (IRGC103544) cross	66
3.2	Distribution of agronomic traits in 312 DH lines evaluated at CIAT, Colombia, 2001 and 158 DH lines at Crowley, LA 2002.....	73
3.3	Fig. 3.3. Assignment of QTLs for 6 agronomic traits on the rice linkage map adjusted by MapDisto program among 312 DH lines derived from <i>O. sativa</i> (Caiapo) x <i>O. glaberrima</i> (IRGC 103544) cross at Colombia and Crowley. Confidence interval for each QTL indicated as bar to the right of chromosome. QTLs in bold indicate positive allelic effects from IRGC 103544. ht = QTL for plant height, hd = QTL for heading date, pan = QTL for panicle length, till = QTL for tillers per plant, gy = QTL for grain yield, gw = QTL for 1000-grain weight. Zu = QTL reported by Zhuang et al. (1997), Mo = QTL reported by Moncada et al. (2001), Lu = QTL reported by Lu et al. (1997), Xi = QTL reported by Xiao et al. (1996), B = QTL reported by Brondani et al. (2002), Y = QTL reported by Yu et al. (2002), M = QTL reported by Mei et al. (2003), Ya = QTL reported by Yano et al. (19978) * = QTL detected in this study.....	80

3.4	Chromosomal locations of markers selected by MR, BSA, DA and Interval Analysis among 312 DH lines derived from <i>O. sativa</i> (Caiapo) x <i>O. glaberrima</i> (AC# IRGC103544), from Crowley. <i>ht</i> : Plant height ; <i>dh</i> : Days to heading; <i>till</i> : Tillers per plant; <i>pan</i> : Panicle length; <i>gy</i> : Grain yield: <i>gw</i> : 1000-grain weight	114
3.5	Chromosomal locations of markers selected by MR, BSA, DA and Interval Analysis among 312 DH lines derived from <i>O. sativa</i> (Caiapo) x <i>O. glaberrima</i> (AC# IRGC103544), from Colombia. <i>ht</i> : Plant height ; <i>dh</i> : Days to heading; <i>till</i> : Tillers per plant; <i>pan</i> : Panicle length; <i>gy</i> : Grain yield: <i>gw</i> : 1000-grain weight.....	115

ABSTRACT

Wild relatives of cultivated rice varieties offer new genetic sources for enhancing economic value, but traditional interval mapping techniques have not gained widespread support among applied researchers for marker assisted selection. The objectives of this study were to detect quantitative trait loci (QTL) for agronomic traits in a hybrid mapping population and compare the non-parametric Discriminant Analysis (DA) procedure with traditional approaches for accuracy and precision. In addition, the effects of population structure on marker-assisted classification were explored. A molecular linkage map comprising 100 SSR markers that spanned the rice genome at intervals of 10.5 cM on the average was constructed based on 312 doubled haploid lines derived from the cross interspecific *Oryza sativa* x *O. glaberrima*.

The mapping population was evaluated in replicated field plots in Colombia and Louisiana in 2001 and 2002, respectively. QTLs were identified for grain, milling and eating qualities and important agronomic traits such as heading date, plant height, number of tillers per plant, panicle length, grain yield and 1000-grain weight. A total of 28 QTLs were detected for 10 grain quality traits, and 22 QTLs for six agronomic traits were detected that were significant in at least one environment, but only seven were significant in both environments. SSR markers that best discriminated between pre-defined groups of high and low trait values were selected by stepwise DA. Using a k -nearest neighbor algorithm, the largest phenotypic differentiation (3 standard deviations) between two contrasting phenotypic groups resulted in 100% correct classification. Adjustments for population structure resulted in a 5-fold decrease in number of markers needed to achieve the same level of accuracy. These results demonstrated that procedures such as DA and consideration of population structure can be used for efficient marker-based allocation of the doubled haploid lines into pre-defined groups for yield and other agronomic traits. Finally, DA-selected markers pointed to the same or closely linked regions on the linkage map that in turn underscored the validity of the DA approach for genetic mapping.

CHAPTER 1: INTRODUCTION

Rice is the world's most important food crop, providing nutrition for more people than any other crop. Rice accounts for 23 percent of the world's supply of calories (Brar and Khush 2002) and it is unique among cereals by having a storage protein primarily made of glutelin. This protein has a more balanced amino acid profile than the prolamine-rich storage proteins found in most cereals (Juliano 1985). Rice is planted on about 150 million ha of land annually, (11 percent of the world's arable land); it is one of the most versatile crops cultivated worldwide as it is grown under a wide range of agro-climatic conditions ranging from irrigated, rain-fed lowland, rain-fed upland and flooding ecosystems.

Although the world's rice production has more than doubled from 257 million tons in 1966 to 600 million tons in 2000, the increase has not kept up with the demand for rice because of the corresponding increase in the human population during this time. It is estimated that rice production must increase by at least 40 percent during the next 25 years to meet ever-increasing demands. This calls for the development of rice varieties with higher yield potential, tolerance to biotic and abiotic stresses and superior grain quality. These goals can be achieved through plant breeding.

The genetic variability for some of the traits needed for high yield performance and stress tolerance is limited in cultivated germplasm. This is because a small core of adapted progenitors has been used repeatedly in rice breeding programs such that the genetic base of rice has become narrow (Moncada et al. 2001; Hargrove et al. 1980; Dilday 1990). Introgression of genes from other rice species can provide genetic variation to improve rice and meet the challenges affecting rice production. Major efforts have been made by rice breeding programs at institutions such as the West Africa Rice Development Association (WARDA), the International Rice Research Institute (IRRI), and the International Center for Tropical Agriculture (CIAT) to

introgress useful genes from various sources into elite breeding lines. For instance at WARDA, attempts have been made to introgress useful genes from the African rice *O. glaberrima* into *O. sativa* elite breeding lines (Jones et al. 1997). A total of 1130 accessions were evaluated for a range of morphological and agronomic traits over two years, 1991 and 1992. Wider genetic variation was found for some of the traits, especially growth duration, in *O. glaberrima* than in traditional or improved *O. sativa* varieties, thus revealing the valuable gene pools in the *O. glaberrima* germplasm that can be utilized for crop improvement. Clearly, the greatest challenge to plant breeders is to consolidate positive genetic elements into the most favorable combinations.

The reservoir of natural genetic variation found in a plant species is the primary tool for crop improvement. This genetic variation is not limited to cultivated species, but can also be found in wild relatives. The understanding, management and efficient utilization of the genetic variability found in both cultivated species and their wild relatives is the key to an efficient crop improvement program.

Crop improvement begins with the selection of parental lines having the desired traits that will meet the objective of the breeder. This is followed by making crosses between the parents to generate a segregating population. Selection of progenies showing the desired traits then commences and continues as the population is advanced from one generation to the next. For simply inherited traits controlled by single genes with major effects, the selection process can be made by traditional breeding methods. However, plant breeders are often confronted with problems when trying to improve a trait that is controlled by many genes through traditional breeding methods. In rice breeding, most agronomic and grain quality traits are controlled by many genes each of which has a relatively small effect on the overall phenotype. These traits do not show discrete phenotypes, consequently they are often measured and given a quantitative

value and are referred to as quantitative traits. Quantitative traits are difficult to study because the phenotypes do not give an insight into the genotype. The expression of genes controlling quantitative traits can be greatly influenced by the environment (Lynch and Walsh 1998). Consequently, the improvement of polygenic traits by traditional breeding methods is time consuming and the gains are harder to realize. Breeders usually overcome this problem by multi-environmental evaluation of replicated trials to capture the effect of the environment (Talbot 1997).

Because of the intractable problem encountered while trying to improve quantitative traits by conventional means, breeders and geneticists have considered the potential use of DNA markers to identify chromosomal regions harboring genes that influence the quantitative traits. Morphological markers were the first generation of markers to be used for identification and selection for quantitative traits loci. Very slow progress was made in the use of these markers due to the undesirable effects of many of the markers on the target phenotype. The effect of the marker genes on the quantitative traits was often larger than that of the linked quantitative trait loci (QTL), hence it was difficult to effectively and extensively use these markers to study quantitatively inherited traits (Tanksley et al. 1998). Moreover, there was a dearth of available segregating genetic markers.

Falconer and Mackay (1966) observed that if molecular markers were found that are linked to the genes responsible for quantitative traits, indirect selection for such markers might improve the efficiency of breeding for such traits. This is because molecular markers are unaffected by the environment and as such can be used to identify loci that affect such quantitative traits.

Various markers have been used to identify chromosomal regions that affect quantitative traits in crops. For instance, Stuber and Edwards (1986) first used isozymes as molecular

markers to identify QTL in maize. Isozymes have also been used for genetic analysis and to study linkage relationships in rice (Pham et al. 1990). The utility of isozymes was reduced by the numbers of such markers available. The advent of DNA markers such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) and microsatellites (also known as simple sequence repeats or SSRs) has greatly enabled the molecular dissection of quantitative traits. Molecular markers are polymorphic and their potential application includes marker assisted selection (Lee 1995), measurement of genetic diversity among lines (Dudley et al., 1991; Smith et al., 1991; Yu and Nguyen, 1994; Mackill, 1995), selection of parents for crossing (Dudley et al., 1992), fingerprinting of lines for legal purposes, control of genetic purity in the seed production process (Smith and Smith 1989), and establishment of genealogical relationships among modern elite rice varieties (Cao and Oard, 1997).

Using appropriate statistical models, molecular markers are useful for predicting quantitative variation within a set of germplasm. Wang et al. (1996) used multiple regression to identify strong associations between RAPD markers and a set of quantitative traits. In the multiple regression models, markers were used as predictor variables to study quantitative variation within a set of rice germplasm (Virk et al. 1996).

Several efforts have been made to introgress useful genes into elite rice varieties through interspecific hybridization with varying amount of success (Jones et al. 1997; Martinez et al. 1997; Brar and Khush 1997; Moncada et al. 2000). Recent breakthroughs in anther culture and molecular biology provide greater opportunities for rice breeders to develop a new generation of rice varieties that are better adapted and high yielding. It is now possible to select for molecular markers linked to traits of interest (marker assisted selection) rather than selecting for the traits themselves.

A concerted effort has been made on the improvement of agronomic characters in rice, but very little effort has been made to study the genetic basis of rice grain or milling quality (Tan et al. 2001). Most of the studies on grain quality have focused on eating, cooking and appearance qualities with an emphasis on amylose content, alkali spreading scores, gel consistency and chalkiness. (Lanceras et al. 2002; Tan et al. 1999; McKenzie and Rutger 1983). A study is therefore needed to investigate the genetic basis of milling quality in rice. It is well established that for general consumer acceptance, it is essential that improved lines of rice possess good grain, cooking and eating qualities in addition to high grain yield potential.

Chromosomal regions controlling various agronomic traits have been identified in rice. QTL have been revealed through their association with molecular markers, and by the year 2000, more than 1000 QTL have been documented in rice (Xu 2001). However, the use of these QTLs by breeding programs is still in its infancy. Only a few of the reported QTLs have been used for crop improvement (Yano et al. 2000, heading date; Zhou et al 2001, amylose content). Most of the QTLs reported in rice have not been used widely because there is no general agreement on the true location of QTLs, most have wide confidence intervals and QTLs have not been adequately tested over many environments to gain the confidence of breeders. To utilize available information from QTL and molecular marker data, a more comprehensive combination of useful markers to identify quantitative traits is urgently needed.

Researchers have recently considered the method of selective genotyping as a means of reducing the cost of genotyping for QTL analysis without losing much power of detection. This method implies the selection of the extremely high and low scoring individuals from the continuous distribution of a quantitative trait. This method is based on the premise that extreme individuals provide the most linkage information (Lander and Botstein, 1989). Van Gestel S et

al. (2000) concluded that the method of selective genotyping is a powerful method to detect associations for a quantitative trait.

During this dissertation research, chromosomal regions associated with milling, eating and cooking qualities of rice were detected in the same regions as other mapping studies, despite the occurrence of segregation distortion of the microsatellite markers. Moreover, new QTL positions for percent rice bran were detected for the first time. None of the earlier mapping studies has reported QTL for rice bran yet the bran serves as a reservoir for essential fatty acids needed for energy and cell repair. An elucidation of the genetic basis of inheritance of rice bran is necessary for formulating a breeding strategy for this trait.

In this study, a non-parametric model based on Discriminant analysis, (DA) was evaluated for its potential to identify candidate markers associated with agronomic traits in a doubled haploid population of 312 lines derived from a cross between IRGC103544 (an African rice species) and Caiapo, an elite line from Brazil. The effect of population structure on the accuracy of correct classification of lines was also evaluated. A genetic map was constructed from the same doubled haploid (DH) lines using traditional QTL mapping methods. Markers were detected that pointed to the same regions on the rice genetic map and new markers associated with all six agronomic traits were also detected. These results show that detection of markers through DA has a genetic basis. Based on the results of this study, the use of DA as a complement to traditional QTL analysis is recommended for the rapid identification of markers associated with agronomic and grain quality traits in rice breeding. Several studies have identified QTLs for many quantitative traits in rice. However, most breeding programs are not using this information for marker-assisted selection of superior genotypes. In view of the deluge of information on molecular and phenotypic characteristics available through many years of rice

research, DA could be a useful tool for the integration of marker information in the selection of breeding lines and also for germplasm improvement.

This dissertation is divided into three parts: The first part addresses the mapping efforts of QTL for grain quality traits in a population of 312 doubled haploid lines (DH) derived from a cross between *O. sativa* and *O. glaberrima*. The results of this research are reported in Chapter 2. In Chapter 3, mapping of agronomic traits in the same 312 DH population is reported. The same chapter also investigates the use of a non-parametric procedure known as discriminant analysis (DA) for marker assisted classification of DH lines into pre-defined groups of high and low trait values. The application of DA as a tool for allocating rice lines into groups was first proposed in 2000 (Balzarini et al. 2000) and its application to identify rice germplasm with contrasting phenotypes was reported by Capdevielle et al. (2000). These pioneering researchers gave the impetus for further research into the application of DA. The effect of population structure was also investigated to determine its effects on mapping of agronomic traits and classification of lines into predefined groups of high and low trait values. The phenotypic based DNA pooling methodology of Michelmore et al. (1991) was also used to simulate bulked segregant analysis (BSA) to refine the detection of these six agronomically important traits.

Chapter four gives an overview of the results of the various studies in this dissertation and the overall conclusion. The DA procedure once established would assist rice breeders in marker assisted selection since various results (Balzarini et al. 2000; Capdevielle et al. 2001) have suggested that this procedure has a genetic basis.

The first objective of this study was to detect chromosomal regions that control plant height (PLtHt), days to heading (HD), tillers per plant (Till/Plt), panicle length (Panlen), thousand grain weight (TGW) and grain yield (GY) using the DH mapping population described

previously. The second objective was to assess the use of DA as a tool for marker assisted classification in rice. Thirdly, because association of molecular markers with phenotypic traits may be biased by the presence of cryptic population structure, the effect of adjusting for population structure on the efficiency of DA classification of lines into pre-defined groups was investigated. The fourth objective was to assess multiple regression models as an alternative method for selecting markers associated with agronomic traits. Finally, BSA was compared with multiple regression and DA in the selection of markers for allocation of rice lines into pre-defined groups. In summary, this chapter specifically examined the potential of association genetics and population structure, two topics first developed in human research, to detect markers in rice that distinguish defined agronomic groups or lines with contrasting agronomic traits.

CHAPTER 2: QTL MAPPING OF GRAIN QUALITY TRAITS FROM THE INTERSPECIFIC CROSS *ORYZA SATIVA* (L.) X *ORYZA GLABERRIMA* (STEUD.)

2.1 Introduction

Rice constitutes a principal source of calories for Asia, Africa, and South America, and this trend will continue well into the 21st century (Pingali et al. 1997). World rice production has more than doubled in the last three decades from 257 million tons in 1966 to 560 million tons in 1997 as a result of the introduction of high yielding varieties and improved agricultural practices (FAO 2000). High grain yield is an important consideration for commercial and private producers, but demand for superior grain quality is increasingly a priority for international export markets in all rice producing areas of the world (Juliano et al. 1990; Unnevehr et al. 1992; Tan et al. 2000). The primary components of rice grain quality include appearance, eating, cooking, milling and nutritional qualities. Each of these components consists of attributes whose values are determined by their physical-chemical properties and other socio-cultural factors such as the history and traditions of the localities where rice is grown. Appearance quality is determined by grain dimension as stipulated by grain length, width, width-length ratio, grain size and shape, and translucency of the endosperm (Unnevehr et al. 1992; Juliano and Villareal 1993). Grain dimension can be used to classify rice into so-called short, medium and long grain types. Medium and short grain cultivars tend to exhibit low amylose content, high gelatinization temperatures, and moist, chewy cooking properties. Long grains generally possess high amylose content and remain separate after cooking with a dry, fluffy consistency (McKenzie and Rutger 1983).

Eating and cooking qualities are primarily influenced by the physical properties of starch in the endosperm, composed of amylose and amylopectin. The amylose content (AC) of rice, recognized as one of the most important determinants of eating and cooking qualities (Bao et al.

2002; Juliano 1985; Webb 1980; Zhou et al. 2003), has been reported to be governed by the waxy (Wx) locus and mapped to chromosome 6 (Tan et al. 1999; Zhou et al. 2003). Other studies have reported AC to be specified by a single major gene with modifications by minor genes (Bollich and Webb 1973; IRRI 1976; McKenzie and Rutgers 1983; Okuno et al. 1983; Kumar and Khush 1988). Additional reports have indicated that AC exhibits a complex genetic basis due to the triploid nature of the endosperm that results in additional cytoplasmic and epistatic effects (Mo 1993; Pooni et al. 1993). Three independent studies reported that the waxy locus was linked to a gene for alkali spreading score, a measure of the temperature at which the rice grain becomes gelatinous during cooking (McKenzie and Rutgers 1983; Ghosh and Govindaswamy 1972; Sano 1984).

Milling quality is assessed using three principal characteristics, namely brown rice, milled rice, and head milled rice. Brown rice consists of grains from which the bran has not been removed by milling. Milled rice is made up of whole and broken rice grains that have the bran removed. Head rice, or the proportion of whole kernel including broken kernels that are 75-80 percent of the whole kernel, is a major factor determining rice market value and is one of the most important criteria for milled rice.

Rice is known to be unique among cereals by having a storage protein primarily made of glutelin, which has a more balanced amino acid profile than the prolamine-rich storage proteins found in most cereals (Juliano 1985). Increasing the protein content of rice may increase and balance the protein intake of people who depend on rice as a staple food. Protein content has been shown to exert a direct impact on the chemical and physical properties of cooked rice (Hamaker and Griffin 1990, 1991; Marshal et al. 1990; Juliano 1993; Hamaker 1994).

The complexity of inheritance and environmental effects on grain quality traits has hampered rapid gains in traditional breeding programs that rely on phenotypic selection criteria.

Recent developments in simple sequence repeats (SSR) or microsatellite molecular markers coupled with linkage maps (McCouch et al. 1997; Harushima et al. 1998; Temnykh et al. 2000; McCouch et al. 2002) have increased the prospects for genetic improvement of cooking quality in rice via marker assisted selection (Causse et al. 1994; Harushima et al. 1998). The potential for this marker-based approach was recently reported for enhanced eating characteristics of the Chinese variety Zhenshan 97 by introgression of the waxy gene region from the restorer line Minghui 63 (Zhou et al. 2003).

Undomesticated rice has been shown to harbor useful genes for agronomic traits among progeny of *O. sativa* x *O. rufipogon* crosses as revealed by quantitative trait loci (QTL) analysis (Xiao et al. 1996, 1998; Moncada et al. 2001). However, little attention has been given to wild rice species as potential sources of genes for grain quality. QTL mapping studies of grain and milling quality have focused exclusively within the *O. sativa* germplasm (Tan et al. 1999; He et al. 1999; Tan et al. 2000; Tan et al. 2001; Zhou et al. 2003). *O. glaberrima* Steud., the native cultivated rice of economic importance in West Africa with domestication from the ancestral *Oryza breviligulata* A. Chev. & Roehr is believed to be independent to that of *O. sativa* (Porteres 1950; Second 1982) and may harbor useful genes.

Deviation of observed frequencies of individuals in a given genotypic class from their expected Mendelian frequencies within a segregating population has been defined as segregation distortion (Sander and Novitski 1957; Lyttle 1991; Xu et al. 1997; Lu et al. 2002). The phenomenon of segregation distortion has been frequently reported in wide crosses in many crop species (Helentjaris et al. 1986; Zivy et al. 1992; Xu et al. 1994; Lefebvre et al. 1995; Liu et al. 1994) and documented in crosses between rice sub-species (Xu and Shen 1992), inter-species (Lorieux et al., 2000; Heuer and Miezán 2003) and in additional studies where different

molecular markers have been used for genetic mapping (McCouch et al. 1988; Causse et al. 1994; Xu et al. 1995; Harushima et al. 1996; Yamagishi et al. 1996).

The genetic basis of segregation distortion in rice was first reported to be associated with the occurrence of a gametophytic gene (*ga-1*) on chromosome six (Iwata et al. 1964). Nakagahra (1972; 1986) further reported the presence of *ga-2* and *ga-3* on chromosome three. Lin et al. (1992) showed that segregation distortion may be due to abortion of male and female gametes or selective fertilization in certain interspecific or inter-sub specific crosses. The work of Nakagahra (1972) led to the suggestion that the gametophytic loci were responsible for the partial or total elimination of gametes carrying one of the parental alleles. Segregation distortion at a marker locus would occur as a result of linkage between the marker and the gametophyte gene (*ga*) which would confer low pollinating ability (Nakagahra 1972). Xu et al. 1997 reviewed the occurrence of segregation distortion across a wide range of species and mapping populations. Recombinant inbred lines (RILs) were found to display the highest frequencies of distorted markers (40 percent) compared to F_2 , doubled haploid (DH) or backcross populations (20-30 percent). In the RILs there are more cycles of meiosis compared to DH lines and F_2 , thus giving more opportunity for segregation distortion in RILs than F_2 s or DH lines.

Segregation distortion has been identified as a problem often encountered in mapping that can produce deviations of single locus segregation ratios from expected frequencies. (Lyttle 1991; Zivy et al. 1992). Several authors have discussed methods to test linkage or estimate recombination frequencies between genes showing aberrant segregation ratios (Bailey 1949; Garcia-Dorado and Gallego 1992; Lorieux et al. 1995). The classical estimate of linkage (Garcia-Dorado and Gallego 1992) is defined as the ratio of the number of recombinant individuals over the total number of individuals in a population. In this case, the proportionality between the expected frequencies of the parental and recombinant classes remains the same. The Bailey's

estimate takes into consideration the viability of the dominant allele relative to the recessive and uses maximum likelihood procedures to estimate linkage (Bailey, 1949). Lorieux et al. 1995 examined various models developed for estimating recombination fractions for backcrosses and found that Bailey's estimate was more consistent and efficient than the classical estimate of Garcia-Dorado (1992). Bailey's method is used for statistical software like Mapmaker (Lander et al. 1987), Joinmap (Stam 1993), QTL Cartographer (Zeng 1994), and others have been designed for constructing genetic maps. However, these programs are designed for markers that exhibit Mendelian segregating ratios. There is no option in these programs to adjust for genetic markers which show deviations from the expected Mendelian frequencies. The software program "Mapdisto" was recently developed by Lorieux et al. (2000) for the analysis of molecular marker data showing aberrant segregating ratios.

Sterility barriers between *O. sativa* and *O. glaberrima* in early hybrid generations have limited the transfer of useful genes between these species (Jones et al. 1996; Second 1984). Heuer and Miezán (2003) recently assessed hybrid sterility in an *O. glaberrima* x *O. sativa* crosses using a segregating BC₂F₃ population, and microsatellite markers mapped to the waxy promoter region of chromosome six. These researchers found that fertile plants were homozygous for the microsatellite marker while semi-sterile plants were heterozygous, suggesting the presence of a gene causing hybrid sterility. Despite the sterility barriers that hinder *O. sativa* x *O. glaberrima* crosses, natural gene flow between *O. glaberrima* and *O. sativa* or *O. longistaminata* have been reported and analyzed by several workers (Takeoka 1965; Second 1984 and De Kochko 1987). Several useful genes from *O. glaberrima* germplasm have been introgressed into adapted *O. sativa* cultivars. For example, Jones et al. 1997 characterized 1,130 accessions of *O. glaberrima* that harbored several desirable morphological and agronomic traits including seedling vigor, growth duration, plant height, panicles m⁻², and grain shape, a

determinant of milling quality. Substantial levels of transgressive segregation were detected for these traits among progenies of *O. sativa* and *O. glaberrima* crosses. Good milling and eating quality traits were also found in this population (WARDA 1998). Moreover, accessions of *O. glaberrima* often have shown superior attributes for several agronomic traits under poor management conditions as well as resistance to biotic (viruses, nematodes and insects) and abiotic (acidity, iron toxicity, drought) stresses (Attere and Fatokun 1983; Sanoet al.1984; Ghesquiere et al. 1997; Reversat and Destombes 1998). The recently released NERICA varieties for West Africa, derived from *O. sativa* x *O. glaberrima* hybrids, show increased resistance to several pests, enhanced levels of protein, and a three-fold increase in grain yield over traditional *O. glaberrima* varieties (Linares 2002).

In most rice breeding programs, a small core of adapted lines are used repeatedly to develop varieties. Consequently, the genetic base of rice in most rice producing countries is narrow. There is need to widen this genetic base through introgression of genes from wild relatives of rice. These studies suggest that interspecific hybridization of the African (*O. glaberrima*) and Asian rice (*O. sativa*) can help provide useful variation for future advances in rice grain quality.

In this study, a molecular marker-based analysis of QTL for traits that determine milling and grain quality was carried out using DH lines derived from the cross between *O. sativa* and *O. glaberrima*. These six traits include milling quality, protein content, amylose content, alkali spreading score, grain length, grain width and length/width ratio.

The primary objective of this research was to map the QTLs for milling, eating and cooking qualities of rice using a population of DH lines derived from the interspecific cross between African (*O. glaberrima*, IRGC 103544) and Asian rice (*O. sativa*, cv. Caiapo). Information about molecular markers tightly linked to QTLs that control these traits will facilitate breeding strategies to improve rice milling, eating and cooking qualities. In addition, detection and

assessment of segregation distortion are needed for accurate QTL mapping involving the DH material. Finally, detection of *O. glaberrima*-derived QTLs and their comparison with previous studies will shed light on the potential of African rice as new sources for enhanced grain quality traits.

2.2. Materials and Methods

2.2.1 Plant Materials

Caiapo, a tropical *japonica* commercial variety developed by the Brazilian national rice program for upland acid soil conditions (Anonymous, EMBRAPA 1997), was used as the recurrent parent in this study. This variety is characterized by good milling and eating characteristics, long grain type, early maturity, low tiller number, 80 cm height, and typical yields of 2.5 t/ha under upland conditions (Anonymous, EPAMIG 1994). *Oryza glaberrima*, IRGC accession # 103544 originally collected in the wild from Mali, Africa, grows to ~ 95 cm height with resistance to several biotic and abiotic stresses (Dr. Brar IRRI pers. communication). Excellent vegetative growth helps suppress growth of weed pests (Dingkuhn et al. 1996).

2.2.2. Population Development

Production of DH lines was carried out by Dr. Cesar Martinez, CIAT, Cali, Colombia. IRGC accession # 103544 served as the male parent in crosses to Caiapo. F₁ plants were grown in 1997 in the greenhouse at the Centro Internacional de Agricultura Tropical (CIAT) in Cali, Colombia. A total of 200 F₁ seeds were produced. Since all F₁ plants were completely sterile 20 individuals were used as the female in backcrossing to Caiapo. A total of 154 BC₁F₁ plants were produced and transplanted under irrigated conditions to the field in 1998. All plants were sterile, and 103 BC₁F₁ plants were selected and backcrossed to Caiapo to generate the BC₂ generation, which was grown under irrigated field conditions. A high level of sterility in this material was observed. The final backcross to Caiapo was completed in 1999 that generated 97 BC₃F₁ plants

grown under irrigated conditions to ensure survival and good plant development. Anthers were collected from each BC₃F₁ plant and used in anther culture as described by Lentini et al. (1995). A total of 695 DH plants were obtained and grown under irrigated field conditions in 2000. Wide genetic variability was observed in the R1 generation in terms of number of tillers per plant, plant height, days to flowering, number of grains per panicle, panicle size, grain weight, presence of awns, yield potential and sterility. A set of 312 DH lines representing the observed genetic variability was chosen for further agronomic and molecular characterization. Seed from each DH line was grown one generation to produce sufficient seed for subsequent phenotypic and genetic analyses.

2.2.3 Field Experiments

The first field trial was conducted at Palmira, Colombia in August 2001 by Dr. Martinez. The 312 DH lines were planted under irrigated conditions in a randomized complete block design, in two-row plots, five meters long with three replications. Twenty-five day old seedlings were transplanted at a spacing of 30 x 30 cm. Caiapo and *O. glaberrima*, IRGC accession #103544, were also included as controls. Fertilizer was applied at the rate of 120 kg N ha⁻¹, 73 kg K₂O ha⁻¹, 63 kg P₂O₅ ha⁻¹, and 4 kg ZnSO₄ ha⁻¹. A post emergence application of Butaclor and Bentazol each at the rate of 3 Lha⁻¹ was used to control weeds supplemented by manual weeding as needed. Data on agronomic traits including flowering, plant height, tiller number, panicle sterility, grain weight and plot yield were taken. Experimental plots were harvested in December 2001. The following grain quality data was taken after harvest, percent brown rice, percent head rice, percent rice bran, percent milled rice, percent amylose, alkali spreading score and percent protein. Grain dimensions such as grain length, grain width and grain length/width ratio were also measured.

In April 2002, a second field experiment with the DH lines and the two parents was planted in a randomized complete block design with two replications at the Rice Research Station, Crowley, Louisiana. Seeds were planted mechanically using a Kincaid cone planter. Fertilizer (NPK 8-24-24) was applied at the rate of 275 kg ha⁻¹ at planting and top dressed with urea at 40 kg ha⁻¹ at 35 days after planting. Thinning of seedlings was performed three weeks after planting to produce 20 cm within and 30 cm distances between rows. Clincher (Cyhalofop) herbicide was applied at 1 L ha⁻¹ 30 days after planting. Insects were controlled by spraying Karate at 0.15 L ha⁻¹ 30 days after planting. Irrigation was applied during prolonged dry periods that occurred in June and July. Harvesting was completed in October 2002. IRGC #103544 and ~30 percent of the DH lines did not flower at the Louisiana location, presumably due to photoperiod sensitivity. Therefore, a high percentage of missing data from this site precluded its use in the analysis.

2.2.4 Measurement and Evaluation of Grain Quality and Milling Traits

2.2.4.1 Milling Quality

Rice grown at the Colombia location was harvested and stored at room temperature for at least 3 months before processing. The length and width of 20, fully formed paddy rice grains from each DH line were measured using a vernier caliper. The length-width ratio was calculated as the grain length divided by the grain width. Hulls were removed from 50 g of rough rice from each line using a Model TH035A Satake Huller (Houston, TX) to yield brown rice. The embryo and the bran layer were removed from brown rice by passing thorough a McGill miller, model #1, (Phillip Rahm International). Long grains and medium grains were milled separately for 30 seconds each. For the long grain, an 858 g weight was applied 9.1 cm from the fulcrum of the pressure plate saddle of the milling machine. For medium grains this weight was applied at 16.7 cm from the fulcrum of the pressure plate saddle. The whole plus broken kernels obtained after

milling were defined as the total milling yield. The hulls and bran were collected during milling and weighed. Total milled rice was separated into whole rice plus grains that were at least 75 percent of whole rice to constitute the head rice while the remainder was regarded as broken rice.

2.2.4.2 Amylose Content

Well-mixed samples and standards were ground through a 0.40 mm screen, using a model #3010-018 UDY mill (UDY Corporation, Fort Collins, Co.) and allowed to equilibrate over night at room temperature in a sample holding cabinet. Sixty milligrams of milled ground rice were weighed and transferred to sample culture tubes (Fisher Scientific, item #T3062-8). One ml of 100 percent ethanol was added to each sample that was agitated gently for 5 minutes. Samples were covered with plastic wrap and allowed to stand at room temperature overnight in 1N NaOH. A total of 54 ml of distilled, deionized water were added to samples and vortexed for 10 –15 seconds using a Maxi-Mix 1 mixer, (type 16700, Barnstead/Thermolyne, Dubuque, Iowa,). Samples were held overnight at room temperature and apparent AC was determined on an auto-analyzer 3 (Bran and Luebbe, Roselle, IL, model AA3) using automated analyzer control and evaluation software AACE, Version 5.24 (Bran and Luebbe, Roselle, IL,).

2.2.4.3 Alkali Spreading Score

The method of Little et al. 1958 was used for conducting the alkali spreading test. Ten milled rice grains from each parent or DH line were immersed in 1.7 percent potassium hydroxide solution at room temperature for 23 hours. Grains were carefully separated using forceps, and the spreading value of the grains was scored by visual assessment using the method of Jennings et al. 1979. Alkali spreading score is known to be inversely related to the temperature at which gelatinization of rice occurs (Lanceras et al. 2000).

2.2.4.4 Protein Content

Protein content was determined by a Nitrogen Gas Analyzer (model 528, LECO Corporation). 1mg Samples were placed into a quartz combustion tube in an induction furnace at 900°C. Total crude protein was calculated from the nitrogen content of the processed grain (%N x 5.95 = % protein).

2.2.5 Molecular Marker Analysis

The microsatellite genotyping of the materials was carried out by Dr. Joe Tohme from CIAT, Colombia. The population of 312 DH was analyzed using a total of 100 polymorphic microsatellite markers between the parents distributed randomly throughout the genome at an average distance of 10.5 cM. Modifications to the polymerase chain reaction (PCR) profile as described by Chen et al. 1997 were as follows: initial denaturing step, 94°C, 15 seconds, annealing for 15 seconds varied between 55°C and 64°C depending on the primer used. Extension at 72°C, 15 seconds, total 35 cycles, final extension at 72°C, 5 minutes.

2.2.6 Statistical Analysis, Map Construction and QTL Detection

Correlation among traits was evaluated using PROC CORR (SAS Institute 1998). QTL analyses associated with markers for each trait were performed using MAPMAKER 3.1 (Lander et al. 1987; Lincoln et al. 1992). Linkage groups were created with a minimum LOD score of 3.0 and a recombination fraction of 0.4 using the “group” command. The marker order within the linkage groups was determined using the “compare”, “try” and “ripple” commands. Map distances were calculated by utilizing the Kosambi function. Due to distorted segregation (i.e. departure from Mendelian segregation) detected on all chromosomes, the “MapDisto” software program (Lorieux 2000) was used to adjust for non-Mendelian inheritance of certain microsatellite markers. Transformation of data for normality of each trait was performed using inverse and log transformations in Excel 2002 (Microsoft Excel 2002) and normality was

checked using PROC UNIVARIATE in SAS (SAS Institute 1998). QTLs were detected by interval (Lander and Botstein 1989) and composite interval mapping procedures (Zeng et al. 1994). A model using five co-factors was selected in the QTL Cartographer software program (Basten et al. 1994, 1997; <http://statgen.ncsu.edu/qtlcart/cartographer.html>) to control for genetic background and applied to the composite interval mapping procedure. Only QTLs detected both by interval and composite interval mapping were used in the analysis.

2.3. Results

2.3.1 Mean and Frequency Distribution of Traits

2.3.1.1 Percent Brown Rice

The Caiapo parent produced 5.4 percent higher brown rice than IRGC 103544. The mean percent brown rice of the DH lines was approximately 1 percent lower than that of the Caiapo parent (Table 2.1), but 4 percent higher than IRGC 103544. The unimodal frequency distribution of percent brown rice among the 312 DH lines shown in Figure 2.1a suggested polygenic inheritance of this trait. Distribution was skewed towards the Caiapo, most likely due to the three backcrosses made to this parent. The large variation in percent brown rice among DH lines and the pattern of the frequency distribution (Fig. 2.1a) suggested ample opportunity for improvement of this trait. Nearly one-quarter of the DH lines showed transgressive segregation for high percent brown rice. In a separate RIL mapping population derived from *O. sativa* x *O. sativa* cross (Tan et al. 2001), the differences between the parents were relatively small and the RILs showed 50 percent less variability than the DH lines used in this study.

2.3.1.2 Percent Head Rice

Percent head rice for the Caiapo parent was 43 percent higher than IRGC103455 (Table 2.1). DH lines on average produced 10 percent lower head rice than the Caiapo parent, although six percent of the lines showed transgressive segregation for this trait. As was the case for

percent brown rice, the frequency distribution of percent head rice was unimodal and skewed towards Caiapo (Fig 2.1b). The continuous pattern of the distribution suggested quantitative inheritance, a result similar to that of Tan et al. 2001 who observed a range of 23 – 75 percent among the RILs tested.

2.3.1.3 Percent Rice Bran

The DH lines had 11 percent more bran than the Caiapo parent, but 15 percent lower than IRGC103544 (Table 2.1). This trait displayed apparent discrete distribution that suggested simple genetic control (Fig. 2.1c), but inadvertent selection during three backcrosses to Caiapo may explain this result. For this trait, 12 percent of the DH lines showed transgressive segregation for high percent bran while 10 percent displayed values lower than Caiapo.

2.3.1.4 Percent Milled Rice

The percent milled rice of 312 DH lines is presented in Table 2.1. The mean percent milled rice of the DH lines was 2 percent lower than the Caiapo parent, but approximately 6 percent higher than IRGC103544. From Table 2.1, percent milled rice had a range of values from 33 percent to 89 percent. However, the frequency distribution in Fig 2.1d, showed that 70 percent of the DH lines was skewed towards the Caiapo parent. The discrete distribution observed among the DH lines indicated simple genetic control. Three backcrosses were made to the Caiapo during population development. Inadvertent selection during backcrossing may be responsible for this result. Ten percent of the DH lines showed transgressive segregation for high milled rice while only 2 percent were transgressive segregants for low milled rice. The distribution of milled rice among RILs reported by Tan et al. 2001 (RILs) showed a narrower range of values (61.2-77.5) compared to (33-89) for the interspecific progenies used in this study, but the mean of the RIL was 3 percent higher than that of DH lines.

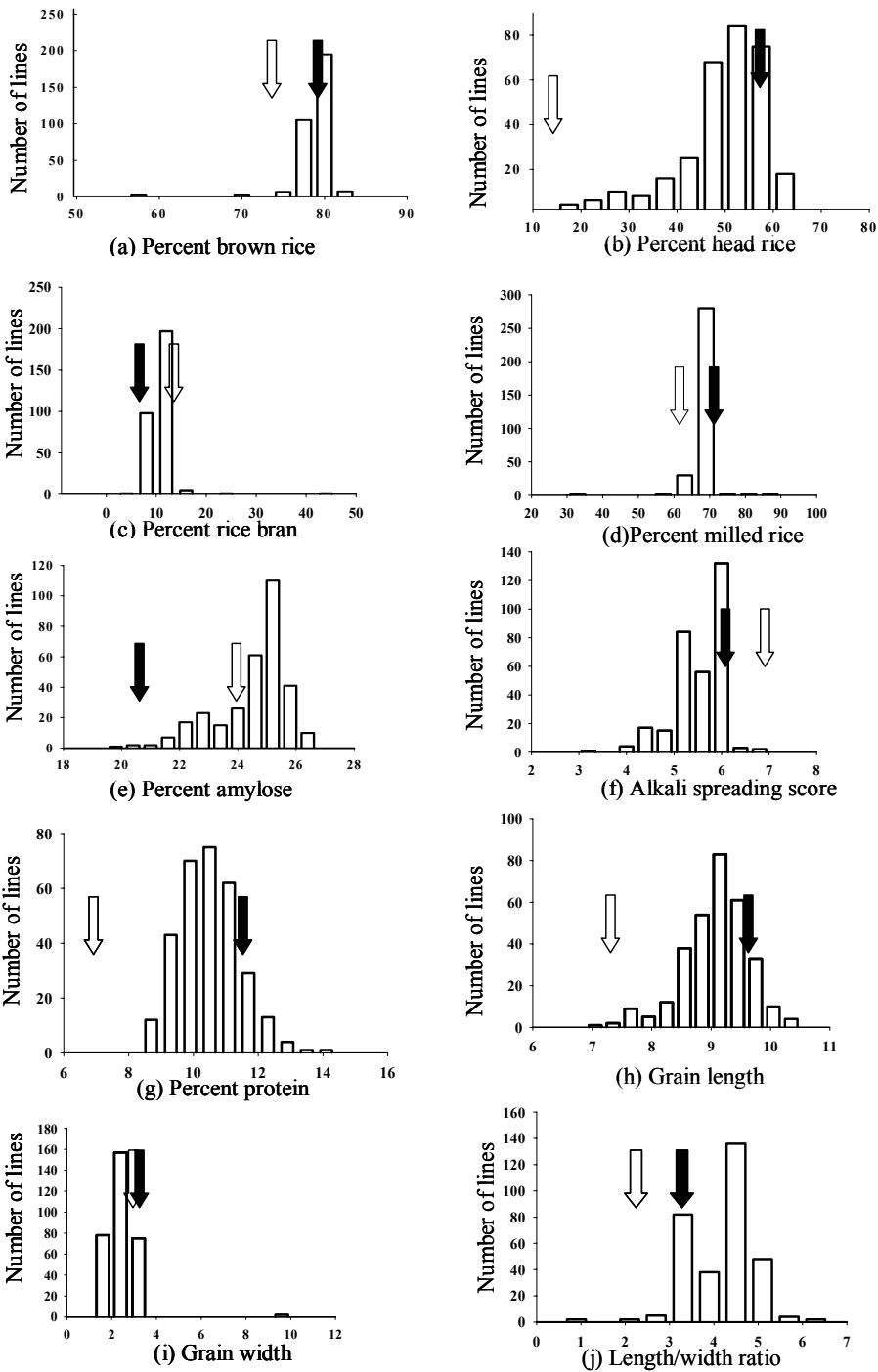


Figure 1. Distribution of milling and grain quality characteristics among 312 DH rice lines evaluated at CIAT, Colombia, 2001. Filled arrows represent mean values for the Caiapo parent while empty arrows represent mean values for the *O. glaberrima* parent (IRGC 103544).

2.3.1.5 Amylose Content

The mean amylose content of DH lines was 4 percent higher than the Caiapo parent, but similar to that of IRGC103544 (Table 2.1). Amylose content showed continuous distribution suggesting that this trait was not simply inherited in this cross. McKenzie and Rutgers (1983) reported amylose content was controlled by 1-3 major genes with additional modifying genes. Lanceras et al. (2000) reported quantitative inheritance while He et al. (1999) detected a single major gene controlling amylose content. More than three-quarters of the DH lines in the current study showed transgressive segregation for high amylose content (Fig 2.1e).

2.3.1.6 Alkali Spreading Score

The two parents did not differ in their alkali spreading score, and the DH lines were generally intermediate in value to the parents with some individuals equal to the high *O. glaberrima* parent. The continuous distribution observed for this trait indicated quantitative inheritance (Fig 2.1f). Transgressive segregants for low alkali spreading score were predominant as 60 percent of the DH lines produced scores between 3.2 and 5.7. The distribution of alkali spreading score was skewed towards IRGC 103544. The range of values obtained in this study was similar to that of Bao et al. (2002), but three times higher than values reported by Lanceras et al (2000).

2.3.1.7 Protein Content

There was a 4 percent difference between the two parents in their protein content (Table 2.1). The mean protein content of the DH lines was 1.2 percent lower than the Caiapo parent, but 3.7 percent higher than IRGC 103544 (Table 2. 1). The protein content of DH lines showed continuous unimodal distribution with transgressive segregation observed for 8 percent of the DH lines. (Fig 2.1g).

2.3.1.8 Grain Length

The Caiapo parent was ~2 mm longer than IRGC103544 in grain length, and the resulting DH lines exhibited a wide range. However, mean grain length of the DH lines was essentially identical to that of Caiapo. Grain length showed continuous distribution indicating quantitative inheritance with positive skewness observed towards Caiapo, the long grain parent (Fig. 2.1h). Ten percent of the DH lines showed transgressive segregation for longer grains.

2.3.1.9 Grain Width

For this trait, no difference was observed between the two parents (Table 2.1). The mean grain width of the DH lines was, 0.6 mm less than both parents. Approximately 80 percent of the DH lines had grain widths between 1.4 and 2.8 (Fig. 2.1i). The frequency distribution was unimodal with 9 percent of the lines showing transgressive segregation for wider grains.

2.3.1.10 Length/Width Ratio

The mean grain size of the DH lines as indicated by their L/W ratio was ~25 percent higher than the Caiapo parent and 40 percent than IRGC 103544. The range of grain sizes showed that the smallest grains among the DH lines were similar in size to the IRGC103544 Table 2.1. The frequency distribution of this trait indicated quantitative inheritance with ~ 80 percent of the lines showing transgressive segregation for high L/W ratio.

2.3.2 Correlation of Traits

Correlation coefficients among the 10 traits were determined both at the $P < 0.05$ and $P < 0.01$ levels. Table 2.2 shows ~ 40 percent of all pair wise correlations were statistically significant ($P < 0.01$) with an additional 8 percent of the correlations significant at the $P < 0.05$ level. A positive correlation between alkali spreading score and protein content was detected, an association not previously reported. Negative correlations were detected between protein content and length-width ratio, and a strong negative correlation between rice bran percentage and milled

rice percentage. Correlations among the remaining traits in this study were consistent with previous research using both inter sub-specific and *O. sativa* x *O. sativa* crosses (McKenzie & Rutger 1983, Tan et al. 2000).

Table 2.1. Mean values for 10 grain quality traits of parents and 312 DH lines derived from *O. sativa* (Caiapo) x *O. glaberrima* (IRGC103544) cross, Colombia, 2001

	BR	HR	Brn	MR	AC	Alk	PRO	GL	GW	LWR
Caiapo	79.6a	59.0a	9.4b	70.2a	20.5b	6.0a	11.7a	9.6a	2.9a	3.2a
IRGC1035 44	74.2b	15.2b	12.2a	62.0b	23.8a	7.0a	6.8b	7.4b	2.9a	2.5a
DH lines	78.9	49.2	10.6	68.2	24.5	5.5	10.5	9.0	2.3	4.1
Range	57-82	15-67	4-45	33-89	20-26	3-7	7-14	7-10	1.5-3.3	2.2-6.3
S.E	0.10	0.53	0.18	0.18	0.07	0.03	0.5	0.03	0.03	0.04

^a BR = Percent brown rice, HR = Percent head rice, Brn = Percent rice bran, MR = Percent milled rice, AC = Percent amylose, Alk = Alkali spreading score, PRO = Percent protein, GL = Grain length (mm), GW = Grain width (mm), LWR = Grain length/width ratio.

S.E. = Standard error

Means bearing the same alphabet do not differ significantly.

Table 2.2. Correlation coefficients among 10 grain quality characteristics of 312 DH lines derived from *O. sativa* (Caiapo) x *O. glaberrima* (IRGC103544) cross, Colombia, 2001

	BR	HR	BRN	MR	AC	ALK	PRO	GL	GW
BR									
HR	0.18**								
BRN	-0.01	-0.31**							
MR	0.30**	0.39**	-0.84**						
AC	-0.03	-0.02	0.10	0.11*					
Alk	0.10	0.22**	0.03	0.02	0.18**				
PRO	0.01	0.11	-0.15**	0.15**	-0.12*	0.28**			
GL	0.03	-0.18**	0.02	-0.01	-0.13*	-0.12*	0.16**		
GW	0.001	0.07	-0.05	0.05	-0.07	0.08	0.19**	0.26**	
LWR	-0.01	-0.15**	0.10	-0.10	0.06	-0.17**	-0.33**	-0.07	-0.86**

^a BR = Percent brown rice, HR = Percent head rice, BRN = Percent rice bran, MR = Percent milled rice, AC = Percent amylose, ALK = Alkali spreading score, PRO = Percent protein, GL = Grain length (mm), GW = Grain width (mm), LWR = Grain length/width ratio.

* Significant at P < 0.05

** Significant at P < 0.01

2.3.3 Segregation of Marker Loci

A total of 39 out of 100 markers used in this study showed varying amounts of segregation distortion on all 12 chromosomes for this interspecific cross (Table 2.3). The majority of markers (34/39) were skewed towards Caiapo which may be explained by the three backcrosses and subsequent selection against sterile plants during population development. Five of the most distorted markers were detected on chromosomes 3 and 6. All three distorted markers on chromosome 3 and the two most distorted markers on chromosome 6 were skewed toward the *O. glaberrima* parent, which was not expected. This aberration may have occurred during inadvertent selection in the three backcrosses and/or *in vitro* anther culture procedures that produced the doubled haploid material. Similar marker skewness toward the unadapted parent in an *O. sativa* x *O. rufipogon* cross was previously reported that was attributed to the large genetic distance between parental lines (Moncada et al. 2001). In the study reported here, the order and position of 39/100 markers on the linkage map (see below) were grossly distorted or switched in position when compared to Lorieux et al. (2000) or other published maps (data not shown). Five markers (RM267, RM31, RM274, RM194 and RM169) (Table 2.3) located within a 63 cM region on chromosome 5 showed high levels of skewed segregation that has not been reported in previous mapping studies (Harushima et al. 1995; Liu et al. 1997; Lin et al. 1992), and probably represents a hot spot for non-Mendelian gene action in *O. glaberrima*.

2.3.4. Interval Length Between Markers and Map Length

The genetic map produced from this study, after adjustment with the MapDisto software program, showed an average interval length of 10.5 cM between markers, with total map length of 1,050 cM. Previous maps constructed from other mapping populations were 33 percent longer (Chen et al. 1997) or approximately twice the length of the map reported here (Lorieux et al. 2000), even after adjustments made for segregation distortion. Additional maps derived within *O. sativa* germplasm (Panaud et al. 1996, Temnykh et al. 2000 and Causse et al. 1994) showed

Table 2.3 Chi square values and chromosome location of microsatellite markers showing segregation distortion among 312 doubled haploid lines derived from the cross *O. sativa* (Caiapo) x *O. glaberrima* (IRGC 103544)

Markers	Chrom.	Chi square	Prob F	Skewness*	Chrom. Position**
RM5	1	6.958	0.0083	Caiapo	24.3
RM297	1	8.007	0.0047	Caiapo	60.7
RM315	1	8.889	0.0029	Caiapo	86.5
RM226	1	11.50	0.0007	Caiapo	37.8
RM128	1	14.89	0.0001	Caiapo	35.2
RM236	2	9.188	0.0024	Caiapo	13.9
RM110	2	9.85	0.0017	Caiapo	0.0
RM301	2	12.25	0.0005	Caiapo	56.3
RM71	2	13.14	0.0003	Caiapo	40.1
RM174	2	13.66	0.0002	Caiapo	33.0
RM85	3	28.42	0.0001	IRGC 103544	129.8
RM60	3	193.0	0.0001	IRGC 103544	0.0
RM81B	3	231.0	0.0001	IRGC 103544	14.5
RM280	4	7.23	0.0072	Caiapo	39.0
RM124	4	9.788	0.0018	Caiapo	28.0
RM348	4	11.50	0.0007	Caiapo	8.0
RM349	4	13.7	0.0007	Caiapo	14.8
RM241	4	16.22	0.0001	Caiapo	3.2
RM131	4	16.48	0.0001	Caiapo	23.2
RM317	4	18.2	0.0001	Caiapo	6.7
RM267	5	8.49	0.0036	Caiapo	24.0
RM31	5	12.4	0.0001	Caiapo	75.0
RM274	5	12.57	0.0001	Caiapo	84.0
RM194	5	14.76	0.0001	Caiapo	27.0
RM169	5	14.89	0.0001	Caiapo	30.0
RM103	6	8.54	0.0035	Caiapo	225.1
RM253	6	81.81	0.0001	IRGC 103544	115.9
RM190	6	32.21	0.0001	IRGC 103544	0.0
RM10	7	14.89	0.0001	Caiapo	15.0
RM308	8	7.23	0.0072	Caiapo	86.0
RM42	8	8.47	0.0036	Caiapo	46.0
RM25	8	9.97	0.0016	Caiapo	12.0
RM149	8	10.32	0.0013	Caiapo	78.0
RM256	8	18.06	0.0001	Caiapo	67.0
RM316	9	14.70	0.0001	Caiapo	5.0
RM184	10	7.23	0.0072	Caiapo	22.0
RM239	10	9.60	0.0019	Caiapo	0.0
RM229	11	18.72	0.0001	Caiapo	46.0
RM277	12	12.76	0.0004	Caiapo	26.0

*skewed marker segregation towards Caiapo or IRGC 103544 parent

** Chromosomal map location of marker in cM.

map lengths ~ 1 to 2-fold longer than the map shown in Fig. 2.3. It has been observed that marker distortion without additional adjustments was associated with a 27 percent reduction in the length of chromosome 8 (Lorieux et al.1995). The MapDisto program used in the present study was able to successfully restore marker order when compared to the map of Lorieux et al. (2000), but adjustments were not sufficient to produce a map similar in length to those reported previously by Panaud et al. 1996 and Chen et al. (1997).

2.3.5 QTL Analysis of Grain Quality Traits

2.3.5.1 Percent Brown Rice

Three QTLs, *br1*, *br7* and *br8* were detected on chromosomes 1 in the interval RM297-RM315), 7 (RM10-RM351) and 8 (RM126-RM13) explaining 2.8, 4.9 and 3.6 percent of the phenotypic variance, respectively (Table 2.4). Altogether, these three explained 11.3 percent of phenotypic variation for this trait. Alleles from the Caiapo parent were associated with increased percent brown rice at all three loci. In contrast to this work, Tan et al.2001 detected one QTL for brown rice on chromosome 5 that explained 10 percent of the phenotypic variation.

2.3.5.2 Percent Head Rice

Five chromosomal regions were associated with QTLs for percent head rice (Table 2.4). The Caiapo parent contributed all five QTLs for this trait. On chromosome 1, *hr1* was detected in the interval RM140 – RM5 that explained 17.8 percent of phenotypic variation with a LOD score of 3.4. The QTL *hr3* was detected on chromosome 3 which accounted for 12 percent of phenotypic variation (LOD = 4.0). A major QTL *hr6* with LOD of 11.9 was located on chromosome 6 in the interval RM190 – RM253 accounting for 54.1 percent of total variation.

On chromosome 8, *hr8* was detected with a LOD score of 3.2 and this QTL explained 7.6 percent variation. QTL *hr11* mapped to chromosome 11 in the interval RM209-RM229. This

QTL produced a LOD of 4.0 and explained 17.4 percent of phenotypic variation. The QTLs *hr3* mapped 8 cM away from a QTL for head rice reported by Tan et al. 2001. The QTL *hr3* on chromosome 3 in the interval RM388-RM55 accounted for 12 percent variation while one QTL *hr11* was mapped to chromosome 11 in the interval RM209 – RM229 explaining 17.4 percent of total variation. Tan et al. 2000 also reported a QTL for percent head rice that mapped 8.7 cM from the QTL *hr3* detected in this study (Fig.2.2).

2.3.5.3 Percent Rice Bran

Four QTLs, *rb2*, *rb4*, *rb7* and *rb10*, were significantly associated with percent rice bran with LOD values of 55.6, 28.4, 3.2 and 52.9, respectively (Table 2.4). The QTLs *rb2*, *rb4* and *rb10* mapped to chromosomes 2, 4, and 10, explaining 32.7, 39.7 and 32.8 percent variation, respectively. The fourth QTL, *rb7* was detected on chromosome 7 that accounted for 4 percent of phenotypic variation. Alleles *rb4* and *rb7* were contributed by IRGC103544 while *rb2* and *rb10* were derived from the Caiapo parent. The QTL on chromosome 2 (*br2*) was found in the interval (RM236 – RM279) that explained 32.7 percent of total variation. Alleles from the Caiapo parent increased rice bran at this locus.

2.3.5.4 Percent Milled Rice

Two regions were found associated with QTLs for percent milled rice on chromosomes 5 and 7 (Table 2.4). The QTL *mr5* on chromosome five (LOD = 3.2) was detected in the interval RM188-RM26, 47 cM distance from that reported by Tan et al. (2001). This QTL was derived from IRGC 103544 and accounted for 6.2 percent of phenotypic variation. Another QTL *mr7* was detected on chromosome 7 (LOD = 3.71), which accounted for 5.3 percent of phenotypic variation. The Caiapo parent increased percent milled rice at this locus. The other QTL of smaller effect was found in the interval RM125 – RM111 accounting for 5.3 percent of phenotypic variation with alleles coming from Caiapo. Tan et al. (2000), detected two minor

QTLs, one each on chromosome 3 and 5 that explained 4.8 and 7.0 percent of phenotypic variation, respectively.

2.3.5.5 Amylose Content

Three QTLs were detected to have an effect on amylose content. One major QTL, *amy6* was detected on chromosome 6 in the interval RM190 – RM253 with a LOD score of 19.3. This QTL accounted for 73.7 percent of phenotypic variation. Other researchers have also detected QTL for amylose content at the same region near the Waxy gene on chromosome 6 (He et al. 1999; Tan et al. 1999; Lanceras et al. 2000). Two additional QTLs were also detected on chromosomes 3 and 8, *amy3* and *amy8*, which explained 21.5 percent and 10.9 percent of phenotypic variation with LOD 4.0 and 3.2 respectively. This result corroborates the suggestion that genes of major effects control amylose content and modifying genes (McKenzie and Rutger 1983).

Earlier studies based on mapping populations derived from *O. sativa* x *O. sativa* crosses detected QTLs controlling amylose content on chromosomes other than chromosome 6. For instance, He et al. (1999) found an additional QTL on chromosome 5 that explained 11.8 percent of phenotypic variation. Lanceras et al. (2000), in a study to map genes for cooking and eating qualities in Thai Jasmine rice, found QTLs on chromosomes 3, 4, 6 and 7. In all these studies, there was a general agreement that chromosome 6 harbored a major gene for amylose content in the region of the waxy locus. They concluded that the presence of one major QTL as well as several modifiers was responsible for expression of this trait.

2.3.5.6 Alkali Spreading Score

The two QTLs detected for alkali-spreading score mapped to chromosome 6 (*alk6-1* and *alk6-2*) (Table 2. 4 and Fig. 2.2). The QTL *alk6-1* flanked by markers RM190 and RM253

accounted for 50.1 percent of phenotypic variation with a LOD score of 32.5. Alleles from IRGC 103544 were associated with increased trait value at this locus. He et al. (1999) also detected QTL for alkali spreading score in this region. The other QTL, *alk6-2* detected on chromosome 6 was flanked by RM 253 and RM162. This QTL explained 44 percent of the variations with LOD score of 10. Alleles associated with increase in Alkali spreading scores were contributed by both IRGC103544 and Caiapo. Lanceras et al. (2000) detected two QTLs on chromosomes 6 and 7 for GC accounting for 57 percent of variation. QTL *alk6-1* was detected at the same position as the major QTL controlling amylose content. Similar results were obtained by other researchers (Tan et al. 1999; He et al. 1999) that a single QTL for amylose content and that for alkali spreading score were found at the same locus on chromosome 6. The same observation was made by Bao et al. (2002). This may be due to the pleiotropic effect of a single gene.

2.3.5.7 Protein Content

Four QTLs, *pro1*, *pro2*, *pro6* and *pro11* were detected for protein content. One *pro1* mapped to chromosome 1 in the interval RM226 – RM297 explaining 15 percent of phenotypic variation (LOD = 5.8) (Table 2.4). At this locus, the Caiapo allele was associated with increased protein content. On chromosome 2, a QTL *pro2* was detected in the interval RM112 – RM6 explaining 6.4 percent variation with LOD score of 3. This QTL was from the IRGC 103544 parent. Another QTL *pro6* on chromosome 6 in the interval RM190-RM253 was detected and this explained 8.8 percent of variation.

The smallest QTL effect *pro11* was detected on chromosome 11 explaining 4.8 percent variation with alleles contributed by Caiapo (LOD=3.3). Tan et al. (2001) detected two QTLs for

protein content on chromosome 6 near the waxy locus at the same position *pro6* was mapped in this study. In our study, this QTL explained 8.8 percent of phenotypic variation while Tan et al. (2001) reported 13 percent of variation.

2.3.5.8 Grain Length

Two QTLs were detected to have an effect on grain length. QTL *gl3* on chromosome 3 in the interval RM251 - RM338 explained 12.5 percent of phenotypic variation. One QTL (*gl6*) was detected on chromosome 6 in the interval RM190-RM253 (Table 2.4 and Fig. 2.2). This QTL explained 4.7 percent of phenotypic variation for this trait. Both QTLs reported for grain length were contributed by alleles from the Caiapo parent.

2.3.5.9 Grain Width

No major QTL was found for grain width (Table 2.4). The QTL *gw4* found on chromosome 4 exhibited a LOD score below the threshold value. This result occurred because grain width of both parents was similar. Tan et al. (2000) detected QTLs for grain width on chromosomes 5 and 6 in both F_{2:3} and RIL populations.

2.3.5.10 Length /Width Ratio

Two chromosomal regions were significantly associated with grain length/width ratio. QTLs *lwr1* and *lwr6* were detected on chromosomes 1 and 6, respectively (Table 2.4 and Fig. 2.2). QTL *lwr1* was detected in the interval RM5 – RM165 and accounted for only 4 percent of variation (LOD=2.8). This allele was from IRGC 103544, while on chromosome 6 in the interval RM190 – RM253, a QTL was detected which explains 14 percent of total variation and a LOD score of 13.4. The Caiapo parent contributed this allele.

2.3.6 Epistatic Interactions

A test of epistatic effects was carried out to identify chromosomal regions that by themselves expressed no discernible effects, but interacted with other loci to produce the observed

Table 2.4 Quantitative Trait Loci for Grain Quality Traits among 312 doubled haploid lines derived from the cross *O. sativa* (Caiapo) x *O. glaberrima* (IRGC 103544)

Trait	QTL	Chr.	Marker Interval	Marker Position	Additive	LOD	R ²	Allelic Source
% Brown rice	<i>br1</i>	1	RM297-RM315	60.7	-0.74	3.4	2.8	Caiapo
	<i>br7</i>	7	RM10-RM351	21.0	-2.73	3.4	4.9	Caiapo
	<i>br8</i>	8	RM126-RM137	21.0	-1.38	3.3	3.6	Caiapo
% Head rice	<i>hr1</i>	1	RM140-RM5	24.0	-24.11	3.0	17.8	Caiapo
	<i>hr3</i>	3	RM388-RM55	80.0	-12.51	4.0	12.0	Caiapo
	<i>hr6</i>	6	RM190-RM253	10.0	-2.41	11.9	54.1	Caiapo
	<i>hr8</i>	8	RM126-RM137	21.0	-11.68	3.2	7.6	Caiapo
	<i>hr11</i>	11	RM209-RM229	43.0	-20.60	4.06	17.4	Caiapo
% Rice bran	<i>rb2</i>	2	RM236-RM279	19.0	-18.25	55.6	32.7	Caiapo
	<i>rb4</i>	4	RM349-RM131	22.8	34.49	28.4	39.7	IRGC 103544
	<i>rb7</i>	7	RM10-RM351	21.0	1.02	3.2	4.0	IRGC 103544
	<i>rb10</i>	10	RM184-RM171	26.0	-18.23	52.9	32.8	Caiapo
% Millied rice	<i>mr5</i>	5	RM188-RM26	61.0	35.17	3.2	6.1	IRGC 103544
	<i>mr7</i>	7	RM125-RM11	0.0	-3.62	3.71	5.3	Caiapo
% Amylose	<i>amy3</i>	3	RM7-RM251	60.0	-2.73	3.73	21.5	Caiapo
	<i>Amy6</i>	6	RM190-RM253	36.0	-2.60	19.3	73.7	Caiapo
	<i>Amy8</i>	8	RM230-RM264	104.0	-1.85	3.1	10.9	Caiapo
Alkali spreading score	<i>alk6-1</i>	6	RM190-RM253	6.0	0.87	32.5	50.1	IRGC 103544
	<i>alk6-2</i>	6	RM253-RM162	156.0	-0.87	10	44.0	Caiapo
Protein	<i>pro1</i>	1	RM226-RM297	55.0	-1.01	5.8	15.0	Caiapo
	<i>pro2</i>	2	RM6-RM112	98.0	1.06	3.0	7.4	IRGC 103544
	<i>pro6</i>	6	RM190-RM253	10.0	1.41	3.6	8.8	IRGC 103544
	<i>pro11</i>	11	RM209-RM229	35.0	-0.67	3.3	4.8	Caiapo
Grain length	<i>gl3</i>	3	RM251-RM338	72.2	-0.67	3.8	12.5	Caiapo
	<i>gl6</i>	6	RM162-RM30	204.0	-0.61	3.29	4.7	Caiapo
Grain width	<i>gw4</i>	4	RM131-RM124	45.2	3.83	1.70	3.0	IRGC 103544
L/w ratio	<i>lwr1</i>	1	RM5-RM246	24.3	0.71	2.8	4.0	IRGC 103544
	<i>lwr6</i>	6	RM253-RM162	118.0	-0.74	13.4	14.0	Caiapo

Marker position = position of peak marker in cM within the interval; Additive = additive effects expressed in terms of estimated change in phenotype expected from introgression of the *O. glaberrima* allele into Caiapo genome LOD= Log₁₀ (probability of linkage/probability of no linkage); R²= percentage of variation accounted for by the QTL.

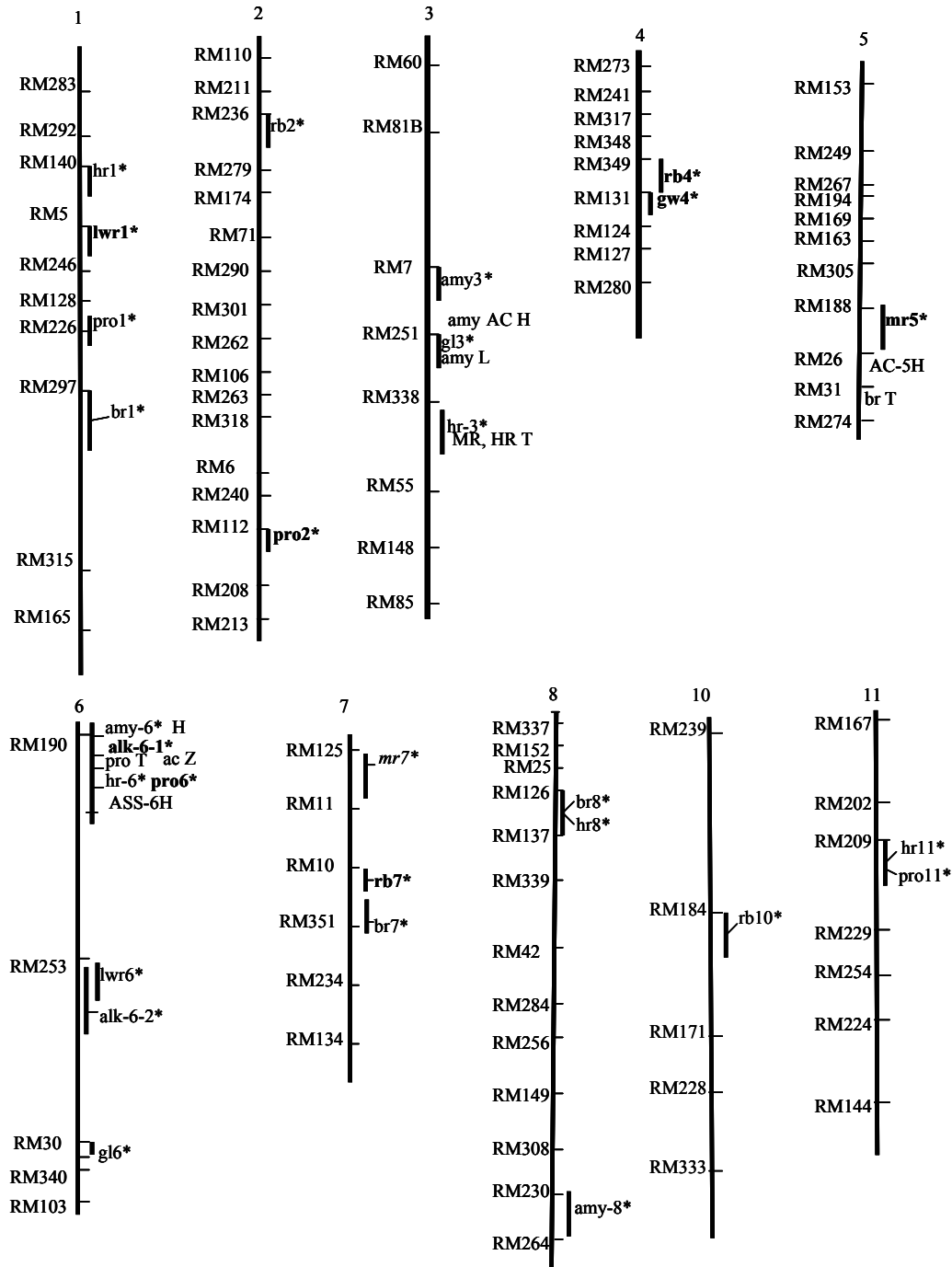


Fig. 2.2. Assignment of 28 QTLs for 10 grain and milling traits on the rice linkage map adjusted by MapDisto program among 312 DH lines derived from *O. sativa* (Caiapo) x *O. glaberrima* (IRGC 103544) cross. Confidence interval for each QTL indicated as bar to the right of chromosome. QTLs in bold indicate positive allelic effects from IRGC 103544. H = QTL reported by He et al. (1999), T = QTL reported by Tan et al. (2001), L = QTL reported by Lanceras et al. (2000), Z = QTL reported by Zhou et al. (2003). For QTLs, *br*: percent brown rice; *hr*: percent head rice; *rb*: percent rice bran; *mr*: percent milled rice; *amy* : percent amylose; *alk*: alkali spreading score ; *pro*: protein content; *gl*: grain length; *gw*: grain width; *lwr*: length/width ratio. * = QTL detected in this study.

phenotype. Table 2.5 shows the list of digenic interactions obtained using the Epistat program (Chase et al. 1997) that subdivides the DH population into four genetically distinct sub-populations determined by the two possible alleles present at each of two homozygous loci. The significance of the interactions between pairs of QTLs was evaluated by likelihood methods and Monte Carlo simulations.

A total of 12 markers were detected on 6 chromosomes that produced 11, two-way interactions. Eight of the 11 interactions were comprised of markers that mapped to different chromosomes. Two general categories of interactions were observed: where loci associated with QTL interacted with ones that were not associated with QTL (non-effect locus) (Tan et al. 2001), and the interaction between loci that were both not associated with QTL (non-effect loci). For example, during QTL analysis, RM253 was found in the interval controlling percent head rice. For this trait there was no demonstrable effect of RM297. However, an interaction between RM253 and RM297 produced an effect on the variation in percent head rice. This interaction may bias the variation explained by the QTL in the interval containing RM253.

Examples of non-effect loci in this study were RM148 and RM85, both not associated with amylose content, but their interaction produced an effect on the trait (amylose). RM253 was found in the interval associated with QTL for percent head rice. There was no demonstrable effect of RM297, but the interaction with RM253 produced an effect on the variation in percent head rice. This interaction may bias the variation explained by RM253. In the same manner, the effect of RM148 on chromosome 3 was conditional upon the presence of RM253. To illustrate digenic interactions observed in this study, data from two locus pairs that showed significant were plotted. To illustrate the epistatic interactions in Table 5, two cases were considered. In Fig. 2. 3a, we see that at even though RM148 produced no main effect on percent head rice, the interaction with RM253 produced an effect which led to variation in head rice percent. This is a

case where a non- effect locus interacted with a locus associated with QTL to produce an effect.

In Fig 2.3b, neither RM81B nor RM333 was associated with QTL for amylose content, but an interaction of the two loci produced variation in amylose content. This is an example of a non-effect locus.

Table 2.5. Significant two-way interaction between different loci.

Trait	Marker 1		Marker 2		F-test ^a P	MC-Test ^b
	Name	Chrom	Name	Chrom		
% Head rice	RM297	1	RM253	6	0.002	0.024
	RM148	3	RM253	6	0.000	0.000
% Amylose	RM283	1	RM253	6	0.005	0.048
	RM81B	3	RM333	10	0.001	0.012
	RM148	3	RM85	3	0.015	0.015
Grain width	RM297	1	RM81B	3	0.028	0.028
	RM297	1	RM167	11	0.024	0.024
	RM297	1	RM209	11	0.002	0.002
	RM81B	3	RM288	9	0.011	0.011
% Rice bran	RM224	11	RM144	11	0.006	0.072
% Milled rice	RM224	11	RM144	11	0.023	0.021

a = F test for the four subgroups of the two marker alleles. b = Monte Carlo simulation using EPISTAT program (Lark et al. 1995) to evaluate the significance of interactions.

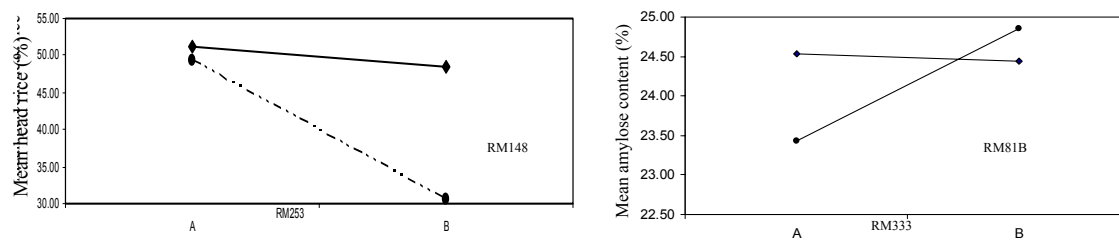


Figure 2.3. The effect of interaction between the marker loci RM253 and RM148 on head rice percent and marker loci RM333 and RM81B on amylose content.

The presence of epistasis between different loci has been shown to be very important in accounting for transgressive segregation in rice (Yu et al. 1977, Yi et al. 1977; Tan et al. 2001) and in soybean (Lark et al. 1995). In the present study, epistatic effects were detected in head rice, amylose content, grain width, percent bran and percent milled rice. All these traits showed transgressive segregation (Fig. 2.1; Table 2.1). However, not all traits displayed hybrid vigor, so epistasis alone, as measured by the Epistat program, cannot explain transgressive segregation for traits showing epistasis in this study. Monte Carlo (MC) simulations are used in situations where some traits and genotypes do not satisfy the normality assumption. In this case MC simulations were used specifically for a given trait and pair of loci at a time to verify the predicted probability of exceeding a given log-likelihood ratio.

It is worthy to note that the effects of interaction estimated for the 2-locus genotypes may not be independent of each other especially when they belong to the same linkage groups (Zeng 1994) and may bias the estimation of digenic genotypes. This potential drawback in the detection of epistasis would occur for only three cases: RM148 and RM85 for amylose content, RM224 and RM144 for percent rice bran, and, RM224 and RM144 for milled rice.

For certain traits, the total variances explained by the loci were high. For example, in the case of head rice which had an R^2 of 54.1 percent, this may be due to epistasis whereby two loci interact to bias the estimated R^2 . Significant interaction was found between RM253 and RM148 on the one hand and RM253 and RM297 on the other. These interactions may have inflated the R^2 value. Recent genetic studies using molecular markers have clearly demonstrated the role that epistatic interactions may play in the genetic basis of quantitative traits (Lark et al. 1995; Li et al. 1997; Yu et al. 1997, 2002). In these various studies, epistatic interactions were reported between loci that did not show significant effects by single locus analyses and those involving

QTLs that were detected by single locus analyses. In situations like these, the interpretations based on single locus analyses become biased since the effect of a genotype at one locus would depend on the genotypes at other loci with which it interacts. For grain width, two non-effect loci interacted to produce an effect on this trait. This may explain the transgressive segregation observed for this trait. Two markers on chromosome 11 interacted to produce a pleiotropic effect on percent rice bran and percent milled rice. It is interesting to note that among the 12 markers that showed epistatic interactions, RM297, RM81B and RM253 were also affected by segregation distortion (Table 2.3).

2.4. Discussion

This study was conducted to detect QTL associated with milling, eating and cooking qualities in DH population developed from *O. sativa* x *O. glaberrima* cross. The mean percent brown rice of the DH lines was similar to the Caiapo parent most likely because the DH lines acquired more alleles from this parent than IRGC 103544 as a result of three backcrosses made to the Caiapo parent. Earlier work by Tan et al. (2001), using RILs derived from an *O. sativa* x *O. sativa* cross, reported that the percent brown rice was similar to one of the parents. The QTL reported by these researchers was detected on chromosome 5 which explained 10 percent of variation. In the DH lines used here, three QTLs were detected on chromosomes 1, 7 and 8. These are potentially new chromosomal regions that may be associated with percent brown rice, but additional data over years and locations are needed to verify this initial result. The low R^2 values obtained for this trait may be related to the low heritability of the trait. Heritability estimate of 29.8 percent has been reported for brown rice (Tan et al. 2001). Therefore, it would be difficult to improve this trait by traditional breeding methods. However, transgressive segregation for percent brown rice observed in our study indicated that it would be possible to select lines with high percent brown rice within this interspecific population.

Percent head rice is a very important milling quality trait. The mean head rice percent of the DH lines was lower than the Caiapo parent, but the wide range of values observed for this trait indicated that the trait could be improved through interspecific hybridization. The transgressive performance of some of the DH lines could have resulted from the detected epistatic interactions. Several studies have reported the importance of epistatic interaction in accounting for transgressive segregation in rice (Moncada et al. 2001; Tan et al. 2001). One of the QTLs detected for head rice, *hr3* was 8 cM away from the interval reported by Tan et al. (2001) who reported an R^2 of 10.1 percent which is similar to 12 percent observed here. One of the QTLs for head rice was also detected for brown rice on chromosome 8 in the interval RM126- RM137. This may be a case of linkage or pleiotropy even though these two traits were not highly correlated (Table 2.2).

The range of values obtained for percent milled rice indicated greater variability compared to an earlier study where *O. sativa* x *O. sativa* population was used (Tan et al. 2001). The mean percent milled rice reported by these researchers for an RIL population was 4 percent higher than the value obtained for DH lines reported here. Two chromosomal regions were detected for percent milled rice *mr5* and *mr7* on chromosomes 5 and 7. The Caiapo parent contributed alleles that increased milled rice in this study, and this material may be a potential source for improving milling yield.

It is generally believed that amylose content is controlled by the effect of one major gene and several modifiers (Bollich and Webb 1973; McKenzie and Rutger 1983; Kumar and Khush 1988; Tan et al. 1999). Other reports have also proposed two major genes plus modifying factors (McKenzie and Rutger 1983) while another study found quantitative inheritance of this trait (Lanceras 2000). One major QTL, *amy6* in the waxy region which controlled 73.7 percent of

phenotypic variation was detected in this study. However, other regions were also found (*amy3* and *amy8*) which explained 21 percent and 10.9 percent of variation, respectively. In spite of segregation distortion, one major QTL for amylose content was detected in the same region on chromosome 6 as other researchers. A QTL for amylose content that accounted for 91 percent of phenotypic variation in the same region has been reported (He et al. 1999). The occurrence of transgressive segregants was postulated to be due to presence of modifying genes (Kumar and Khush 1988). However, the range of amylose content reported in this study (33-89 percent) showed greater variability than among the RIL population of He et al. (1999).

Rice grain quality is usually evaluated according to its suitability for a specific end user. However, good cooking and eating qualities appear to be the most important qualities in view of the fact that they are most widely studied. Most of the mapping studies on rice grain quality have concentrated on cooking quality such as amylose content, gelatinization temperature as stipulated by alkali spreading score and gel consistency (McKenzie and Rutger 1983; Tan et al. 1999, He et al. 1999; Lanceras et al 2000). Many of these studies have identified AC, GC and GT as inherent characteristics of rice that determine cooking and eating qualities as well as processing properties of rice (He et al. 1999; Tan et al. 1999; Lanceras et al. 2000; McKenzie and Rutger 1988; Bao et al. 2002; Webb 1980). It is interesting that an allele from IRGC 103544 was found that increased alkali spreading score which is a measure of gelatinization temperature. This parent may be a potential source for improving gelatinization temperature.

Grain length, grain width and the length/width ratio are important attributes determining the appearance quality of rice grains. The mean grain length of 9.0 mm showed that the DH lines had long grain types on average. In a mapping population developed from a cross between Black Gora and Labelle, Redona and McKill (1998), found transgressive segregation of the same

magnitude reported in this study for grain length, grain width and grain size indicated by L/W ratio. The frequency distribution in both studies indicated polygenic inheritance of this trait. Redona and McKill (1998) detected two QTLs on chromosome 3, while Xu et al. 2000 also detected a QTL for grain length on this chromosome. The QTL *gl3* detected here was found in the same region on chromosome 3 as one of the QTLs detected by these researchers. The gene, Fusayoshi long grain (LK-f), mapped to the same region in the central region of chromosome 3 (Takeda and Saito 1980; Takamure and Kinoshita 1991). It is possible that the *gl3* might be related to Lk-f detected earlier.

The length/width ratio is a measure of grain size (Redona and McKill 1998). This trait mapped to the same region as one of its component traits, i.e. grain length on chromosome 6. Redona and McKill (1998) also observed that grain size tended to be associated with loci for its component traits. The *lwr1* identified in this study appears to be a new region of the genome associated with grain size.

2.4.1 Segregation Distortion

Earlier studies have reported segregation distortion in rice mapping studies on chromosomes 1, 3, 6, 8, 9, and 10 (Harushima et al. 1995), 2, 3, and 12 (Liu et al. 1997), 3, 7, 8, 11, 12 (Lin et al. 1992), and chromosome 6 (Lorieux et al. 2000). High levels (38 percent) of skewed marker segregation were also detected in an *O. sativa* x *O. rufipogon* backcross population (Moncada et al. 2001). The distortion at the proximal end and middle of chromosome 3 has been attributed to the segregation of the gametophyte gene *ga-2* (Nakagahra 1972; Nakagahra et al. 1972). Distorted markers were found in this region (proximal end of chromosome 3) in this mapping study which could be due to the same gametophyte gene. Maekawa et al. (1981) also found a gametophyte gene *ga-6* located on chromosome 4, which

caused gamete abortion in *japonica-indica* hybrids. Liu et al. 1997 estimated genomic regions showing significant effects on hybrid sterility and found these on chromosomes 2 (6.1-29.5cM), 6 (21.8-23.0cM) and 12 (7.0-23.0cM). For this mapping study, we used *O. glaberrima* accession IRGC# 103544 as a parent in this interspecific cross. Though different from TOG5681 used by Lorieux et al. (2000), there is no report to show that differences observed in segregation distortion pattern or map distances could be due to differences in accession. The distorted map observed in this study may be due to altered segregation, or inadvertent selection during backcrossing or a combination of these.

The segregation pattern of microsatellite markers used in this study showed extensive distortion, which necessitated an adjustment with the MapDisto program. The resulting map restored the order of markers, but could not fully adjust the distances between markers, so the map was ~ 46 percent shorter than other rice maps constructed from interspecific mapping populations. This could be due to the extent of segregation distortion of the markers. For instance, chromosomes 3 and 6 were highly distorted (see X^2 values on Table 2.3 determined before the use of MapDisto).

Moncada et al. (2001) also observed 47 skewed markers out of 125 (37.6 percent). However, the chromosomal locations of these markers were not reported. Twenty-eight percent of these markers were skewed towards the Caiapo parent while 8.8 percent were skewed towards *O. rufipogon*. Thirty-nine of the markers used in our present study were skewed and these were found in different positions on the twelve chromosomes. The distribution of these distorted alleles may have contributed to the difference in our map compared with previous work.

Twenty eight QTLs were detected in this study. The most important milling trait is head rice (Tan et al. 2000). Five QTLs were detected for this trait with percent variation ranging from

7-54 percent. It is worthy to note that the same major QTL was detected for two important traits that determine cooking and eating qualities of rice, amylose content and alkali spreading score. Segregation for these traits was controlled by a major gene on chromosome 6, which probably exhibited pleiotropic effects.

2.4.2 Breeding Implications

The present investigation has important implications for rice breeding. The first concerns the inheritance of amylose content and gelatinization temperature represented by alkaline spreading score. These two components are the major determinants of rice eating quality. We observed that the two traits appear to be tightly linked together; it should be possible to simultaneously improve these two in a breeding program. In this work, amylose content, alkaline spreading score and protein content were mapped to the same chromosomal regions as earlier reported by others.

Secondly, interspecific hybridization of the Asian and African rice can be used to improve the milling, cooking and eating qualities of rice. This has been shown by the transgressive segregation found in most of the traits. The undesirable characteristics usually observed in *O. glaberrima* such as grain shattering, lodging and prolonged heading which may cause some of the progenies to be unfit (due to linkage drag) can be minimized during introgression of QTLs by marker assisted backcrossing (Kearsey and Farquhar 1998). This will help break unfavorable correlations between quantitative characters of interest (Verhaegen et al. 1997). In view of the significant epistatic effects reported here, the assessment of QTLs across genetic backgrounds would be important for elucidation of the distribution of genetic variation in this population. The QTLs that were reported here, especially *hr3*, *amy6* and *alk6-1* which have been consistent in other studies would be most useful in marker assisted selection (MAS) and are also very interesting candidates for positional cloning.

Oryza glaberrima (IRGC 103544) alleles were associated with increase in percent rice bran for two of the four QTLs *rb4* and *rb7* detected for this trait. There is no record of QTLs controlling percent rice bran in the literature for comparison with these detected QTLs. However, this result suggests that *O. glaberrima* introgressions at these loci may be advantageous for improving the oil content of rice since most of the oils are found in the bran. This is a promising research area and could be investigated further especially to determine if these QTLs would be stable in another environment or another background. Alleles from IRGC 103544 were also associated with increase in alkali spreading score, a measure of gelatinization temperature. This QTL *alk 6-1* has also been reported by Zhou et al. (2003).

In spite of the phenotypic inferiority of *O. glaberrima* and low yield potential, transgressive segregants were produced for most of the grain quality traits. This interspecific cross offers good opportunity for selection of lines that out-perform the Caiapo parent. This study was carried out in one location and one year, but the data are generally consistent with previous QTL studies while at the same time, potential new QTLs for grain quality traits were identified both from Caiapo and IRGC 103544 (Table 2.4).

2.5 References

AACC (1995) Approved methods of the American Association of Cereal Chemists, 9th edn. Method 46–16 approved December 1988, reviewed October 1994. The Association, St. Paul, Minnesota

Anonymous, EPAMIG (Empresa do Pesquisa Agropecuaria de Minas Gerais) (1994) Caiapo. Nova apcao de arroz de sequeiro, Belo Horizonte, Epamig, Embrapa

Anonymous, EMBRAPA (Empresa Brasileira do Pesquisa Agropecuaria) (1997) A producao de sementes no Brasil. Relatorio da Safra: 1995/96

Attere A, Fatokun C (1983) Reaction of *Oryza glaberrima* accessions to rice yellow mottle virus. Plant Dis 67:420-421

Bao JS, Sun M, Corke H (2002a) Analysis of genetic behavior of some starch properties in indica rice (*Oryza sativa* L.): thermal properties, gel texture, swelling volume. *Theor Appl Genet* 104:408-413

Bao JS, Sun M, Corke H (2002b) Genetic diversity of starch physiochemical properties in waxy rice (*Oryza sativa* L.). *Cereal Chem* (in press)

Basten CJ, Weir BS, Zeng ZB (1994) Zmap- a QTL cartographer. *Computing Strategies and Software. Proc 5th Congr on Genetics Applied to Livestock Production: Guelph, Ontario*

Basten CJ, Weir BS, Zeng ZB (1994) Zmap- a QTL cartographer: a reference manual and tutorial for QTL mapping. Department of Statistics, North Carolina State University, Raleigh, North Carolina

Bideaux JM (1978) Screening for horizontal resistance to rice blast (*Piricularia oryzae*) in Africa. In Buddenhagen, I.W. and Persley, G.J. (eds) *Rice in Africa*. London, U.K: Academic Press

Bollich CN, Webb BD (1973) Inheritance of amylose in two hybrid populations of rice. *Cereal Chem* 50:631-636

Carpenter, AJ (1978) Rice History: In Buddenhagen, I.W. and Persley, G.J. (eds) *Rice in Africa*. London, U.K: Academic Press

Causse MA, Fulton TM, Cho YG, Ahn SN, Chunwongse 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 interspecific backcross population. *Genetics* 138:1251-1274

Chase K, Adler FR, Lark KG (1997) EPISTAT: a computer program for identifying and testing interactions between pairs of quantitative trait loci. *Theor Appl Genet* 94:724-730

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

De Kochko, A (1987) Isozymic variability of traditional rice, *Oryza sativa* L. in Africa. *Theor Appl Genet* 73:675-82

De Vicente MC and Tanksley SD (1993). QTL analysis of transgressive segregation in an interspecific tomato cross. *Genetics* 134:585-596.

Dingkuhn, M, David E. Johnson, Monty P Jones, and A Sow (1996) The Physiological basis of developing low- management upland rice plant types. *Interspecific Hybridization: Progress and prospects*. 81-102

FAO (2002) Food and Agricultural Organization Annual Report 2002

- Garcia-Dorado A, Gallego A (1992) On the use of the classical tests for detecting linkage. *J. Hered* 83:143-146
- Ghesquiere A, Sequier J, Second G, Lorieux M (1997) First steps towards a rational use of African rice, *Oryza glaberrima*, in rice breeding through a 'contig line' concept. *Euphytica* 96:31-39
- Gosh AK, S Govindaswamy (1972) Inheritance of starch-iodine-blue value and alkali digestion value in rice and their genetic association. II *Riso* 21:123-132
- Grant V (1975) *Genetics of flowering plants*. Columbia University Press, New York
- Hamaker BR, Griffin VK (1990) Changing the viscoelastic properties of cooked rice through protein disruption. *Cereal Chem* 67:261-264
- Hamaker BR, Griffin VK (1991) Potential influence of a starch granule-associated protein on cooked rice stickiness. *J Food Sci* 56:1327-1329
- Hamaker BR (1994) The influence of rice protein on rice quality. In: Marshall WE, Wadsworth JI (eds) *Rice science and technology*, Marcel Dekker, New York, New York, pp 177-194
- Harushima Y, Kurata N, Yano N, Nagamura Y, Sasaki T, Minobe Y, Nakagahra Y (1995) Detection of segregation distortions in an indica-japonica rice cross using a high-resolution molecular map. *Theor Appl Genet* 92:145-150
- Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiyama H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T (1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics* 148:479-494
- He P, Li SG, Qian Q, Ma YQ, Li JZ, Wang WM, Chen Y, Zhu LH (1999) Genetic analysis of rice grain quality. *Theor Appl Genet* 98:502-508
- Helentjaris T, Slocum M, Wright S, Schaefer A, Nienhuis J (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. *Theor Appl Genet* 72:761-769
- Heuer S and K. Miezán (2003) Assessing hybrid sterility in *Oryza glaberrima* x *O. sativa* hybrid progenies by PCR marker analysis and crossing with wide compatibility varieties. *Theor Appl Genet Springer-Verlag* 2003 10.1007/s00122-003-1325-x
- Hua JP, Xing YZ, Xu CG, Sun LX, Yu SB, Zhang Qifa (2002) Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantages for performance. *Genetics* 162:1885-1895
- International Rice Research Institute (1976) *Annu Rep 1975*. Los Banos, Philippines, pp 85-86

- IRRI (1983) Annual report for 1982. International Rice Research Institute. Los Baños, Laguna, The Philippines
 Juliano BO (1985) Rice chemistry and technology, 2nd ed. American Association of Cereal Chemists, St. Paul, Minnesota
- Iwata T, Nagamatsu T, Omura T (1964) Abnormal segregation of waxy and apiculus coloration by a gametophytic gene belonging to the first linkage group in rice. *Jpn J Breed* 14:33-39
- Jacquot, M. (1977) IRAT and rice genetic resources. Paper presented at 'Rice genetic conservation Workshop', IRRI / IBPGR, December 1977. (unpubl.)
- Jones, MP, Dingkuhn, M., Aluko, GK. and Semon Mande (1997) Interspecific *O. sativa* L. x *O. glaberrima* Steud. progenies in upland rice improvement. *Euphytica* 92:237-246.
- Juliano BO (1985) Criteria and test for rice grain quality. In: Juliano BO (ed) Rice chemistry and technology. American Association of Cereal Chemists, Inc. St. Paul, Minnesota, USA, pp 443-513
- Juliano BO, Perez CM, Kaosa-Ard M (1990) Grain quality characteristics of export rices in selected markets. *Cereal Chem* 67:192-197
- Juliano BO (1993) Rice in human nutrition. FAO Food Nutrition Series No. 26. International Rice Research Institute: Manila, The Philippines
- Juliano BO, Villareal CP (1993) Grain quality evaluation of world rices. International Rice Research Institute, Manila, The Philippines
- Kumar I, Khush GS (1988) Inheritance of amylose content in rice (*Oryza sativa* L.). *Euphytica* 38: 261-269
- Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shomura A, Shimizu T, Lin S-Y, 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 (1994) A 300-kilobase interval genetic map of rice including 883 expressed sequences. *Nature Genet* 8:365-372
- Lanceras JC, Zue-Liu Huang, Onanong Naivikul, Apichart Vanavichit, Vinitchan Ruanjaichon, Somvong Tragoonrung (2000) Mapping of genes for cooking and eating qualities in Thai Jasmine rice (KDML105) *DNA Research* 7:93-101
- Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185-199
- Lander ES, Green P, Abrahamson J, Barlow A, Daley M (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174-181

- Lark KG, Chase K, Alder F, Mansur LM, Orf J (1995) Interactions between quantitative trait loci in which trait variation at one locus is conditional upon a specific allele at another. *Proc Natl Acad Sci USA* 92:4656–4660
- Lentini Z, P Reyes, CP Martinez, WM Roca.(1995). Androgenesis of highly recalcitrant rice genotypes with maltose and silver nitrate. *Plant Science* 110:127-138.
- Lefebvre V, Palloix A, Caranta C, Pochard E (1995) Construction of an intraspecific integrated linkage map of pepper using molecular markers and doubled haploid progenies. *Genome* 38:112-121
- Lin SY, H. Ikehashi, S. Yanagihara, A. Kawashima (1992) Segregation distortion via male gametes in hybrids between Indica and Japonica or wide-compatibility varieties of rice (*Oryza sativa* L.). *Theor Appl Genet* 84:812-818
- Linares, OF (2002). African rice (*Oryza glaberrima*): history and future potential. *Proc Natl acad Sci USA* 99: 16360-16365.
- Lincoln S, Daly M, Lander E (1992) Constructing genetic linkage maps with MAPMAKER/EXP. Whitehead Institute Technical Report. <http://www-genome.wi.mit.edu/ftp/distribution/software>
- Little RR, GB Hilder, and EH Dawson (1958) Differential effect of dilute alkali on 25 varieties of milled white rice. *Cereal Chem.* 35: 111-126
- Liu CJ, Witcombe JR, Pittaway TS, Nash CT, Busso CS, Gale MD (1994) An RFLP-based genetic map of pearl millet (*Pennisetum glaucum*). *Theor Appl Genet* 89:481-487
- Liu KD, Wang J, Li HB, Xu CG, Liu XH, Zhang Q (1997) A genome-wide analysis of wide analysis of wide sompatibility in rice and the precise location of the S₅ locus in the molecular map. *Theor Appl Genet* 95:809-814
- Lorieux M, Goffinet B, Perrier X, Gonzalez de Leon D, Lanaud C (1995) Maximum-likelihood models for mapping genetic markers showing segregation distortion. 1. Backcross populations. *Theor Appl Genet* 90:73-80
- Lorieux M, Petrov M, Huang N, Guiderdoni E, Ghesquiere A (1996) Aroma in rice: genetic analysis of a quantitative trait. *Theor Appl Genet* 93:1145-1151
- Lorieux M, M-N. Ndjiondjop and A. Ghesquiere (2000) A first interspecific *O. sativa* x *O. glaberrima* microsatellite genetic linkage map . *Theor Appl genet* 100:593-601
- Lu H, J. Romero-Severson, R. Bernardo (2002) Chromosomal regions associated with segregation distortion in maize. *Theor Appl Genet* 105: 622-628
- Lyttle T W (1991) Segregation distorters. *Ann Rev Genet* 25:511-557

- Maekawa M, Kinoshita T, Takahashi M (1981) Genetical studies on the rice plant. LXXVI. A new gametophytic gene in the second linkage group of rice Jpn Fac Agric Hokkaido Univ 60:107-114
- Marshall WE, Normand FL, Goynes WR (1990) Effects of lipid and protein removal on starch gelatinization in whole-grain milled rice. Cereal Chem 67:458-463
- McCouch SR, Kochert G, Yu Z H, Wang Z Y, Khush G S (1988) Molecular mapping of rice chromosomes Theor Appl Genet 76:815-829
- 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
- McCouch SR, Cho YG, Yano M, Paul, Blinstrub M (1997) Report on QTL nomenclature. Rice Genet Newsl 14: 11-13
- McKenzie KS, Rutger JN (1983) Genetic analysis of amylose content, alkali spreading score, and grain dimensions in rice. Crop Sci 23:306-311
- Mo HD (1993) Quality improvement of rice quality in China. Sci Agri Sin 26:8-14
- Nakamura N, Yamanouchi H (1992) Nucleotide sequence of a DNA encoding the starch branching enzyme, or Q-enzyme I, from rice endosperm. Plant Physiol 99:1265-1266
- Moncada P, CP. Martinez, J. Borrero, M. Chatel, H. Gauch Jr., E. Guimaraes, J. Tohme, S.R. McCouch. (2001) Quantitative trait loci for yield and yield components in an *Oryza sativa* x *Oryza rufipogon* BC₂F₂ population evaluated in an upland environment. Theor Appl. Genet. (2001) 102:41-52
- Nagamura Y, Antonio BA, Sasaki T (1997) rice molecular genetic map using RFLPs and its applications. Plant mol Biol. 35:78-87
- Nakagahra M (1972) Genetic mechanism on the distorted segregation of marker genes belonging to the eleventh linkage group in cultivated rice. Jpn J Breed 22:232-238
- Nakagahra M, Omura T, Iwata N (1972) Gametophyte genes and their loci on the eleventh linkage group of cultivated rice. Jpn J Breed 22:305-312
- Okuno K, Fuwa H, Yano M (1983) A new mutant gene lowering amylose content in endosperm starch of rice. Jpn J Breed 33:387-394
- Panaud O, X Chen, SR McCouch (1996) Development of microsatellite markers and characterization of sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.) Mol Gen Genet 252: 597-607

- Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335:721–726
- Pingali PL, Hossain M, Pandey S, et al. (1998) Economics of nutrient management in Asian rice systems: Towards increasing knowledge intensity *Field Crops Res* 56:157-176
- Pooni HS, Kumar I, Khush GS (1993) Genetical control of amylose content in selected crosses of indica rice. *Heredity* 70:269–280
- Porteres R (1950) Vieilles agricultures de l’Afrique intertropicale. Centres d’origine et de diversification varietale primaire et berceaux de l’Agriculture anterieure au XVIe siecle. *Agron Trop* 5:489-507
- Redona and Mackill (1998) Quantitative trait locus analysis for rice panicle and grain characteristics. *Theor Appl Genet* 96 (6-7): 957-963
- Reversat G, Destombes D (1998) Screening for resistance to *Heterodera sacchari* in the two cultivated rice species, *Oryza sativa* and *O. glaberrima*. *Fund and Appl Nematol* 21:307-317
- Sandler L, Novitski E (1957) Meiotic drive as an evolutionary force. *Amer Naturalist* 41:105-110
- Sano Y (1984) Differential regulation of waxy gene expression in rice endosperm. *Theor Appl Genet* 68:467–473
- Sano Y, Sano R, Morishima H (1984) Neighbor effects between two occurring rice species, *Oryza sativa* and *Oryza glaberrima* *J Appl Ecol* 21: 245-254
- Second G (1982) Origin of the genetic diversity of cultivated rice (*oryza* spp.): study of the polymorphism scored at 40 isozyme loci. *Jap J Genet* 57: 25-58
- Second, G (1984) Relations ivolutives cheese le genre *oryza* et processus de domestication des riz. Etude et theses, ORSTOM, Paris France.
- Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *The Plant Journal* 3:739-744.
- Takeoka, T (1965) Taxonomy and chromosome numbers of African representatives of *Oryza officinalis* complex. *Bot Mag. (Tokyo)* 78:198-201.
- Tan YF, Li JX, Yu SB, Xing YZ, Xu CG, Zhang QF (1999) The three important traits for cooking and eating qualities of rice grain are controlled by a single locus. *Theor Appl Genet* 99:642–648
- Tan YF, Xing YZ, Li JX, Yu SB, Xu CG, Zhang QF (2000) Genetic bases of appearance quality of rice grains in Shanyou 63, an elite rice hybrid. *Theor Appl Genet* 101:823–829

- Tan YF, M. Sun, YZ Xing, JP Hua, XL Sun, QF Zhang, H. Corke (2001) Mapping quantitative trait loci for milling quality, protein content and color characteristics of rice using a recombinant inbred line population derived from an elite rice hybrid. *Theor Appl Genet* 103:1037-1045
- 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: 698-712
- Unnevehr LJ, Duff B, Juliano BO (1992) Consumer demand for rice grain quality. International Rice Research Institute, Manila, The Philippines, and International Development Research Center, Ottawa, Canada
- Virk, PS, Ford-Lloyd BV, Jackson MT, Newbury, HJ (1995) Use of RAPD for the study of diversity within plant germplasm collections. *Heredity* 74: 170-179
- WARDA 1997 Annual report of the West African Rice Development Association, Bouake, Cote d'Ivoire
- Webb BD (1980) Rice quality and grades. In Luh BS (ed) *Rice: production and utilization*. Avi Publication Company, Incorporated, Westport, Connecticut, USA, pp 543–565
- Xiao J, Li J, Yuan L, Tanksley SD (1996) Identification of QTL affecting traits of agronomic importance in a recombinant inbred population derived from sub-specific rice cross *Theor Appl Genet* 92:230-244
- 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
- Xu GG, Yu SB, Zhang Q, Li JX, Xing YZ, Tan YF (2000) Genetic bases of appearance quality of rice grains in Shanyou 63, an elite hybrid. *Theor Appl Genet* 101: 823-829
- XU GW, Magill CW, Schertz KF, Hart GE (1994) A RFLP linkage map of *Sorghum bicolor* (L) Moench. *Theor Appl Genet* 89:139-145
- Xu Y, Shen ZT (1992) Distorted segregation of waxy gene and its characterization in indica-japonica hybrids. *Chinese J. Rice Sci* 6:89-92
- Xu Y, Shen ZT, Chen Y, Zhu LH (1995) Distorted segregations of RFLP markers and their distribution on chromosomes in an indica-japonica F₂ population of rice (*Oryza sativa* L.). *Acta Bot Sin* 37:91-96
- Xu Y, Zhu L, Xiao J, Huang N, McCouch S (1997) Chromosomal regions associated with segregation distortion of molecular markers in F₂, backcross, doubled haploid, and recombinant inbred populations in rice (*oryza sativa* L.). *Mol Gen Genet* 253:535-545

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

Yu SB, Li JX, XU CG, Tan YF, Gao YJ, Li XH, Zhang QF, Saghai Maroof MA (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid *Proc Natl Acad Sci USA* 90:9226-9231

Zeng ZB (1993) Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci *Proc Natl Acad Sci USA* 90:10972-10976

Zeng ZB (1994) Precision mapping of quantitative trait loci. *Genetics* 136:1457–1468

Zhou PH, YF Tan, YQ He, CG Xu, Q. Zhang (2003) Simultaneous improvement for four quality traits of Zhenshan 97, an elite parent of hybrid rice, by molecular marker-assisted selection. *Theor Appl Genet* 106:326-331

Zivy M, Devaux P, Blaisonneaux J, Jean R, Thiellement H (1992) Segregation distortion and linkage studies in microspore-derived double haploid lines of *Hordeum vulgare* L. *Theor Appl Genet* 83:919-924

CHAPTER 3: GENETIC MAPPING OF AGRONOMIC TRAITS FROM THE INTERSPECIFIC CROSS *ORYZA SATIVA* (L.) AND *ORYZA GLABERRIMA* (STEUD.) USING TRADITIONAL AND NON-PARAMETRIC METHODS

3.1 Introduction

Most agronomic traits of interest to plant breeders show continuous variation, and the effects of individual genes controlling such traits cannot be readily identified. Quantitative traits are controlled by many genes and their expressions are modified by the environment. Consequently, it has been challenging to enhance these traits using traditional breeding methodology that relies primarily on phenotypic selection criteria. The relationship between the genotype of a crop (genetic constitution) and the trait values observed in the field (phenotype) is not directly estimable (Falconer and Mackay, 1996), making selection difficult for quantitative traits. An added level of complexity is introduced when genotypes perform differently in different environments via genotype x environment interactions. Breeders are compelled to evaluate genotypes using large replicated trials and multiple environments which further lengthen the breeding process.

Recent advancements in rice breeding have led to the search for molecular markers associated with the various traits of interest with a view to improving the efficiency of breeding for such traits. Indirect selection using molecular markers linked to genes of interest holds promise as a means to increase the efficiency of selection for quantitative traits. Various markers have been used recently to identify loci within the rice genome affecting agronomic traits (Moncada et al., 2001; Yu et al., 2002).

3.1.1 Molecular Markers and Their Use in Crop Breeding

The basic theory of using genetic markers to manipulate loci controlling plant traits was introduced by Sax in 1923 (Mazur and Tingey, 1995) who reported the association of

quantitatively inherited seed size with simply inherited genetic markers that governed seed coat pigmentation and pattern in common bean (*Phaseolus vulgaris* L.). Subsequent reports of linkage between single gene markers and quantitative trait loci (QTLs) used morphological mutations as genetic markers, the nature of which posed major limitations for the study of quantitative variation (Rasmusson 1953; Everson and Schaller 1955; Thoday 1961). In these studies, only a few such markers were available in any given cross, and the effect of marker genes on quantitative traits was often larger than that of the linked QTLs, making it difficult to effectively study quantitatively inherited traits extensively (Tanksley et al. 1998). Application of these association studies was therefore limited by the lack of available segregating genetic markers.

Recent advances have produced segregating genetic markers in many crop species including rice. Earlier studies on linkage of simply inherited genes with quantitative traits provided the insight into the potential of markers. However, the slow progress made in the use of morphological markers was due to limited number of markers available and the undesirable effects on phenotype of many of the morphological markers.

Later development led to the use of isozymes as genetic marker, which are multiple forms of enzymes arising from genetically determined differences in primary structure (Ostenhof et al. 1971). Isozymes were used successfully as markers to identify QTLs in maize (Stuber and Edwards 1986) and rice (Pham et al. 1990). However, the low number of markers available reduced the utility of isozymes as markers.

DNA markers have been used in a number of crops to determine the number of genes controlling trait inheritance and for gene tagging (Anderson et al. 1993; Ma et al. 1993; Schachermayer et al. 1994). The utility of DNA based markers is determined to a large extent by the technology that is used to reveal DNA polymorphism (Mazur and Tingey, 1995). The

available assays fall into two broad categories, *viz.* restriction enzyme based assays and DNA amplification based assays. The first category comprises restriction fragment length polymorphism (RFLP) which detects DNA polymorphism through restriction endonuclease digestion followed by visualization via DNA blot hybridizations. Many researchers still prefer to use RFLP markers because each polymorphic co-dominant allele at a locus is detected in the assay. Moreover, they have been found useful for detecting locus-specific polymorphisms across species boundaries (Liu et al., 1994). Restriction fragment length polymorphisms (RFLPs) have been useful for constructing saturated genome maps in rice (Panaud 1992; Causse et al. 1994). However, these markers require relatively large amounts of DNA for the assay, are time consuming, labor intensive, and therefore only low-resolution maps have been developed (O'Brien, 1993; Liu et al., 1994).

Different types of DNA amplification-based markers have been developed over the years for use in rice studies, including Randomly Amplified Polymorphic DNA (RAPD) (Virk et al. 1995), Amplified Fragment Length Polymorphism (AFLP) (Virk et al. 2000) and microsatellites (McCouch et al. 1997). The discovery that short primers (usually 10-mers) of an arbitrary nucleotide sequence could be used to amplify segments of genomic DNA from a wide variety of species led to the application of RAPD technology (Williams et al. 1990). RAPD markers are faster and easier to assay than RFLPs, and only a very small amount of DNA template is required for RAPD assays. Nevertheless, the major limitation of RAPD is that it shows dominant inheritance and marker/marker homozygotes cannot be distinguished from marker/null heterozygotes.

AFLP has been recognized as a reliable and efficient DNA marker system compared to RFLP and RAPD methods (Vos et al. 1995; Powell et al. 1996; Russell et al. 1997). AFLP markers have been found useful in the study genetic diversity of different plant species due to

high reproducibility and throughput (Maughan et al. 1996; Ellis et al. 1997; Erschadi et al. 2000).

The AFLP technology is not species specific and so can be used as an alternative to microsatellites markers. Moreover, AFLP has a high number of polymorphic marker fragments amplified by PCR in one reaction, resulting in a sufficient genome coverage using only a small number of primer combinations. However, AFLP markers are dominant, and this is a major drawback for genetic analysis, unless a high number of loci are investigated.

Microsatellite markers, also known as simple sequence repeats (SSR), are short, tandemly-repeated DNA sequences that are very abundant in the genome of eukaryotes including rice. SSR markers are codominant, require only a small DNA template, and are technically simple to assay (Morgante and Olivieri 1993; Wang et al. 1994; McCouch et al., 1997).

Microsatellites are evaluated using PCR primers targeted to unique sequences flanking a microsatellite motif. The resulting PCR products are separated according to size by gel electrophoresis using either agarose or acrylamide gels or more recently by capillary electrophoresis systems. Because microsatellites are randomly distributed throughout the rice genome, saturated genome coverage would be possible using these markers. A major limitation of the microsatellite marker is the tendency for physical breakage within the extended repeat motif which may lead to smaller length of some motifs (Yang et al. 1994). A positive relationship between the length of microsatellite motif and allelic diversity in rice can occur, so small sized, broken motifs would be of limited utility in the study of allelic diversity. Moreover, microsatellites are species specific.

The use of molecular markers to enhance rice breeding efforts has great potential due to the availability of different types of polymorphic markers, genetic maps and on-line database resources (examples: Ricegenes (<http://stein.cshl.org/ricegenes.html>) and Gramene (<http://www.gramene.org>)). Molecular markers may be effectively utilized in a breeding

program depending on the value, ease and cost of measurement and, the nature of the genetic control of the quantitative trait. Various markers have been developed and used recently to identify loci within the rice genome affecting agronomic traits (Paterson et al. 1988; Paterson et al. 1996; Moncada et al., 2001; Yu et al., 2002; Hua et al. 2002). Recent advances in molecular marker technology and development of high-density linkage maps in rice (Causse et al. 1994; Harushima et al. 1998; Chen et al. 1997) have provided a powerful tool for elucidating the genetic basis of several agriculturally important traits, many of which are quantitatively inherited. However, results from these different studies have not been routinely applied to public breeding programs, due to incomplete agreement on the location and utility of putative QTLs.

Moreover, reported QTLs have not been adequately tested under different environments, so a more comprehensive combination of useful markers and genetic approaches are needed to accurately and predictably identify quantitative traits. Non-parametric methods, discussed below, do not strictly demand normal distribution of variables, and have been developed to analyze both quantitative and categorical data (Gnanadesikan, 1987; Hand, 1997). These methods and the advances in computer technology leading to the availability of fast and very powerful computers have increased the opportunities for application of multivariate classification techniques.

3.1.2 Bulk Segregant Analysis

A bulk segregant analysis (BSA) procedure also known as trait-based analysis was first introduced by Lebowitz et al. (1987). The feasibility of BSA to identify markers linked to any specific gene or genomic region has been demonstrated on several occasions (Michelmore et al. 1991; Giovannoni et al. 1991; Wang and Paterson 1994). Based on the assumption that markers adjacent to the targeted gene would be in linkage disequilibrium, BSA has been used to rapidly identify markers linked to QTLs in crop plants. To perform BSA, a population resulting from a

cross that segregates for the gene of interest is required. Plants from such segregating population are grouped according to phenotypic expression of a trait and tested for differences in allele frequency between the population bulks. Quarrie et al. (1999) observed that when QTL has a large effect, a molecular marker which is closely linked to that major QTL (regulating a particular trait) will segregate according to the phenotype of the trait. These researchers used BSA to locate QTLs associated with the yield of maize under severe drought conditions.

Because bulks are formed from the two extremes of the distribution of a trait, the reliability of the BSA strategy may be affected by the size of the bulk. As the bulks utilized become smaller, there is a tendency for the false positives to increase. However, Michelmore et al. (1991) stated that BSA with only small numbers of individuals in one or both bulks would provide great enrichment for markers to the targeted loci. Observations from Michelmore et al. (1991) have shown that the minimum size of the bulk would be determined by the frequency with which unlinked loci might be detected as polymorphic between the bulked samples. These researchers concluded that this frequency depends on the type of marker and population used to generate the bulk.

Bulk segregant analysis can be readily applied where a genetic map approaches saturation and the continued mapping of polymorphisms detected by arbitrarily selected markers becomes less efficient (Bishop et al. 1983). Moreover, if there are regions of interest in the genome that are sparsely populated with markers, BSA can be used to focus on such regions. Because of the robustness of BSA, it can be used to construct a high-resolution map. In a study by Michelmore et al. (1991), RAPD and RFLP markers were used to test the sensitivity of BSA to determine the genetic distance at which there would be sufficient recombinants to cause BSA to become monomorphic. They found that both RFLP and RAPD were equally sensitive to detect markers

closer than 15 cM from the target allele, but the limit of detection was found to be approximately 25 cM. BSA has been used widely to map different disease resistance genes (Michelmore et al. 1991; Dong et al. 2000; Liu et al. 2002; Mammadov et al. 2003).

3.1.3. Discriminant Analysis as a Tool for Marker Assisted Classification

Discriminant analysis (DA) is a multivariate non-parametric statistical procedure that can be used to generate functions from a sample of observations for which group membership is known. Functions can then be applied to new individuals with a view to assign them to predefined groups. Fahima et al. (2002) recently used DA to correctly classify 88 percent of 135 wild emmer wheat populations into sites of origin based on microsatellite profiles.

The focus of DA is concerned with the efficient separation, or discrimination, of two or more groups of cases (individuals) based on the observed attributes of the cases. With the two-group case, the usual approach is to develop a rule known as the discriminant function or algorithm from a combination of the phenotypic and molecular profiles of the training data set. The value of this function is then calculated and used to classify cases as belonging to one of the two groups.

In assessing the effectiveness of discriminant functions for predicting group membership, a procedure known as cross-validation is often used. In this procedure the probability of misclassification error in the allocation of n lines into predefined groups is estimated using $n-1$ lines to predict allocation of the one left out. This hold-out procedure, developed by Hand (1967), starts with “ x ” group of observations. Let us assume two groups “ n ” and “ n_2 ”, where one observation is omitted from the “ n ” group and a classification function is developed based on the remaining observations. The observation left out is then classified using the function constructed earlier. This procedure is repeated until all of the “ x ” observations have been classified. The

number of misclassified cases is then noted. This process is repeated for all the materials in the training sample as well as the test data comprising the lines to be classified. Estimates of the conditional misclassification probabilities are used for calculation of the actual error rate. The procedure of cross-validation is offered as an option in readily available computer programs such as SAS version 8 using the “crossvalidate testlist” option in the “Proc Discrim” command.

DA has many applications in both basic and applied research. Ebdon et al. (1998) evaluated DA in the classification of Kentucky bluegrass into groups of low and high water use cultivars. These researchers identified two and three-variable discriminant rules that correctly classified cases into their true water use groups with 75 percent correct classification. The authors concluded that DA might be an efficient and useful tool for predicting the water use patterns of new cultivars on the basis of a few plant measurements that are routinely assessed by breeders. Recently, Barrella and Petrere (2003) used DA to identify parameters of water (dissolved oxygen, conductivity and nitrogen) influencing fish in two Brazilian rivers with different levels of pollution. DA has also been used to establish multivariate quality controls in the production of pharmaceutical proteins (Bylund et al. 2003). In the field of human medicine, DA was used as a statistical tool to discover genes differentially expressed between malignant oral epithelium and normal tissues in micro-array experiments and to construct a robust classifier using the identified discriminatory genes (Hwang et al. 2003). Soil scientists have used DA to separate humic acids extracted from peats, brown coals and lignites, and to make predictions on samples of unknown origin (Francisco et al. 2003). All these studies demonstrate the relevance, robustness, and application of DA in contemporary scientific research. Nevertheless, the DA approach has not been extensively used in plant genetic mapping and linkage studies. The application of DA as a tool for allocating rice lines into groups was proposed by Balzarini et al. (2000), and its application to identify rice germplasm with contrasting phenotypes was reported

by Capdevielle et al. (2000). These preliminary studies indicated the potential of DA as a useful tool for rice breeding and indicated the need for further evaluation of the procedure.

3.1.4. Multiple Regression

Multiple regression is a statistical procedure that has been used to explore associations between molecular markers and quantitative traits in rice. Using RAPD markers as predictor variables, Virk et al. (1996) identified markers that showed strong associations with QTLs. Furthermore, selected markers were used for predicting the phenotypic performance of other accessions. The assumption was made of a linear relationship between the markers and the quantitative trait of interest. One of the drawbacks in using multiple regression for selecting markers is the possibility of correlation among markers used as regressor variables. This multicollinearity is assessed by the variance inflation factor (VIF) (Neter et al. 1996). Semagn et al. (2000) found severe multicollinearity among selected variables when two or more variables were selected by the multiple regression model.

3.1.5. Analysis of Molecular Variance (AMOVA)

Analysis of molecular variance (Excoffier et al. 1992) is a statistical method for studying molecular variation within a species. The software ARLEQUIN (<http://lgb.unige.ch/arlequin/>), developed by Laurent Excoffier, performs the AMOVA procedure using RFLP, DNA sequences, SSR, and standard multilocus frequency data (Cottrel and White, 1996). AMOVA is a hierarchical analysis of variance procedure that separates and tests tiers of genetic diversity to elucidate variation among groups of populations, diversity among populations within groups and diversity among individuals within a population. Hypotheses are tested using permutational analysis of variance, so that the assumption of normal distribution is not required, resulting in robust testing for molecular variance. Input data are used to create a distance matrix between

samples to measure genetic structure of the population from which the samples are drawn. Huff et al. (1993) determined the pattern and extent of RAPD marker variation within and among buffalograss populations using AMOVA. Their results demonstrated the power of AMOVA in separating important and interesting regional and population differences against a background of extensive within-population polymorphisms.

3.1.6. Association Mapping and Linkage Disequilibrium

Linkage disequilibrium (LD), a concept that describes the non-random association of alleles across two or more loci, has been recently exploited as a useful tool for map-based association studies in humans (Boehnke 1994; Bennet et al. 1995; Chagnon et al. 1998). Due to the complexity of common human diseases and difficulties in family-based analysis, traditional genetic approaches generally used in plants are not appropriate. Instead, most human geneticists favor association studies in which genetic and phenotypic variation are compared in large population samples to identify correlations implicating genetic risk factors (Briscoe et al. 1994; Hsu et al. 2003; Clark 2003).

Human LD studies are usually designed within a “case-control” arrangement in which a group of sufferers from a condition is compared to an unrelated group of healthy individuals (Blum et al. 1990; Holden 1994). The extent of LD is affected by genetic sub-structure within a population, natural selection, as well as genetic drift due to the finite size of a population (Chakraborty et al. 1992). Linkage disequilibrium can also result from mutation, population admixture (i.e. gene flow between two or more genetically distinct populations) and population bottlenecks (Deng et al. 2000).

Although association genetics has been used to map genes underlying complex diseases, the results of independent studies are often inconsistent. Conflicting results were reported in the association between the dopamine D2 receptor gene and alcoholism (Blum et al. 1990; Carpenter

et al. 1993; Pato et al. 1993) and vitamin D receptor genotypes and bone mass (Morrison et al. 1994; Eisman 1995; Gong et al. 1999). These inconsistencies have been attributed to population admixture, (Chakraborty and Smouse 1988; Lander and Schork 1994; Deng and Chen 2000), often leading to spurious associations between the marker alleles and the complex traits. The presence of subgroups within the population (population structure) may lead to highly significant marker-phenotype associations, even when the marker is not physically associated to any causal loci (Spielman and Ewens 1996). The possibility of these spurious associations leading to false positives has so far restricted the utility of association genetics studies. Recently, Pritchard et al. (2000) developed statistical methods for standard association studies that use independent marker loci to detect the occurrence of cryptic population structure and to assign individuals to subpopulations that share similar marker genotype frequencies. The software program (STRUCTURE; <http://pritch.bsd.uchicago.edu>) was developed by Pritchard et al. (2000) to infer population structure using multilocus genotype data. With this program the entire sample is assumed to contain K populations each of which is characterized by a set of allele frequencies at each locus. Individuals in the sample are assigned to populations or jointly to two or more populations if their genotypes indicate that they are admixed. Adjustment for structure has been used to improve association mapping in human genetics (Pritchard et al. 2000b; Mott and Flint 2001).

Plant molecular biologists and breeders are beginning to consider association genetics as a potential tool for the improvement of crops. Recently, Thornsberry et al. (2001) conducted association studies based on a candidate gene in the analysis of flowering time and the *dwarf 8* (*d8*) gene in maize. In this study, variation in *d8* was evaluated for association with flowering time and plant height in 92 maize inbred lines. These researchers found nine polymorphisms including a MITE insertion in the promoter region of *d8* and a two-amino acid deletion adjacent

to the SH2-like domain, a potentially key binding domain in this putative transcription factor, associated with flowering time. Also very recently, Whitt et al. (2002) used association methods to map the mutation which occurred at the *sugary1* (*su1*) locus in maize to a single nucleotide, even though ~ 150 mutations were segregating within a 12 kb region of this locus.

The objectives of this study were: (1) to identify QTLs linked to economically important agronomic traits among DH lines derived from the interspecific cross *Oryza sativa* (L.) and the African rice *O. glaberrima* (Steud.) (2) identify potential molecular markers from *O. glaberrima* that contribute to agronomic traits measured in this study (3) determine the effect of population structure on power and precision to detect markers associated with the measured agronomic traits (4) evaluate DA for correct allocation of DH lines into pre-defined groups based on their microsatellite profiles (5) compare genetic map locations of DA-selected markers with those markers detected by traditional QTL and BSA approaches.

3.2 Materials and Methods

3.2.1. Plant Material

Caiapo, a tropical *Oryza japonica* commercial variety developed by the Brazilian national rice program for upland acid soil conditions (Anonymous, EMBRAPA 1997), was used as the recurrent parent in this study. This variety is characterized by good milling and eating characteristics, long grain type, early maturity, low tiller number, 80 cm height, and typical yields of 2.5 t/ha under upland conditions (Anonymous, EPAMIG 1994). *Oryza glaberrima*, IRGC accession # 103544, (Dr. Brar IRRI, pers. communication) originally collected in the wild from Mali, Africa is ~ 95 cm tall with resistance to several biotic and abiotic stresses including, drought, soil acidity, rice blast, and rice stripe necrosis virus. Excellent vegetative growth of *O. glaberrima* helps suppress weeds (Dingkuhn et al. 1996).

3.2.2. Population Development

Genetic material for this study was developed by Dr. Cesar Martinez, rice breeder at the Centro Internacional de Agricultura Tropical (CIAT) in Cali, Colombia. IRGC accession # 103544 served as the male parent in crosses to Caiapo (Fig.3.1). F₁ plants were grown in 1997 in the greenhouse, at CIAT, Colombia. A total of 200 F₁ seeds were produced. Since all F₁ plants were completely sterile 20 individuals were used as the female. A total of 154 BC₁F₁ plants were produced and transplanted under irrigated conditions to the field in 1998. All plants were sterile and 103 BC₁F₁ plants were selected and backcrossed to Caiapo to generate the BC₂ generation which was grown under irrigated field conditions. A high level of sterility in this material was observed. The final backcross to Caiapo was completed in 1999 that generated 97 BC₃F₁ seeds. BC₃F₁ plants were grown under irrigated conditions to ensure survival and good plant development. Anthers were collected from each BC₃F₁ plant and used in anther culture as described by Lentini et al. (1995).

A set of 312 DH lines representing the observed genetic variability was chosen for further agronomic and molecular characterization. Seed from each DH line was advanced one generation by selfing to produce sufficient seed for subsequent phenotypic and genetic analyses.

3.2.3. Field Experiments

The first field trial was conducted by Dr. Martinez at Palmira, Colombia in August 2001. The 312 DH lines were planted under irrigated conditions in a randomized complete block design, in two-row plots, five meters long with three replications. Twenty-five day old seedlings were transplanted at a spacing of 30 x 30 cm. with Caiapo and *O. glaberrima* (IRGC 103544) also included as controls. Fertilizer was applied at the rate of 120 kg N ha⁻¹, 73 kg K₂O ha⁻¹, 63 kg P₂O₅ ha⁻¹, and 4 kg ZnSO₄ ha⁻¹. A post emergence application of Butaclor and Bentazol each at the rate of 3 Lha⁻¹ was used to control weeds supplemented by manual weeding as needed.

Population Development

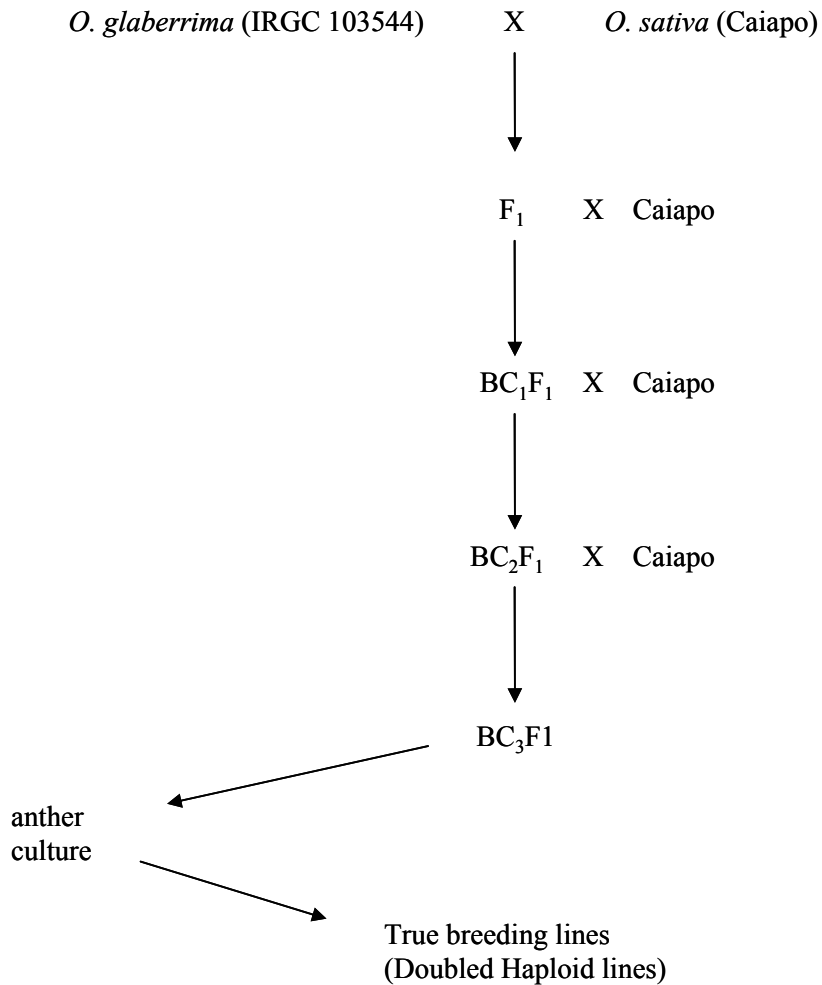


Fig 3.1 Schematic diagram of the development of 312 doubled haploid lines derived from *O. sativa* (Caiapo) x *O. glaberrima* (IRGC 103544) cross.

Data on agronomic traits including flowering, plant height, days to heading, tiller number, panicle sterility, and plot yield were taken. Experimental plots were harvested in December

2001. After drying to 15 percent moisture, 1000 grains were counted on per plot basis and weighed.

In April 2002, a second field experiment with the DH lines and the two parents was planted in a randomized complete block design with two replications at the Rice Research Station, Crowley, Louisiana. Seeds were planted mechanically using a Kincaid cone planter. Fertilizer (NPK 8-24-24) was applied at the rate of 275 kg ha⁻¹ at planting and top dressed with urea at 40 kg ha⁻¹. Thinning of seedlings was performed three weeks after planting to produce 20 cm within and 30 cm distances between rows. Clincher (Cyhalofop) herbicide was applied at 1 L ha⁻¹ 30 days after planting. Insects were controlled by spraying Karate at 0.15 L ha⁻¹ 30 days after planting. Irrigation was applied during prolonged dry periods that occurred in June and July. Harvesting was completed in October 2002. IRGC accession #103544 parent and ~ 40 percent of the DH lines did not flower at the Louisiana location, presumably due to photoperiod sensitivity. Therefore, a high percentage of missing data from this site precluded its use in the analysis.

3.2.4. Molecular Marker Analysis

The population of 312 DH lines was analyzed using a total of 100 microsatellite markers distributed randomly throughout the genome at an average distance of 10.5 cM. A total of 125 microsatellite markers were randomly chosen over the twelve linkage groups of the rice genome out of which 100 that were polymorphic between the parents were chosen. These 100 microsatellites were used to derive molecular profiles of the 312 DH lines, by Dr. Tohme, CIAT, Cali, Colombia.

3.2.5. Statistical Analysis, Map Construction and QTL Detection

Correlation among traits was evaluated using PROC CORR (SAS Institute 1998). QTL analyses associated with markers for each trait were performed using MAPMAKER 3.1 (Lander

et al. 1987; Lincoln et al. 1992). Linkage groups were created with a minimum LOD score of 3.0 and a recombination fraction of 0.4 using the “group” command. The marker order within the linkage groups was determined using the “compare”, “try” and “ripple” commands. Map distances were calculated by utilizing the Kosambi function. Due to distorted segregation detected on all chromosomes, the “MapDisto” software program (Lorieux 2000) was used to adjust for non-Mendelian inheritance of certain SSR markers. Data transformation for each trait was performed using inverse and log transformations to improve normality of the trait distributions (Excel 2002). QTLs were detected by interval and composite interval mapping procedures (Liu 1997; Zeng et al. 1994). A window size of 10 cM was used as a default for the QTL Cartographer used for mapping. The window blocks out a region of 5 cM on each side of markers flanking the target interval. Background markers were selected by a stepwise regression procedure. The inclusion of background markers in the analysis helps in one of two ways, depending on whether the background markers and the target interval are linked. If they are not linked, inclusion of the background marker makes the analysis more sensitive to the presence of a QTL in the target interval. If they are linked, inclusion of the background marker may help to separate the target QTL from other linked QTLs (Zeng 1993, 1994). Only QTLs detected both by interval and composite interval mapping procedures were used in the analysis.

3.2.6. Discriminant Analysis

Datasets for each trait were checked for normality using “Proc univariate” in SAS version 8.1 (SAS Institute, 2000). Deviations from normality were corrected by inverse and log transformations using Microsoft Excel 2002. Data from Colombia and Louisiana were analyzed separately. Groups were then defined for each phenotypic trait based on the degree of phenotypic differentiation of the two groups. Two groups were formed by taking materials at opposite tails of the distribution. These two groups were separated by 1 standard deviation (1SD)

i.e. 1 standard deviation between the high and low groups, designated as reduced differentiation, 2 standard deviations between the two groups (2SD), designated as intermediate and 3 standard deviations, the largest differentiation between the two groups (3SD). These groups constituted the “training samples” (Hand, 1997). Analysis of Molecular Variance (AMOVA; Excoffier et al. 1992; Arlequin software Schneider et al. 2000) was performed on microsatellite marker profiles between each group to test statistical differences for allelic frequencies. Groups that differed significantly were analyzed by stepwise discriminant analysis (Gnanadesikan 1987; Hand 1997) using the “stepdisc procedure” forward stepwise selection option in SAS (SAS Institute, 2000) to identify and select a sub-set of markers that revealed statistical differences between the pre-defined groups. Markers selected by Stepwise DA were then used in a classificatory, non-parametric, nearest-neighbor procedure to predict and allocate RIL lines into groups of low and high trait values. The DA analysis was performed using the DISCRIM procedure of SAS/STAT software version 8.1 (SAS Institute, 2000). The process was used to obtain discriminant rules for each of the six agronomic traits. The efficiency of the discriminant procedure was evaluated using the leave-one out method to determine the percent correct classification using the “crossvalidate testlist” option in the SAS “Proc Discrim” (version 8.1, SAS Institute, 2000).

The 312 DH lines were used for Interval Mapping (IM) and Composite Interval Mapping (CIM) using the “QTL Cartographer” software (Zeng 1994). A bulk segregant analysis (Wang and Paterson, 1994; Mitchelmore et al. 1991) was also performed to detect markers associated with these traits. In the comparison among these four mapping methods, the backward elimination option of DA was used to select informative markers.

For analysis of potential population structure among the RIL lines, the “Structure” program (STRUCTURE; <http://pritch.bsd.uchicago.edu>) was used to assign individual lines to

subpopulations inferred from allelic frequencies at each locus. The DA approach was compared in the allocation of lines into groups of high and low trait values corrected for population structure and alternatively when population structure was not assumed.

3.3 Results

3.3.1. Phenotypic Variation for the Different Agronomic Traits

The mean phenotypic values for each trait in the 312 DH lines are shown in Table 3.1. The Caiapo parent was ~33 percent taller than the IRGC 103544 parent. For this trait, the DH lines were skewed towards the Caiapo parent (Fig.3.1). This result is expected due to the three backcrosses made to the Caiapo parent. However, the DH lines were 3 cm shorter on average than the Caiapo parent. Transgressive segregation for tall plants among the DH lines was detected as 35 percent of these progenies were taller than Caiapo. Similar observations of transgressive segregation for plant height have been reported by other researchers (Moncada et al. 2001; Mei et al. 2003) in an interspecific cross between *O. rufipogon* and Caiapo. The continuous unimodal frequency distribution of plant height both at Crowley and Colombia suggested polygenic inheritance of this trait. At Crowley, the Caiapo parent was taller than IRGC 103544 by 11.4 percent. The mean height of the DH progenies was 23 cm taller than Caiapo. Forty-eight percent of the DH lines showed transgressive segregation for tall and only 1 percent exhibited segregation for short plant types

The distribution of days to heading of the DH population was bimodal at Colombia as shown in Figure 3.2. Although the range of heading dates among the DH lines was as wide as 28 days, the parents did differ by only two days (Table 3.1). On the average, the DH lines headed 6-8 days later than both parents and transgressive segregation for late heading among the DH lines was observed for 73 percent of the DH lines that headed later than the two parents.

IRGC 103544 produced 15 more tillers than the Caiapo parent (Table 3.1). The DH lines had an average of two tillers more than the Caiapo parent. This observation was consistent for both Louisiana and Colombia locations. *O. glaberrima* is known to have profuse tillering (Jones et al. 1997), and the reduced tillering of some of the progenies may have been due to suppression of tillering ability resulting from backcrosses made to the Caiapo parent. The IRGC 103544 parent was very photosensitive at Crowley and did not flower. However, the DH lines on average headed 18 days later than the Caiapo parent, and even though only 2 percent of the DH lines were earlier, 48 percent were later in flowering than the recurrent parent.

At Crowley, IRGC103544 produced 20 more tillers than Caiapo and a wide range of 16 tillers between the lowest and the highest tillering DH lines was observed. There was an improvement in number of tillers per plant among the DH lines as 83 percent of these lines produced from 10 to 21 tillers per plant (Fig.3.2d).

The Caiapo panicles were 5 cm longer than those of IRGC 103544, but similar to the average length of those of the DH lines. There was positive transgressive segregation for this trait among the DH lines and in 57 percent of the DH lines, panicle length ranged from 24.2 to 29.2 cm. The distribution of panicle length is shown in Table 1 and Fig. 3.2c.

At Crowley, the average panicle length of the DH progenies was 2 cm longer than the Caiapo parent. Since IRGC 103544 did not flower, data for panicle length was not recorded (Table 3.2). Grain yield of the Caiapo parent was more than five times greater than that of IRGC 103544 in Colombia (Table 3.1). The mean grain yield among the DH lines was more than twice as much as IRGC, but ~ half of the yield of Caiapo. Transgressive segregation for high grain

Table 3.1 Mean values for 6 agronomic traits of parents and 312 DH lines derived from *O. sativa* (Caiapo) x *O. glaberrima* (AC# IRGC103544) cross, Colombia, 2001

Parent/Line	PLHt	HD	Till/Plt	Panlen	Grain yield	TGW
Caiapo	120.0a	84.0a	9.0b	24.0a	2680.0a	29.0a
IRGC 103544	84.0b	82.0a	24.0a	18.7b	465.8b	19.8b
DH lines	117.0	90.0	11.0	25.0	1819.0	28.2
Range	84-143	81-109	7-24	19-29	466-2896	19.8-38
S.E	0.54	0.38	0.11	0.10	25.6	0.16

^a PLHt = Plant height, HD = Heading date, Till/Plt = Tillers per plant, Panlen =Panicle length
 TGW = 1000-grain weight, S.E = Standard error of the mean
 Means bearing the same alphabet do not differ significantly

Table 3.2 Mean values for 6 agronomic traits of parents and 158 DH lines derived from *O. sativa* (Caiapo) x *O. glaberrima* (AC# IRGC103544) cross, Crowley, LA, 2002

Parent/Line	PLHt	HD	Till/Plt	Panlen	Yield	TGW
Caiapo	80.0	85.0	6.0	21.0	1250.0	20.0
IRGC 103544	71.0	-	26.0	-	-	-
DH lines	103.0	103.0	8.0	23.0	684.0	22.0
Range	51-124	62-127	3-19	4-29	160-1520	5-33
S.E	0.92	0.73	0.23	0.56	21.7	0.35

^a PLHt = Plant height, HD = Heading date, Till/Plt = Tillers per plant,
 Panlen =Panicle length TGW = 1000-grain weight, S.E = Standard error of mean

yield was observed among the DH lines. At Crowley, as shown in Table 3.1, the Caiapo parent yielded twice as much as the mean yield of the DH lines. In spite of the non-adaptability of the DH lines at Crowley, 8 lines out-yielded this parent by 4.0-270 kg. Earlier researchers have reported the introgression of yield enhancing QTLs from other relatives of *O. sativa* (Moncada et al. 2001; Brondani et al. 2002) resulting in transgressive segregation for yield.

The grain weight of Caiapo parent was 10 g heavier than IRGC 103544, but only 0.85 g more than the DH lines. Transgressive segregation for heavier grains was observed among the DH lines as 28 percent of the lines produced grains at least 9.7 g and 0.4 g heavier than IRGC103544 and Caiapo, respectively. Average 1000-grain weight (TGW) of the DH lines was

2 g more than that of Caiapo at Crowley. However, there was transgressive segregation for both low (16 percent of DH lines) and high (32 percent of DH lines) TGW.

Results from Crowley showed weak phenotypic correlation among traits even though some associations were statistically significant ($p < 0.05$) (Table 3.3). Heading date and plant height, panicle length and plant height, grain yield and tillers per plant were significant ($p < 0.01$). Correlations between tillers per plant and tillers per plant, tillers per plant and panicle length, TGW and panicle length, TGW and Tillers per plant and, yield and panicle length were also statistically significant ($p = 0.05$). The weak correlations observed in this study are an indication of the complexity of the agronomic traits being evaluated.

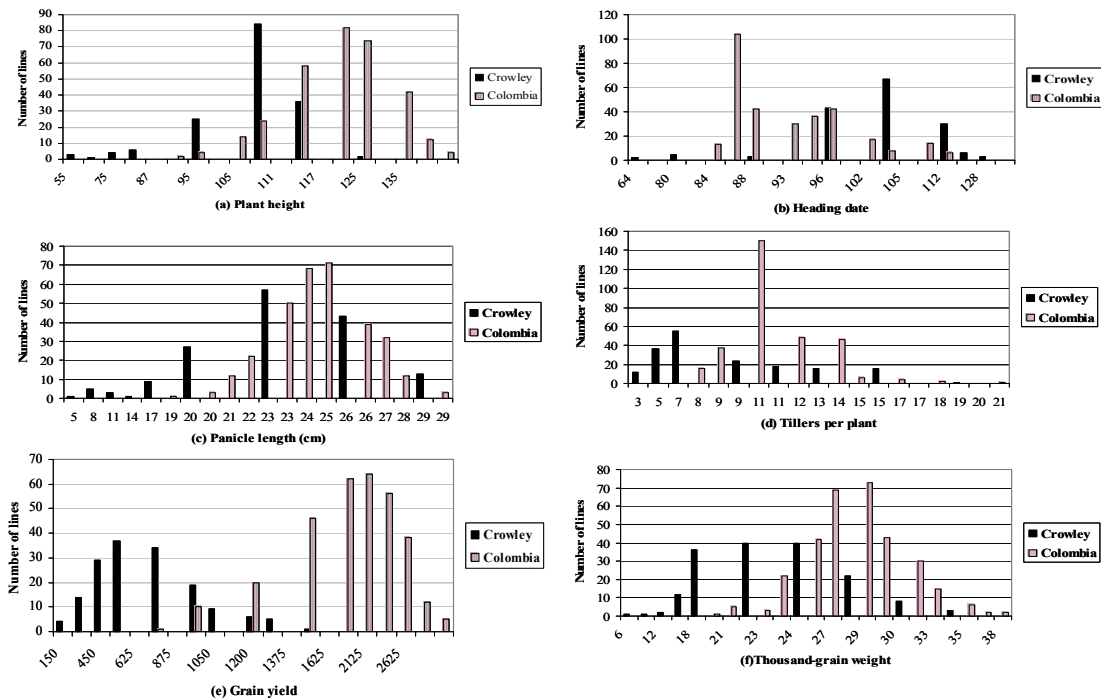


Figure 3.2. Frequency distribution of agronomic traits in 312 DH lines at CIAT, Colombia, 2001 and 158 DH lines at Crowley, LA 2002.

Table 3.3. Correlation coefficients among grain quality characteristics of 312 DH lines derived from *O. sativa* x *O. glaberrima* cross at Colombia, 2001 and Crowley, 2002

	PLHt	HD	Till/Plt	Panlen	Yield	TGW
Colombia						
PLHt						
HD	0.55**					
Till/Plt	0.084	0.22**				
Panlen	0.55**	0.27**	-0.07			
Gy	0.37**	0.17*	0.40**	0.11		
TGW	-0.16	-0.31**	-0.10	-0.05	-0.12	
Corwley, LA						
PLHt						
HD	0.38**					
Till/Plt	-0.19	0.05				
Panlen	0.30**	0.13	-0.20*			
Gy	0.02	0.04	0.38**	0.18*		
TGW	0.06	0.03	-0.16	-0.16*	-0.14	

^a PLHt = Plant height (cm), HD = Heading date (days), Till/Plt = Tillers per plant, Panlen = Panicle length (cm), Gy = Grain yield (kg/ha), TGW = 1000-grain weight (g)

* = significant at $p < 0.05$

** = significant at $p < 0.01$

3.3.3. Genetic Mapping of QTLs for Agronomic Traits

3.3.3.1 Plant Height

QTLs reported here were those detected by both simple interval and confirmed by composite interval analysis. Result from Columbia showed that QTLs of moderate to minor effects were detected for plant height (Table 3.4). For example, the QTL, *ht8*, was detected on chromosome 8 that accounted for 7.3 percent of phenotypic variation with a log of the odds (LOD) score of 3.5. This QTL was detected 15 cM from Qph8, reported by Mei et al. (2003). QTL *ht10* for plant height was detected on chromosome 10 while *ht11* was detected on chromosome 1, accounting for 6.1 and 7.4 percent of phenotypic variation, respectively. The QTL, *ht11* was detected 8 cM from a QTL for plant height reported by Yu et al. (2002). Alleles from the Caiapo parent were associated with increased plant height at all loci except the *ht11* locus (Table 3.4). For the

Louisiana location, four QTLs were detected for plant height, *ht1*, *ht4*, *ht2* and *ht8* which accounted for 38.1, 17.5, 24.6 and 16.7, respectively (Table 3.5). Alleles from Caiapo were associated with increased plant height at these loci. Of these, only *ht8* was detected at both locations. QTL *ht2* was detected 2.2 cM from that detected by Li et al 1995 and 19.2 cM from the position reported by Xiao et al. QTL *ht8* was detected by Xiao et al. (1996) that mapped 1.3 cM away from this QTL *ht8* detected in this study.

3.3.3.2 Days to Heading

Two chromosomal regions on chromosome 4 were associated with days to heading in Colombia (Table 3.4 and Fig. 3.4). QTL *hd4* accounted for 4 percent of the phenotypic variation with a LOD score of 2.5, and this QTL accounted for the 4 percent of variation in days to heading at this locus. A similar QTL was detected 3 cM from this position (Xiao et al 1995). Brondani et al. (2002) detected QTL of medium effect for days to flowering which accounted for 12.39 percent of phenotypic variation 3.7cM from this position. Another QTL, *hd6* was detected on chromosome 6 with a LOD score of 8.0 which accounted for 6.4 percent of variation. Yu et al. (2002) detected QTL for heading date 2 cM from this position that explained 17.1 percent and 25 percent of phenotypic variation, respectively, in two different experiments. Alleles from Caiapo were associated with heading date at this locus. Three QTLs, *hd1*, *hd2*, and *hd6*, found on chromosomes 1, 2 and 6, respectively, were detected for days to heading among the DH lines (Table 3.5 and Fig. 3.4). These QTLs were detected with LOD scores of 3.0, 2.0 and 3.0 for the respective QTLs. Of these three, 2 loci were contributed by IRGC 103544. Days to heading due to IRGC103544 were expected in view of the fact that this parent did not flower throughout the duration of the experiment. The QTL *hd6* was located 10 cM away from *Hd-3* reported by Yano et al. (1997).

Table 3.4 Quantitative Trait Loci for agronomic traits in 312 DH lines derived from *O. sativa* (Caiapo) x *O. glaberrima* (AC# IRGC103544) cross; Colombia, 2001 and Crowley, LA, 2002

Trait	QTL	Chr.	Marker Interval	Site ^c	Marker Position	Additive	LOD	R ²	Allelic Source
Plant height	<i>ht8</i>	8	RM230-RM264	Co	120.0	-11.4	3.5	7.3	Caiapo
	<i>ht10</i>	10	RM184-RM228	Co	38.0	-8.7	3.3	6.1	Caiapo
	<i>ht11</i>	11	RM224-RM144	Co	65.0	7.4	3.0	7.4	IRGC103554
	<i>ht1</i>	1	RM226-RM297	Cr	49.8	-26.3	3.4	38.1	Caiapo
	<i>ht4</i>	4	RM349-RM1311	Cr	18.8	-30.2	2.8	17.5	Caiapo
	<i>ht2</i>	2	RM208-RM213	Cr	127.8	-21.4	4.0	24.8	Caiapo
	<i>ht8</i>	8	RM152-RM25	Cr	7.0	-39.5	7.5	16.7	Caiapo
Day to heading	<i>hd4</i>	4	RM349-RM124	Co	25.0	26.2	2.5	4.0	IRGC103554
	<i>hd6</i>	6	RM190-RM253	Co	12.8	-2.7	8.0	6.4	Caiapo
	<i>hd1</i>	1	RM297-RM315	Cr	72.7	47.8	2.0	20.5	IRGC103554
	<i>hd2</i>	2	RM211-RM236	Cr	10.4	36.2	2.0	8.9	IRGC103554
	<i>hd6</i>	6	RM190-RM253	Cr	12.8	-3.5	3.0	9.5	Caiapo
	<i>till4</i>	4	RM273-RM241	Co	2.0	1.6	2.0	5.0	IRGC103554
Tillers /plant	<i>till7</i>	7	RM351-RM234	Co	26.0	2.2	2.0	7.6	IRGC103554
	<i>till10</i>	10	RM184-RM171	Co	28.0	1.8	3.0	7.2	IRGC103554
	<i>till1</i>	1	RM297-RM315	Cr	74.7	3.7	2.0	10.1	IRGC103554
	<i>till4</i>	4	RM273-RM241	Cr	2.0	-3.0	3.0	15.0	Caiapo
	<i>till11</i>	11	RM224-RM144	Cr	69.0	-2.4	2.0	5.8	Caiapo
Panicle length	<i>pan2</i>	2	RM240-RM208	Co	108.0	-2.2	2.5	4.6	Caiapo
	<i>pan3</i>	3	RM60-RM81B	Co	6.0	2.8	4.5	8.1	IRGC103554
	<i>pan5</i>	5	RM169-RM163	Co	30.0	-2.5	3.5	4.6	Caiapo
	<i>pan7</i>	7	RM11-RM10	Co	12.0	3.9	5.5	11.8	IRGC 103544
	<i>pan8</i>	8	RM337-RM152	Co	0.0	9.9	3.8	8.2	IRGC103554
	<i>pan11</i>	11	RM224-RM144	Co	71.0	-3.5	8.0	9.4	Caiapo

c = Experimental location, Co = Colombia Cr = Crowley; Marker position = position of peak marker in cM within the interval, Additive = additive effects expressed in terms of estimated change in phenotype expected from introgression of the *O. glaberrima* allele into Caiapo genome
 LOD= Log₁₀ (probability of linkage/probability of no linkage), R²= percentage of variation accounted for by the QTL.

Table 3.4 cont'd

Trait	QTL	Chr.	Marker Interval	Site ^c	Position	Additive	LOD	R ²	Allelic Source
Panicle length	<i>pan1</i>	1	RM226-RM297	Cr	45.0	-5.6	3.0	4.9	Caiapo
	<i>pan3</i>	3	RM81B-RM7	Cr	40.0	-4.7	2.5	5.4	Caiapo
	<i>pan7</i>	7	RM11-RM10	Cr	12.0	9.3	3.2	3.0	IRGC103554
	<i>pan8</i>	8	RM337-RM152	Cr	0.0	9.9	3.8	10.2	IRGC103554
	<i>pan10</i>	10	RM184-RM171	Cr	28.0	-4.5	3.0	5.3	Caiapo
	<i>pan11</i>	11	RM224-RM144	Cr	71.0	-4.7	9.0	13.4	Caiapo
Grain Yield	<i>gy1</i>	1	RM246-RM128	Cr	33.0	12.8	2.0	4.7	IRGC103554
	<i>gy5</i>	5	RM163-RM188	Co	44.0	-30.5	2.5	3.5	Caiapo
	<i>gy7</i>	7	RM125-RM11	Co	28.0	-39.8	2.5	4.0	Caiapo
Grain Yield	<i>gy2</i>	2	RM174-RM290	Cr	42.0	15.5	2.0	5.9	IRGC103554
	<i>gy9</i>	9	RM316-RM257	Cr	7.0	18.3	4.2	12.0	IRGC103554
1000-Grain weight	<i>gw2</i>	2	RM208-RM213	Co	131.0	2.4	2.5	4.3	IRGC103554
	<i>gw7</i>	7	RM351-RM234	Co	34.0	-8.2	3.0	7.3	Caiapo
	<i>gw5</i>	5	RM26-RM274	Co	71.0	4.0	2.5	6.8	IRGC103554
	<i>gw8</i>	8	RM284-RM149	Co	8.0	-3.2	3.0	6.4	Caiapo
1000-Grain weight	<i>gw3</i>	3	RM148-RM85	Cr	110.4	3.8	2.5	5.0	IRGC103554
	<i>gw7</i>	7	RM351-RM234	Cr	34.0	-6.2	4.2	13.3	Caiapo
	<i>gw8</i>	8	RM126-RM339	Cr	23.0	5.0	2.5	6.9	IRGC103554
	<i>gw10</i>	10	RM171-RM228	Cr	40.0	-5.0	2.7	7.4	Caiapo

c = Experimental location, Co = Colombia Cr = Crowley; Marker position = position of peak marker in cM within the interval, Additive = additive effects expressed in terms of estimated change in phenotype expected from introgression of the *O. glaberrima* allele into Caiapo genome LOD= Log₁₀ (probability of linkage/probability of no linkage), R²= percentage of variation accounted for by the QTL.

3.3.3.3 Tillers Per Plant

Three QTLs, *till4*, *till7* and *till10* were detected for the number of tillers per plant among the DH lines in Colombia (Table 3.4 and Fig. 3.3). These QTLs were detected on chromosomes 4, 7 and 10 with LOD scores of 3.0, 2.0 and 2.5 and accounted for 5 percent, 7.6 percent and 7.2 percent of phenotypic variation, respectively. In all three cases IRGC 103544 alleles were associated with increased number of tillers. Brondani et al. (2002) detected a QTL for tiller number 6.8 cm from *till4*. At Crowley, three QTLs were detected for tillers per plant namely,

till1, *till4* and *till11* on chromosomes 1, 4 and 11 which accounted for 10.1, 15.0 and 5.8 percent of phenotypic variation, respectively (Table 3.5). The QTL *till4* was the same as the one detected in Colombia. QTLs *till1* and *till4* were contributed by IRGC103544 alleles while Caiapo allele was associated with a minor QTL (*till11*). Brondani et al. (2002) detected QTL for tiller number in the region ~ 10 cM from the *till11* (Fig. 3.4).

3.3.3.4 Panicle Length

Six QTLs were identified in this population affecting panicle length in Colombia and Crowley as shown in Table 3.4 and Fig.3.3. On chromosome 2, *pan2* was detected at a LOD of 2.5, and this QTL accounted for 4.6 percent of variation in this trait. Alleles from Caiapo were associated with long panicles at this locus. The QTL, *pan3*, detected on chromosome 3 at the interval RM60-RM81B with LOD score of 4.5 accounted for 8.1 percent of phenotypic variation. Alleles from IRGC 103544 contributed to panicle length at this locus. A QTL was detected on chromosome 8 at both locations in the interval RM337-RM152 with alleles contributed by both parents. For Colombia and Crowley, *pan8* accounted for 8.2 percent and 10.2 percent phenotypic variation, respectively. On chromosome 11, *pan11* was detected with a LOD score of 8.0 in a different region of the chromosome from that reported by Brondani et al. (2002) which accounted for 9.4 percent of phenotypic variation in their study. In our study at Crowley, six chromosomal regions were found associated with variation in panicle length as shown in Table 3.4. These QTLs, *pan1*, *pan3*, *pan7*, *pan8*, *pan10*, and *pan11* accounted for 3.4 to 13.4 percent of phenotypic variation. It is worthy to note that four out of six QTLs were also detected for this trait in Colombia (*pan3*, *pan7*, *pan8* and *pan11*). Of these QTLs, *pan7* was found 4 cM from *phl7* reported earlier by Brondani et al. (2002) (Fig. 3.3).

3.3.3.5 Grain Yield

Two QTLs, *gy5* and *gy7* that mapped to chromosomes 5 and 7, respectively, were detected for grain yield (Table 3.4 and Fig. 3.3) in this DH population. Both QTLs were detected at LOD scores of 3.0 and accounted for 4.0 and 4.3 percent phenotypic variation, respectively. Alleles from the Caiapo parent were associated with increased grain yield at these loci. Brondani et al. (2002) also reported the association of grain yield with RM11 on chromosome 7. At Crowley, there were three QTLs detected for grain yield (Table 3.4). The QTL *gy1*, *gy2*, and *gy9* were detected on chromosomes 1, 2, and 9 with LOD scores of 2, 2, and 4.2, respectively. The corresponding percent phenotypic variations were 4.7, 5.9, and 12.0. Brondani et al. (2002) also detected QTL for grain yield on chromosome 7 at position 29.8 cM on this chromosome, while Xiao et al. (1996) detected QTL at 32 cM on the same chromosome.

3.3.3.6 Thousand Grain Weight

Four QTLs were detected for thousand-grain weight in Colombia (Table 3.4 and Fig. 3.3). The QTL *gw2* located between RM208 and RM213 was detected on chromosome 2 with a LOD of 2.5 and this accounted for 4.3 percent of phenotypic variation. Two chromosomal regions were found to be associated with grain weight on chromosome 5. The QTL *gw5* was detected with LOD score of 2.5 and accounted for 6.8 percent of phenotypic variation. On chromosome 8, *gw8* was detected in the interval RM284-RM149 with LOD of 3.0, and this QTL accounted for 6.4 percent of phenotypic variation.

At Crowley, four QTLs affecting 1000-grain weight were detected (Table 3.4). These were *gw3* on chromosome 3, *gw7* on chromosome 7, *gw8* on chromosome 8 and *gw10* on chromosome 10. These QTLs were detected at LOD scores of 2.5, 4.2, 2.5, and 2.7, respectively. Percent of phenotypic variation accounted for were 5, 13.3, 6.9 and 7.4 percent, respectively, for *gw3*, *gw7*, *gw8* and *gw10*. QTL for grain weight has been reported 13 cM and

~9 cM from these positions on chromosomes 3 and 7, respectively, (Brondani et al. 2002). It is worthy to note that QTL *gw7* was detected at both locations.

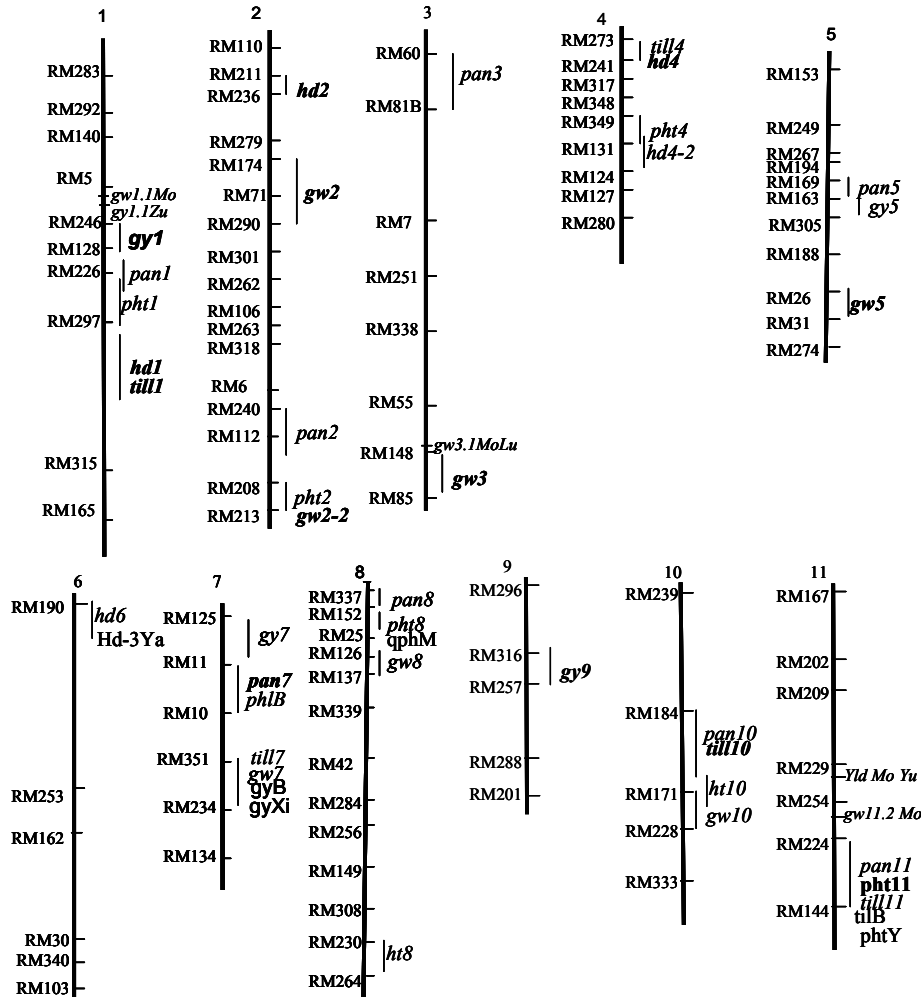


Fig. 3.3. Assignment of QTLs for 6 agronomic traits on the rice linkage map adjusted by MapDisto program among 312 DH lines derived from *O. sativa* (Caiapo) x *O. glaberrima* (IRGC 103544) cross at Colombia and Crowley. Confidence interval for each QTL indicated as bar to the right of chromosome. QTLs in bold indicate positive allelic effects from IRGC 103544. ht = QTL for plant height, hd = QTL for heading date, pan = QTL for panicle length, till = QTL for tillers per plant, gy = QTL for grain yield, gw = QTL for 1000-grain weight. Zu = QTL reported by Zhuang et al. (1997), Mo = QTL reported by Moncada et al. (2001), Lu = QTL reported by Lu et al. (1997), Xi = QTL reported by Xiao et al. (1996), B = QTL reported by Brondani et al. (2002), Y = QTL reported by Yu et al. (2002), M = QTL reported by Mei et al. (2003), Ya = QTL reported by Yano et al. (19978). * = QTL detected in this study.

3.3.4. Classification of 312 DH lines Using Markers Identified by Discriminant Analysis with and without Consideration of Population Structure

The issue of classifying a population of 312 DH lines into pre-defined groups of low and high trait values is a multivariate problem which involves the identification of molecular markers as predictor variables used to make such allocations. For the first series of analyses, the stepwise discriminant procedure forward selection option without adjustment for population structure was used to select a maximum of 20 markers with the largest contribution to differentiate between “high and low” groups. Markers so selected were then used to allocate test lines into either of two groups of contrasting trait values and the percent correct classification was evaluated.

Table 3.5 shows the percent correct classification of tall and short plant types using data from the Colombia location in 2001. The highest accuracy achieved with the K-NN classification using 20 markers was 93 percent when classes of tall and short plants were separated by 3SD. Accuracy of classification was reduced by 5 percent and 13 percent with intermediate (2SD) and minimal differentiation (1SD), respectively. When population structure was assumed, there was a dramatic improvement in the classification of lines as the highest accuracy of classification (100 percent) was achieved with 15 markers at the 2SD level for all subpopulations. However, 3SD proved to be the most efficient level of differentiation as only 5 markers achieved the same level of accuracy as 10-15 markers at the 2SD level.

For days to heading, the same trend was observed as in plant height (Table 3.6). Highest accuracy of classification was achieved with 10 markers when structure was assumed. Ten markers either achieved the same level of accuracy or higher when structure was assumed compared to when it was not assumed. The correct classification of tillers per plant into pre-defined groups of high and low is presented in Table 3.7. For this trait, 2SD level of differentiation was not sufficient for high accuracy of classification of lines. At the highest level

Table 3.5. Percent correct classification for 312 DH lines, Colombia, 2001 assigned to groups of tall and short plants using a *k*-nearest neighbor (*k*-NN) algorithm with and without consideration of population structure, Colombia, 2001

<u>Population Structure</u>	<u>Number of DH lines</u>	<u>Training Samples</u>	<u>Number of DA-markers included in the K-nearest neighbor classification model^b</u>			
			<u>5</u>	<u>10</u>	<u>15</u>	<u>20</u>
Assuming no Structure K = 1	312	1SD	74 ^a	79	80	80
		2SD	83	85	85	88
		3SD	85	90	93	95
<u>Assuming structure</u>						
Sub population 1	135	1SD	74	82	84	90
		2SD	73	97	100	100
		3SD	100	100	100	100
Sub population 2	95	1SD	75	88	91	98
		2SD	93	93	100	100
		3SD	100	ND	ND	ND
Sub population 3	82	1SD	77	86	92	96
		2SD	85	100	ND	ND
		3SD	100	ND	ND	ND

^a Estimated by cross validation using the leave one out method (Hand 1997).

STEPDISC procedure using the forward method (SAS/STAT software version 8), limited to 20 markers with largest contribution to differentiation between low and high groups.

^b Selection of markers included in the K-NN classificatory model based on training samples defined as low or high groups with different degree of phenotypic overlapping (3 SD= three standard deviations between low and high groups; 2 SD = two standard deviations ; 1 SD = one standard deviation). ND = No data because no selection was made.

Table 3.6. Percent correct classification for 312 DH lines, Colombia, 2001 into early and late heading date groups using a *k*-nearest neighbor (*k*-NN) algorithm with and without consideration of population structure, Colombia, 2001

<u>Population Structure</u>	<u>Number of DH lines</u>	<u>Training samples</u>	<u>Number of DA-markers included in the K-nearest neighbor classification model^b</u>			
			<u>5</u>	<u>10</u>	<u>15</u>	<u>20</u>
Assuming no structure K=1	312	1SD	65 ^a	69	74	76
		2SD	78	80	83	83
		3SD	67	74	85	90
<u>Assuming structure</u>						
Sub population 1	135	1SD	67	74	80	80
		2SD	76	76	80	80
		3SD	73	90	100	100
Sub population 2	95	1SD	68	68	85	85
		2SD	82	92	100	100
		3SD	90	100	100	100
Sub population 3	82	1SD	83	88	95	98
		2SD	95	100	100	100
		3SD	100	100	ND	ND

^a Estimated by cross validation using the leave one out method (Hand 1997).

STEPDISC procedure using the forward method (SAS/STAT software version 8), limited to 20 markers with largest contribution to differentiation between low and high groups.

^b Selection of markers included in the K-NN classificatory model based on training samples defined as low or high groups with different degree of phenotypic overlapping (3 SD= three standard deviations between low and high groups; 2 SD = two standard deviations ; 1 SD = one standard deviation). ND = No data because no selection was made.

Table 3.7. Percent correct classification for 312 DH lines into low and high tillers per plant groups using a *k*-nearest neighbor (*k*-NN) algorithm with and without consideration of population structure, Colombia, 2001

<u>Population Structure</u>	<u>Number of DH lines</u>	<u>Training samples</u>	<u>Number of DA-markers included in the K-nearest neighbor classification model^b</u>			
			<u>5</u>	<u>10</u>	<u>15</u>	<u>20</u>
Assuming no structure K=1	312	1SD	70 ^a	76	80	80
		2SD	83	85	85	88
		3SD	75	85	87	93
<u>Assuming structure</u>						
Sub population 1	135	1SD	70	74	76	78
		2SD	89	92	94	100
		3SD	100	100	100	100
Sub population 2	95	1SD	67	82	85	97
		2SD	66	94	94	100
		3SD	100	ND	ND	ND
Sub population 3	82	1SD	71	76	79	86
		2SD	83	100	ND	ND
		3SD	100	ND	ND	ND

^a Estimated by cross validation using the leave one out method (Hand 1997).

STEPDISC procedure using the forward method (SAS/STAT software version 8), limited to 20 markers with largest contribution to differentiation between low and high groups. ^bSelection of markers included in the K-NN classificatory model based on training samples defined as low or high groups with different degree of phenotypic overlapping (3 SD= three standard deviations between low and high groups; 2 SD = two standard deviations ; 1 SD = one standard deviation). ND = No data because no selection was made.

of differentiation however, only 5 markers were required to achieve 100 percent correct classification into high and low trait values. From Table 3.8, the percent correct classification into groups of short and long panicles increased as more markers were included in the model.

Table 3.8. Percent correct classification for 312 DH lines into short and long panicle length groups using a *k*-nearest neighbor (*k*-NN) algorithm with and without consideration of population structure, Colombia, 2001

<u>Population Structure</u>	<u>Number of DH lines</u>	<u>Training samples</u>	<u>Number of DA-markers included in the K-nearest neighbor classification model^b</u>			
			<u>5</u>	<u>10</u>	<u>15</u>	<u>20</u>
Assuming no structure K=1	312	1SD	70 ^a	79	80	80
		2SD	80	83	85	88
		3SD	75	85	87	95
<u>Assuming structure</u>						
Sub population 1	135	1SD	74	79	88	88
		2SD	73	82	100	100
		3SD	100	100	100	100
Sub population 2	95	1SD	75	88	91	98
		2SD	93	93	100	100
		3SD	100	ND	ND	ND
Sub population 3	82	1SD	77	86	92	96
		2SD	85	100	ND	ND
		3SD	100	ND	ND	ND

^a Estimated by cross validation using the leave one out method (Hand 1997).

STEPDISC procedure using the forward method (SAS/STAT software version 8), limited to 20 markers with largest contribution to differentiation between low and high groups.

^b Selection of markers included in the K-NN classificatory model based on training samples defined as low or high groups with different degree of phenotypic overlapping (3 SD= three standard deviations between low and high groups; 2 SD = two standard deviations ; 1 SD = one standard deviation). ND = No data because no selection was made.

Table 3.9. Percent correct classification for 312 DH lines into low and high grain yield groups using a *k*-nearest neighbor (*k*-NN) algorithm with and without consideration of population structure, Colombia, 2001

<u>Population Structure</u>	<u>Number of DH lines</u>	<u>Training samples</u>	<u>Number of DA-markers included in the K-nearest neighbor classification model^b</u>			
			<u>5</u>	<u>10</u>	<u>15</u>	<u>20</u>
Assuming no structure K=1	312	1SD	67 ^a	73	75	75
		2SD	79	77	82	87
		3SD	86	94	97	97
<u>Assuming structure</u>						
Sub population 1	135	1SD	74	82	86	86
		2SD	85	92	97	98
		3SD	100	100	ND	ND
Sub population 2	95	1SD	75	80	81	85
		2SD	86	88	95	100
		3SD	100	ND	ND	ND
Sub population 3	82	1SD	84	94	100	100
		2SD	100	100	ND	ND
		3SD	100	ND	ND	ND

^a Estimated by cross validation using the leave one out method (Hand 1997).

STEPDISC procedure using the forward method (SAS/STAT software version 8), limited to 20 markers with largest contribution to differentiation between low and high groups.

^b Selection of markers included in the K-NN classificatory model based on training samples defined as low or high groups with different degree of phenotypic overlapping (3 SD= three standard deviations between low and high groups; 2 SD = two standard deviations ; 1 SD = one standard deviation). ND = No data because no selection was made.

Table 3.10. Percent correct classification for 312 DH lines, Colombia, 2001 into low and high 1000-grain weight groups using a *k*-nearest neighbor (*k*-NN) algorithm with and without consideration of population structure, Colombia, 2001

<u>Population structure</u>	<u>Number of DH lines</u>	<u>Training samples</u>	<u>Number of DA-markers included in the K-nearest neighbor classification model^b</u>			
			<u>5</u>	<u>10</u>	<u>15</u>	<u>20</u>
Assuming no structure k=1	312	1SD	75 ^a	79	80	80
		2SD	83	85	88	88
		3SD	75	85	87	93
<u>Assuming structure</u>						
Sub population 1	135	1SD	74	82	84	97
		2SD	73	97	100	100
		3SD	100	100	100	100
Sub population 2	95	1SD	75	88	91	98
		2SD	93	95	100	100
		3SD	100	ND	ND	ND
Sub population 3	82	1SD	77	86	92	96
		2SD	85	100	ND	ND
		3SD	100	100	ND	ND

^a Estimated by cross validation using the leave one out method (Hand 1997).

STEPDISC procedure using the forward method (SAS/STAT software version 8), limited to 20 markers with largest contribution to differentiation between low and high groups.

^b Selection of markers included in the K-NN classificatory model based on training samples defined as low or high groups with different degree of phenotypic overlapping (3 SD= three standard deviations between low and high groups; 2 SD = two standard deviations ; 1 SD = one standard deviation). ND = No data because no selection was made.

An increase of 5 markers from 10 led to a corresponding increase of 6 percent in classification accuracy under the assumption of no structure. Increasing the number of markers beyond 15 did not result in increased accuracy of classification. The same trend was observed when structure was assumed as shown in Table 3.9. Generally for this trait, the 3SD level of differentiation produced the highest accuracy of classifying DH lines into groups of high and low yields with as few as 5 markers.

Classification of DH lines into big and small grains based on their weight is shown in Table 3.10. Using the nearest neighbor algorithm, 20 markers gave 93 percent correct classification when structure was not assumed. However, with the assumption of structure, only 5 markers were required to achieve 100 percent correct classification at the same level of differentiation. In some cases, the program could not select any more markers after 100 percent accuracy had been achieved. The same general trend for results from the Columbia location in 2001 was also found in classifying lines planted at Crowley in 2002. (Tables 3.11- 3.16).

3.3.5. Comparison of Discriminant Analysis with Interval, Regression and Bulk Segregant Analyses for Marker Assisted Classification

Markers detected by interval analysis in this study, and those selected by BSA, multiple regression and discriminant analyses were used to classify DH lines into pre-defined groups of high and low values of the six agronomic traits (plant height, days to heading, panicle length, tillers per plant, grain yield and 1000-grain weight). Lines in the highest level of differentiation (3SD) adjusted for structure, were used as training sets for this classification. The percent correct classification was estimated for each case and compared. Chromosomal locations of the markers selected by each method were also noted. For the DA markers, selection was made by the

Table 3.11. Percent correct classification for 158 DH lines into tall and short plant height groups using a *k*-nearest neighbor (*k*-NN) algorithm with and without consideration of population structure, Crowley, LA. 2002

<u>Population Structure</u>	<u>Number of DH lines</u>	<u>Training samples</u>	<u>Number of DA-markers included in the K-nearest neighbor classification model^b</u>			
			<u>5</u>	<u>10</u>	<u>15</u>	<u>20</u>
Assuming no structure K=1	158	1SD	80 ^a	86	87	92
		2SD	86	90	97	100
		3SD	90	97	100	100
<u>Assuming structure</u>						
Sub population 1	96	1SD	74	80	82	89
		2SD	73	97	100	100
		3SD	100	ND	ND	ND
Sub population 2	56	1SD	75	88	91	98
		2SD	93	93	100	100
		3SD	100	ND	ND	ND

^a Estimated by cross validation using the leave one out method (Hand 1997).

STEPDISC procedure using the forward method (SAS/STAT software version 8), limited to 20 markers with largest contribution to differentiation between low and high groups.

^b Selection of markers included in the K-NN classificatory model based on training samples defined as low or high groups with different degree of phenotypic overlapping (3 SD= three standard deviations between low and high groups; 2 SD = two standard deviations ; 1 SD = one standard deviation). ND = No data because no selection was made.

Table 3.12. Percent correct classification for 158 DH lines, Crowley, 2002 into early and late heading date groups using a k -nearest neighbor (k -NN) algorithm with and without consideration of population structure, Crowley, LA. 2002

<u>Population Structure</u>	<u>Number of DH lines</u>	<u>Training samples</u>	<u>Number of DA-markers included in the K-nearest neighbor classification model^b</u>			
			<u>5</u>	<u>10</u>	<u>15</u>	<u>20</u>
Assuming no structure K=1	158	1SD	70 ^a	77	84	94
		2SD	80	85	85	97
		3SD	89	92	97	100
<u>Assuming structure</u>						
Sub population 1	96	1SD	74	82	82	95
		2SD	73	97	100	100
		3SD	100	ND	ND	ND
Sub population 2	62	1SD	75	88	94	98
		2SD	93	95	100	100
		3SD	100	ND	ND	ND

^a Estimated by cross validation using the leave one out method (Hand 1997).

STEPDISC procedure using the forward method (SAS/STAT software version 8), limited to 20 markers with largest contribution to differentiation between low and high groups.

^b Selection of markers included in the K-NN classificatory model based on training samples defined as low or high groups with different degree of phenotypic overlapping (3 SD= three standard deviations between low and high groups; 2 SD = two standard deviations ; 1 SD = one standard deviation). ND = No data because no selection was made.

Table 3.13. Percent correct classification for 158 DH lines, Crowley, 2002 into low and high tillers per plant groups using a *k*-nearest neighbor (*k*-NN) algorithm with and without consideration of population structure, Crowley, LA. 2002

<u>Population Structure</u>	<u>Number of DH lines</u>	<u>Training samples</u>	<u>Number of DA-markers included in the K-nearest neighbor classification model^b</u>			
			<u>5</u>	<u>10</u>	<u>15</u>	<u>20</u>
Assuming no structure						
K=1	158	1SD	72 ^a	79	85	89
		2SD	74	85	91	97
		3SD	86	89	93	100
<u>Assuming structure</u>						
Sub population 1	96	1SD	78	91	92	98
		2SD	73	97	100	100
		3SD	100	100	ND	ND
Sub population 2	62	1SD	75	88	91	98
		2SD	89	90	94	100
		3SD	100	ND	ND	ND

^a Estimated by cross validation using the leave one out method (Hand 1997).

STEPDISC procedure using the forward method (SAS/STAT software version 8), limited to 20 markers with largest contribution to differentiation between low and high groups.

^b Selection of markers included in the K-NN classificatory model based on training samples defined as low or high groups with different degree of phenotypic overlapping (3 SD= three standard deviations between low and high groups; 2 SD = two standard deviations ; 1 SD = one standard deviation). ND = No data because no selection was made.

Table 3.14. Percent correct classification for 158 DH lines into long and short panicles groups using a *k*-nearest neighbor (*k*-NN) algorithm with and without consideration of population structure, Crowley, LA. 2002

<u>Population Structure</u>	<u>Number of DH lines</u>	<u>Training samples</u>	<u>Number of DA-markers included in the K-nearest neighbor classification model^b</u>			
			<u>5</u>	<u>10</u>	<u>15</u>	<u>20</u>
Assuming no structure K=1	158	1SD	70 ^a	75	80	80
		2SD	80	83	85	85
		3SD	75	85	87	93
<u>Assuming structure</u>						
Sub population 1	96	1SD	74	82	82	97
		2SD	76	97	100	100
		3SD	100	100	100	100
Sub population 2	62	1SD	75	88	91	95
		2SD	93	93	100	100
		3SD	100	ND	ND	ND

^a Estimated by cross validation using the leave one out method (Hand 1997).

STEPDISC procedure using the forward method (SAS/STAT software version 8), limited to 20 markers with largest contribution to differentiation between low and high groups.

^cSelection of markers included in the K-NN classificatory model based on training samples defined as low or high groups with different degree of phenotypic overlapping (3 SD= three standard deviations between low and high groups; 2 SD = two standard deviations ; 1 SD = one standard deviation).

Table 3.15. Percent correct classification for 158 DH lines, Crowley, 2002 into high and low grain yield groups using a *k*-nearest neighbor (*k*-NN) algorithm with and without consideration of population structure, Crowley, LA. 2002

<u>Population Structure</u>	<u>Number of DH lines</u>	<u>Training samples</u>	<u>Number of DA-markers included in the K-nearest neighbor classification model^b</u>			
			<u>5</u>	<u>10</u>	<u>15</u>	<u>20</u>
Assuming no structure K=1	158	1SD	72 ^a	82	82	89
		2SD	80	90	95	98
		3SD	85	90	95	100
<u>Assuming structure</u>						
Sub population 1	96	1SD	74	80	85	89
		2SD	78	97	100	100
		3SD	100	100	100	100
Sub population 2	62	1SD	75	88	91	98
		2SD	93	93	100	100
		3SD	100	ND	ND	ND

^a Estimated by cross validation using the leave one out method (Hand 1997).

STEPDISC procedure using the forward method (SAS/STAT software version 8), limited to 20 markers with largest contribution to differentiation between low and high groups.

^cSelection of markers included in the K-NN classificatory model based on training samples defined as low or high groups with different degree of phenotypic overlapping (3 SD= three standard deviations between low and high groups; 2 SD = two standard deviations ; 1 SD = one standard deviation).

Table 3.16. Percent correct classification for 158 DH lines, Crowley, 2002 into low and high 1000-grain weight groups using a *k*-nearest neighbor (*k*-NN) algorithm with and without consideration of population structure, Crowley, LA. 2002

<u>Population Structure</u>	<u>Number of DH lines</u>	<u>Training samples</u>	<u>Number of DA-markers included in the K-nearest neighbor classification model^b</u>			
			<u>5</u>	<u>10</u>	<u>15</u>	<u>20</u>
Assuming no structure K=1	158	1SD	74 ^a	79	80	80
		2SD	83	85	85	88
		3SD	75	85	87	93
<u>Assuming structure</u>						
Sub population 1	96	1SD	70	80	88	88
		2SD	75	97	100	100
		3SD	100	100	100	100
Sub population 2	62	1SD	75	82	91	98
		2SD	82	93	100	100
		3SD	100	ND	ND	ND

^a Estimated by cross validation using the leave one out method (Hand 1997).

STEPDISC procedure using the forward method (SAS/STAT software version 8), limited to 20 markers with largest contribution to differentiation between low and high groups.

^c Selection of markers included in the K-NN classificatory model based on training samples defined as low or high groups with different degree of phenotypic overlapping (3 SD= three standard deviations between low and high groups; 2 SD = two standard deviations ; 1 SD = one standard deviation).

stepdisc procedure with the backward selection option to select the maximum number of informative markers for each trait.

In subpopulation 1 from Crowley, allocation of DH lines into groups of tall and short plants was most accurate with the DA procedure (Table 3.17). Five markers achieved 100 percent correct classification. The multiple regression models achieved 85 percent while BSA

achieved 90 percent correct classification. All of these methods identified at least seven markers. A similar trend was found with the materials in population 2 grown at Crowley (Table 3.18). Discriminant analysis was the most accurate in classification of lines into pre-defined groups of tall and short plants with 100 percent correct allocation. The four methods within two populations from Crowley selected different markers except in the IM method where 40% of the markers were similar.

In all three sub-populations from Colombia, however, multiple regression and DA classified with equal accuracy (Table 3.19), but 9 MR markers were required to achieve the same level of accuracy as five DA markers. Markers selected by BSA and interval analyses showed lower levels of accuracy (77 and 80 percent for IM and BSA, respectively) (Tables 3.19-3.21). It is worthy to note that one of the markers associated with differentiation of plant height, RM144 on chromosome 11, was detected by all four methods. For heading date, DA-selected markers gave the most accurate classification of lines into early and late heading groups (Table 3.17). With only five markers, DA achieved 100 percent correct classification compared to 81 percent, 61 percent and 72 percent for BSA, interval analysis and MR, respectively. Among subpopulation 1 from Crowley, DA and BSA detected two markers which were also detected by interval analysis for this trait (RM253 and RM124). In subpopulation 2, only marker RM297 was among QTL markers selected by the DA, but RM81B was common to all three (DA, BSA and MR) even though this marker was not associated with QTL for heading date. For the population grown in Colombia, 7 markers were sufficient to classify lines into early and late groups with 100 percent correct classification using DA (Table 3.19).

In all sub-populations, DA produced the highest accuracy. In subpopulation 1, DA and BSA selected 2 QTL markers each while in population 3, DA selected 3 QTL markers and BSA

Table 3.17 Chromosomal location of SSR markers and % correct classification of DH lines using IM, BSA and DA procedures, Crowley 2002, subpopulation 1

Trait	MR			IM/CIM			BSA			DA		
	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification
PHT	7	RM234		8	RM230		6	RM253		7	RM234	
	6	RM340		8	RM264		7	RM234		11	RM202	
	8	RM284		10	RM184		11	RM144		6	RM253	
	11	RM202		10	RM228		8	RM284		8	RM264	
	8	RM264		11	RM224		4	RM349		11	RM202	
	2	RM208		11	RM144		3	RM85				
	1	RM297		1	RM226		9	RM288				
	2	RM211	85	1	RM297	87			100			100
HD	11	RM209		4	RM349		1	RM297		4	RM124	
	4	RM241		4	RM124		4	RM124		6	RM253	
	5	RM305		6	RM190		8	RM256		1	RM297	
	3	RM148		6	RM253		1	RM292		8	RM230	
	2	RM6					6	RM253		2	RM318	
	8	RM230					7	RM134				
	2	RM262					11	RM167				
			72			61	10	RM171				100
PAN	1	RM283		2	RM240		1	RM283		8	RM337	
	3	RM60		2	RM208		3	RM60		3	RM60	
	1	RM226		3	RM60		1	RM226		6	RM190	
	1	RM165		3	RM81B		3	RM7		8	RM152	
	2	RM263		5	RM169		4	RM349				
	3	RM251		5	RM163		6	RM190				
	4	RM349		8	RM152		2	RM240				
	7	RM11		7	RM10							
	2	RM236		8	RM337							
	11	RM209	85			84			89			100

IM = interval mapping, BSA = Bulk segregant analysis, MR = Multiple regression, DA = Discriminant analysis
PHT = Plant height; HD= Heading date; PAN= Panicle length.

Table 3.17 Cont'd

Trait	MR		IM/CIM		BSA		DA			
	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	
TILL	2	RM208		1	RM297		10	RM239	10	RM184
	11	RM144		11	RM144		3	RM55	11	RM144
	9	RM296		4	RM273		8	RM42	4	RM280
	3	RM338		4	RM241		4	RM280	8	RM42
	1	RM165		11	RM224		6	RM162	5	RM163
	11	RM209		1	RM315		1	RM297		
	8	RM42								
	5	RM163	84			64			89	100
YLD	11	RM254		5	RM163		8	RM137	1	RM165
	3	RM148		5	RM188		11	RM224	11	RM167
	8	RM42		7	RM125		7	RM125	4	RM349
	8	RM152		7	RM11		2	RM301	8	RM42
	4	RM349		3	RM148		3	RM7	5	RM163
	3	RM55					3	RM55		
	7	RM125					11	RM144		
	2	RM301					1	RM165		
	3	RM148	88			72			92	100
TGW	8	RM284		2	RM208		2	RM106	7	RM351
	8	RM149		2	RM213		1	RM297	8	RM149
	2	RM211		7	RM351		4	RM127	1	RM297
	2	RM236		5	RM274		3	RM7	8	RM264
	1	RM297		8	RM284		8	RM42		
	8	RM284		8	RM149		2	RM110		
							10	RM333		
			85			62	1	RM315	94	100

IM = interval mapping, BSA = Bulk segregant analysis, MR = Multiple regression, DA = Discriminant analysis
TILL= Tillers per plant; YLD= Yield; TGW =1000-grain weight.

Table 3.18 Chromosomal location of SSR markers and % correct classification of DH lines using MR, IM, BSA and DA procedures, Crowley 2002, subpopulation 2

Trait	MR			IM/CIM			BSA			DA		
	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification
PHT	12	RM247	88	1	RM226	77	2	RM263	80	2	RM263	100
	5	RM249		1	RM297		2	RM208		3	RM338	
	11	RM144		4	RM131		11	RM144		8	RM126	
	7	RM125		2	RM208		9	RM288		11	RM144	
	2	RM211		2	RM213		7	RM125		2	RM211	
	3	RM338		8	RM230		5	RM163				
				8	RM264							
			11	RM144								
HD	3	RM81B	82	4	RM349	61	3	RM81B	90	3	RM81B	100
	8	RM149		4	RM124		8	RM149		7	RM134	
	11	RM202		6	RM190		11	RM202		8	RM337	
	5	RM249		6	RM253		5	RM249		1	RM297	
	10	RM333		1	RM297		10	RM333				
	9	RM288		1	RM315		9	RM288				
				2	RM211		8	RM137				
			2	RM236	10	RM239						
					2	RM290						
PAN	3	RM81B	82	3	RM60	74	3	RM81B	80	8	RM230	100
	2	RM290		3	RM81B		2	RM290		8	RM337	
	12	RM19		8	RM152		12	RM19		8	RM152	
	7	RM125		8	RM337		1	RM5		10	RM333	
	8	RM339		3	RM338		4	RM241		3	RM81B	
	8	RM284					7	RM125				
	10	RM333					8	RM337				
				8	Rm230							

IM = interval mapping, BSA = Bulk segregant analysis, MR = Multiple regression, DA = Discriminant analysis.
PHT = Plant height; HD= Heading date; PAN= Panicle length.

Table 3.18. cont'd

Trait	MR		IM/CIM		BSA		DA				
	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification		
TILL	2	RM208		1	RM297		8	RM264		3	RM81B
	5	RM249		11	RM144		11	RM229		2	RM297
	5	RM163		4	RM273		5	RM305		11	RM144
	11	RM209		4	RM241		2	RM263		2	RM174
	5	RM153		11	RM224		11	RM144		5	RM305
	11	RM254		1	RM315		2	RM290		11	RM254
								2	RM174		
			74			70	9	RM257	82		100
YLD	1	RM297		1	RM128		1	RM297		1	RM128
	4	RM127		2	RM174		11	RM224		1	RM226
	8	RM42		2	RM290		8	RM308		1	RM297
	8	RM308		9	RM316		8	RM230		5	RM274
	3	RM55		9	RM257		5	RM274		7	RM125
	1	RM292		7	RM125		4	RM127		7	RM234
	11	RM224					7	RM234		4	RM127
	5	RM274					11	RM254	86		100
			82			72					
TGW	5	RM163		3	RM148		8	RM264		3	RM85
	1	RM297		3	RM85		1	RM297		11	RM144
	11	RM144		5	RM163		11	RM144		2	RM208
	1	RM140		8	RM149		2	RM208		3	RM148
	11	RM167		8	RM126		3	RM85			
	5	RM188		8	RM339		2	RM263			
	1	RM315		10	RM171		11	RM167			
	8	RM42		10	RM228		1	RM165	95		100
				80			62				

IM = interval mapping, BSA = Bulk segregant analysis, MR = Multiple regression, DA = Discriminant analysis.

TILL= Tillers per plant; YLD= Yield; TGW =1000-grain weight.

Table 3.19 Chromosomal location of SSR markers and % correct classification of DH lines using MR, IM, BSA and DA procedures, Colombia 2001, subpopulation 1

Trait	MR			IM/CIM			BSA			DA		
	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification
PHT	1	RM283		1	RM226		2	RM213		2	RM262	
	11	RM209		1	RM297		2	RM262		1	RM283	
	2	RM213		4	RM131		1	RM140		7	RM125	
	3	RM7		2	RM208		2	RM208		11	RM144	
	1	RM292		2	RM213		7	RM125		8	RM230	
	5	RM188		8	RM230		11	RM144				
	5	RM153		8	RM264		1	RM165				
	6	RM190		11	RM144		11	RM224				
	6	RM253	100			72			80			100
HD	5	RM163		4	RM349		5	RM163		1	RM297	
	11	RM209		4	RM124		1	RM140		5	RM249	
	11	RM224		6	RM190		1	RM315		5	RM163	
	3	RM7		6	RM253		10	RM239		1	RM315	
	1	RM292		1	RM297		5	RM274		11	RM209	
	8	RM339		1	RM315		3	RM338		2	RM208	
	5	RM274		2	RM211		1	RM128		6	RM190	
	1	RM297		2	RM236		1	RM292				
	1	RM128					4	RM124				
		8	RM42	91			81			80		
PAN	1	RM283		3	RM60		2	RM290		1	RM283	
	5	RM163		3	RM81B		3	RM81B		3	RM81B	
	3	RM7		8	RM152		12	RM19		11	RM144	
	11	RM167		8	RM337		1	RM5		8	RM230	
	3	RM81B		3	RM338		1	RM283		3	RM60	
	8	RM126					1	RM128		2	RM208	
	5	RM153					1	RM315				
		1	RM297	85			64	8	RM230	80		

IM = interval mapping, BSA = Bulk segregant analysis, MR = Multiple regression, DA = Discriminant analysis.

PHT = Plant height; HD= Heading date; PAN= Panicle length.

Table 3.19 cont'd

Trait	MR			IM/CIM			BSA			DA		
	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification
TILL	1	RM283		1	RM297		4	RM280		3	RM81B	
	7	RM125		11	RM144		11	RM224		6	RM162	
	11	RM144		4	RM273		7	RM351		2	RM263	
	5	RM305		4	RM241		11	RM144		2	RM290	
	2	RM318		11	RM224		5	RM305		11	RM144	
	11	RM209		1	RM315		2	RM208		2	RM174	
	2	RM208										
	1	RM140	95			70			78			100
YLD	2	RM262		1	RM246		2	RM174		1	RM297	
	1	RM283		1	RM128		5	RM305		7	RM125	
	7	RM125		2	RM174		5	RM274		8	RM42	
	5	RM163		2	RM290		7	RM125		8	RM308	
	5	RM305		9	RM316		8	RM42		7	RM234	
	2	RM318		9	RM257		1	RM246		7	RM134	
	11	RM209		5	RM163		1	RM297		1	RM292	
	2	RM208		7	RM125		1	RM292				
1	RM246	82			75			88			100	
TGW	1	RM283		3	RM148		11	RM144		11	RM144	
	7	RM125		3	RM85		1	RM297		8	RM149	
	6	RM340		5	RM163		6	RM340		3	RM148	
	11	RM144		8	RM149		11	RM167		8	RM264	
	5	RM163		8	RM126		10	RM228				
	1	RM315		8	RM339		8	RM264				
	5	RM305		10	RM171		5	RM305				
	11	RM209	72	10	RM228	70			95			100

IM = interval mapping, BSA = Bulk segregant analysis, MR = Multiple regression, DA = Discriminant analysis.

TILL= Tillers per plant; YLD= Yield; TGW =1000-grain weight.

Table 3.20 Chromosomal location of SSR markers and % correct classification of DH lines using MR, IM, BSA and DA procedures, Colombia 2001, subpopulation 2

Trait	MR			IM/CIM			BSA			DA		
	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification
PHT	2	RM262	100	1	RM226	77	8	RM264	80	2	RM262	100
	6	RM340		1	RM297		2	RM262		8	RM264	
	11	RM224		4	RM131		2	RM263		7	RM125	
	1	RM292		2	RM208		7	RM234		6	RM340	
	7	RM234		2	RM213		1	RM292		2	RM163	
	6	RM253		8	RM230		11	RM144		1	RM144	
	2	RM71		8	RM264							
	8	RM42		11	RM144							
HD	6	RM340	80	4	RM349	61	6	RM190	72	2	RM262	100
	11	RM209		4	RM124		11	RM224		6	RM340	
	11	RM224		6	RM190		3	RM338		1	RM297	
	3	RM7		6	RM253		2	RM318		2	RM318	
	10	RM333		1	RM297		1	RM140		2	RM211	
	10	RM239		1	RM315		3	RM85		10	RM228	
	3	RM60		2	RM211		1	RM315		6	RM190	
	1	RM226		2	RM236		10	RM171				
	8	RM42					2	RM226				
3	RM85											
PAN	7	RM125	100	3	RM60	74	5	RM249	80	1	RM283	100
	10	RM171		3	RM81B		2	RM236		8	RM152	
	10	RM228		8	RM152		1	RM5		5	RM249	
	10	RM333		8	RM337		1	RM128		8	RM230	
	6	RM162		3	RM338		8	RM230		3	RM81B	
	1	RM315					6	RM162		6	RM30	
	2	RM236					1	RM315				
	3	RM81B										

IM = interval mapping, BSA = Bulk segregant analysis, MR = Multiple regression, DA = Discriminant analysis.

PHT = Plant height; HD= Heading date; PAN= Panicle length.

Table 3.20 cont'd

Trait	MR			IM/CIM			BSA			DA		
	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification
TILL	6	RM253		1	RM297		2	RM263		3	RM81B	
	2	RM208		11	RM144		1	RM140		6	RM162	
	1	RM315		4	RM273		2	RM290		2	RM263	
	1	RM140		4	RM241		11	RM144		2	RM290	
	11	RM144		11	RM224		2	RM318		11	RM144	
	6	RM162		1	RM315		2	RM208		2	RM174	
	4	RM280					6	RM162				
	6	RM190	80			64			78			100
YLD	6	RM340		1	RM246		1	RM226		1	RM283	
	2	RM6		1	RM128		5	RM305		7	RM125	
	5	RM153		2	RM174		5	RM274		6	RM340	
	1	RM5		2	RM290		7	RM125		11	RM144	
	6	RM253		9	RM316		5	RM249		5	RM163	
	8	RM308		5	RM163		1	RM128		1	RM226	
	1	RM246		7	RM125		1	RM283				
			82			72	2	RM318	80			100
TGW	2	RM106		3	RM148		11	RM144		2	RM262	
	2	RM240		3	RM85		1	RM297		10	RM228	
	1	RM165		5	RM163		6	RM340		6	RM340	
	11	RM202		8	RM149		11	RM167		11	RM209	
	1	RM5		8	RM339		10	RM228		11	RM144	
	1	RM297		10	RM228		8	RM264		5	RM249	
	6	RM253					5	RM305				
	3	RM85	90			62			75			100

IM = interval mapping, BSA = Bulk segregant analysis, MR = Multiple regression, DA = Discriminant analysis.

TILL= Tillers per plant; YLD= Yield; TGW =1000-grain weight.

Table 3.21 Chromosomal location of SSR markers and % correct classification of DH lines using MR, IM, BSA and DA procedures, Colombia 2001, subpopulation 3

Trait	MR			IM			BSA			DA		
	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification
PHT	2	RM208		1	RM226		7	RM234		7	RM125	
	3	RM7		1	RM297		11	RM144		8	RM264	
	10	RM171		4	RM131		8	RM264		2	RM262	
	3	RM338		2	RM208		2	RM263		1	RM140	
	8	RM339		2	RM213		6	RM190		5	RM163	
	6	RM190		8	RM230		1	RM246		2	RM208	
	6	RM253		8	RM264					8	RM230	
		100		11	RM144	77			80	2	RM263	100
HD	7	RM125		4	RM349		11	RM224		2	RM211	
	6	RM340		4	RM124		1	RM140		6	RM190	
	8	RM339		6	RM190		8	RM230		6	RM340	
	3	RM55		6	RM253		1	RM315		2	RM262	
	6	RM190		1	RM297		3	RM148		2	RM6	
	8	RM149		1	RM315		6	RM253		10	RM228	
	9	RM288		2	RM211		6	RM190		1	RM297	
	2	RM301		2	RM236		10	RM171				
	8	RM230					2	RM318				
10	RM239	100			80			72			100	
PAN	3	RM338		3	RM60		2	RM236		8	RM230	
	1	RM5		3	RM81B		1	RM5		3	RM81B	
	6	RM190		8	RM152		5	RM249		5	RM267	
	4	RM349		8	RM337		8	RM230		6	RM340	
	9	RM288		3	RM338		10	RM228		8	RM339	
	10	RM239	80			74	2	RM263	75	1	RM263	100

IM = interval mapping, BSA = Bulk segregant analysis, MR = Multiple regression, DA = Discriminant analysis
PHT = Plant height; HD= Heading date; PAN= Panicle length.

Table 3.21 cont'd

Trait	MR			IM			BSA			DA		
	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification	Chrom location	Markers	% correct classification
TILL	2	RM262		1	RM297		8	RM42		2	RM208	
	2	RM208		11	RM144		6	RM340		9	RM296	
	3	RM55		4	RM273		8	RM126		8	RM126	
	2	RM240		4	RM241		3	RM338		6	RM162	
	1	RM297		11	RM224		1	RM297		1	RM297	
	6	RM190		1	RM315		4	RM124		1	RM5	
	12	RM19	100			78	6	RM162	78			100
YLD	5	RM153		1	RM246		1	RM128		11	RM167	
	6	RM253		1	RM128		2	RM318		5	RM163	
	6	RM340		2	RM174		1	RM283		8	RM284	
	1	RM246		2	RM290		7	RM125		6	RM190	
	1	RM226		9	RM316		1	RM226		11	RM209	
	1	RM5		9	RM257		5	RM153		7	RM351	
	2	RM240		5	RM163		5	RM274				
			76	7	RM125	82	7	RM11	80			100
TGW	2	RM213										
	5	RM26		3	RM148		5	RM163		3	RM148	
	2	RM6		3	RM85		11	RM144		8	RM339	
	2	RM112		5	RM163		2	RM106		2	RM112	
	3	RM338		8	RM149		8	RM42		8	RM149	
	5	RM188		8	RM126		3	RM148		7	RM125	
	1	RM165		8	RM339		2	RM112		9	RM288	
	9	RM288		10	RM171		8	RM264				
	3	RM81B	72	10	RM228	68			80			100

IM = interval mapping, BSA = Bulk segregant analysis, MR = Multiple regression, DA = Discriminant analysis.

TILL= Tillers per plant; YLD= Yield; TGW =1000-grain weight.

selected two. These results show that the selection of markers that correctly classified lines was most accurate using the DA procedure. For heading date, at least 60 percent of selected markers were common among the population from Colombia compared to the populations from Crowley where methods selected different markers in each population.

In both subpopulations grown at Crowley, DA classified lines into groups of long and short panicle length with 100 percent accuracy. None of the other procedures could achieve 90 percent correct classification. Among the markers selected by DA, three were detected by interval analysis, RM337, RM60 and RM152 (Tables 3.17 and 3.18), while RM81B was selected both by BSA and MR in subpopulation 2. In population 1, RM60 was common to BSA and MR, (Table 3.17). The same trend was observed with the three subpopulations from Colombia location. DA showed the highest accuracy of classification, and two of the markers were common with QTL markers for this trait (Tables 3.19-3.21). Comparing among subpopulations, only IM and BSA selected markers that were common to at least two out of the three populations for classification of lines.

Accuracy of classification into groups of high and low tiller number was highest for DA with 5-6 markers and the interval analysis was lowest with six markers (Table 3.17 and 3.18). Two out of six markers detected by IM were among the top three selected by DA for the classification of lines (Table 3.18). BSA also had a marker which was also detected by interval analysis, but this ranked sixth among the selected markers. For the 3 subpopulations from the Colombia location, six markers classified lines into groups of high and low number of tillers per plant using the DA procedure, whereas BSA and MR required 6 to 7 markers to achieve not more than 78 percent accuracy (Tables 3.19-3.21). For this trait, four markers were common to both Colombia

and Crowley using MR. For the other methods, only two markers were common among those selected for classification of lines.

In subpopulation 1 from Crowley, 5 DA markers accurately classified DH lines into predefined groups of high and low grain yield while BSA and MR used 8 and 9 markers to achieve 92 and 88 percent correct classification (Table 3.17). The same trend was observed in subpopulation 2 (Table 3.18). RM128 and RM125 were detected by interval analysis and also selected by DA among the markers used to allocate lines into pre-defined groups of grain yield. In the three subpopulations from Colombia, DA performed best; however, BSA selected more markers among those detected by IM in subpopulations 1 and 3 than DA (Tables 3.19 and 3.21). In all populations, each method selected at least one of the markers detected by the traditional QTL method. Comparing between locations, DA selected three markers that were common, but not found together in one subpopulation.

Classification of DH lines into groups of high and low grain weight was most accurate using DA (100 percent with only 5 markers), but BSA also achieved high accuracy of 94 and 95 percent with 8 markers (Tables 3.17 and 3.18). Three markers were common to interval analysis and DA among the predictors of group membership of the DH lines. In the three subpopulations from Colombia, DA classified with the highest accuracy of 100 percent (Tables 19-21). BSA had 95 percent in subpopulation 1 and 85 percent in population 2

3.3.6. A Comparison of Chromosomal Locations of Markers Selected by MR, BSA, IM and DA Methods

The genetic map constructed for this cross allowed the comparison among the four methods, with regard to the location of selected markers used for classification (Figures 3.4 and 3.5). Clustering was observed at five locations on the chromosomes where all four methods pointed to the same region of the genome associated with the different traits (Fig. 3.4). On chromosome 3,

all four methods clustered around RM60 for the classification of lines into groups of long and short panicles. Another cluster was formed where three of the four methods pointed to the same region of the chromosome. On chromosome 11, DA and BSA clustered with interval analysis in the classification of lines into groups of high and low tillering, and also tall and short plants. Two methods, DA and BSA pointed to the same region as interval analysis for heading date (*hd4-2*) on chromosome 4. Moreover, DA and BSA also clustered with IM to identify chromosomal regions for grain weight on chromosome 3 and grain yield on chromosome 7. For panicles per plant and tillers per plant, DA and interval analysis pointed to the same regions on chromosomes 8 and 10, respectively. On chromosomes 2 and 8, BSA and IM identified the same chromosomal regions for panicle length.

However, in some cases, other methods point to regions that were not detected by IM. For example, on chromosome 6, DA and BSA selected RM253 for classification of plant height and heading date whereas IM did not detect this position. Also, RM283 on chromosome 1 was selected for classification of panicle length by MR and BSA, but not selected by the other two methods. Only DA identified chromosomes 1 and 5 for grain yield.

In Figure 3.5 (Colombia), all four methods clustered on chromosome 11 for tillers per plant, grain yield on chromosome 7, plant height on chromosome 2 and panicles per plant on chromosome 3. For three traits, plant height, heading date and tillers per plant, DA pointed to the same chromosomal region as BSA on chromosomes 8 and 1 while on chromosome 6, DA and MR identified the same region. DA was in agreement with IM in detecting markers associated with panicles per plant and grain weight on chromosome 3 (RM60 and RM148 respectively) (Figure 3.5). Comparing the map from Crowley and Colombia, there was more clustering of the four methods in the latter than the former for tillers per plant, grain yield, panicle length, and plant

height. A general agreement was found on chromosome 11 where the same chromosomal region was identified for plant height and tillers per plant, irrespective of location of the experiment. In the population from Colombia, there was clustering of two methods on chromosome 2 for plant height (DA and BSA), grain yield (MR) and tillers per plant (MR). These were not detected at Crowley and IM did not detect this region. Likewise, *hd4-2* on chromosome 4 was detected by BSA and DA pointing to the same chromosomal region around RM124, but this region was not detected in Colombia.

3.4. Discussion

Various studies have shown that plant height is one of the most important traits related to yield potential in rice (Moncada et al. 2001; Yu et al. 2002). Plant height was positively and significantly correlated with grain yield at Colombia. There was transgressive segregation for plant height, and the result of detected QTLs in Table 3.4 support the theory that both parents donated positive alleles which enhanced the expression of this trait. However, tall varieties are easily prone to lodging. There is probably a critical height for optimum grain yield in rice. The wide range of plant height offers good opportunity for selection in this population. QTLs detected for this trait at Crowley were not detected in Colombia which indicates the complexity of this trait.

Heading date has been reported to be the most important trait for the adaptation of rice to different cultivation areas and crop seasons (Yano et al. 1997; Yu et al. 2002). Heading date was found to vary between the two experiments indicating the effect of location and cropping ecology on the genotypes used in these studies. The flowering date of the Caiapo parent was prolonged for only three days at Crowley compared to Colombia. IRGC 103544 however, exhibited a dramatic environmental response as this parent did not flower at the Louisiana location. Tillers per plant were improved in this cross as many of the DH lines tillered more than the Caiapo parent. This

result indicates the potential for an increase in tiller number through interspecific hybridization of these two species. The significant positive correlation between tillers per plant and grain yield suggests that grain yield can be improved by increasing the number of tillers per plant. The number of tillers per plant has been shown to be an important component of grain yield in rice (Dingkuhn et al.1996). It is interesting to note that some of the progenies yielded more than 1000 kg per ha at Crowley in spite of the fact that they were not bred for this environment. The transgressive segregation for heavier grains indicated the potential for increased sink capacity in this cross. One of the main advantages of interspecific crosses is the possibility to introgress genetic variability (Brondani et al. 2002). The transgressive segregation observed for most of the traits in this study is a clear evidence of the favorable effect of such introgressions. Other studies involving interspecifics have reported transgressive segregation and the introgression of yield enhancing loci from the wide relative of cultivated rice (Martinez et al. 1997; Moncada et al. 2001; Brondani et al. 2002).

Forty-three QTLs were detected at the two locations altogether for the six traits evaluated in this study. In Colombia, 23 QTLs were detected out of which five had been reported previously (Brondani et al. 2002; Xiao et al. 1996; Yano et al. 1997). Forty-two percent of the detected QTLs originated from IRGC 103544 while 58 percent were from Caiapo. At Crowley however, 20 QTLs were detected out of which 9 corresponding regions were previously reported by others. Fifty percent of these QTLs were contributed by alleles from IRGC 103544. It is interesting to note that QTLs for grain yield were contributed by the IRGC 103544 in Colombia while the only 2 QTLs detected in Crowley for grain yield were contributed by Caiapo. Even though there were no yield-improving alleles detected from IRGC 103544 in the experiment at Crowley, the finding of transgressive segregation for this trait showed that yield improving alleles must have been

introgressed by the IRGC 103544 parent, but were not detected at Crowley. IRGC 103544 contributed at least one allele for five of the six traits evaluated at both locations. The differential performance of the DH lines at these two locations indicated that there are some genotypes adapted to each location and also some that had wide adaptability to both locations.

The first step in the discriminant analysis procedure was to classify 312 DH lines with and without the assumption of population structure. The software program Structure was used to assign the 312 DH lines into subpopulations based on their allelic frequencies. The analysis of structure was carried out with this software which uses a Markov Chain Monte Carlo (MCMC) algorithm to sample from the posterior distribution of the subpopulation allelic frequencies. This analysis was carried out to determine if cryptic structure existed within the 312 DH lines and to enable the use of more homogenous groups for the marker assisted classification of the lines. The result here showed that there were three subpopulations among the DH material.

The performance of the discriminant procedure to classify lines with high degree of accuracy has shown that this method could be used to select an informative set of markers to complement QTL mapping. Similar results were reported by Capdevielle (2001) who found that DA was efficient in classifying rice lines into groups of high and low cold tolerance. The subdivision of DH lines into subpopulations based on the result of the “Structure” program helped remove the confounding effect of associations which might be due to clustering of the lines themselves and not necessarily association of lines with the quantitative trait of interest. In this study, when the effect of population structure was removed, classification efficiency improved greatly. The results presented in Tables 3.5-3.16 show the feasibility of selecting a few set of markers using the SAS Stepdisc procedure for high degree of classification accuracy.

This result has some relevant implications for breeding. The maximum number of markers to be selected was limited to 20 for practical reasons. The breeder would save a lot of cost by using a few selected and informative markers rather than working with many markers, few of which are informative. The few markers selected may be applied to hundreds of lines in a germplasm improvement program to save cost and time. In Tables 3.5-3.16, as few as 5 markers achieved 100 percent correct classification of lines into groups of low and high trait values when adjustment was made for population structure.

To compare other methods of associating markers with traits, markers selected by multiple regression, bulk segregant, and interval analyses were used for classification of lines into predefined groups using the K-nearest neighbor algorithm. From the results in Tables 3.17-3.21, DA produced the highest accuracy of classification for all the traits studied. Therefore depending on the level of genetic linkage involved in associations, DA may be very useful in the selection of parents for producing populations for mapping. In addition, the discriminant analysis may be used as a complementary tool with QTL for marker assisted selection.

The use of a pooled population of predefined groups at the two tails of the distribution for each of the traits to explore marker-QTL associations has been used to increase statistical power which is often low in many mapping and association studies (Lebowitz et al. 1987). In this study we found that the BSA method selected some of the markers found to be associated with QTL for the traits evaluated. Nevertheless, BSA identified other markers that were equally identified by both DA and MR procedures.

Zeng (1994) cautioned that the direct use of multiple regression analysis is not a suitable way for mapping QTL because a partial regression coefficient is a biased estimate of the relevant QTL effect. Here we found that MR selected the least number of markers associated with QTL

even though markers were selected on the basis of their maximum R^2 . However, determining the reason why MR selected few QTL markers is beyond the scope of this work. In view of the robustness of DA procedure, it should be possible to use this method to select the most informative among markers known to be associated with QTL for allocating germplasm into predefined groups of agronomic interest. The K-NN algorithm for the classification of lines into groups is robust. One of its advantages is that this procedure is non-parametric and so does not require the assumption of normality. The method used to construct the classification was built on the assumption that lines that are close together in the matrix defined using marker information will have the same classification. This reflects a similar tendency for the posterior distribution of groups, and this is why the value of k (neighbors) is often chosen to be odd (k=1 in our procedure) so that ties were less likely to occur.

A comparison of positions of markers selected by the three methods with those selected by interval analysis was necessary to determine if there were similarities and differences among the markers selected by the various methods. From Figures 3.4 and 3.5, it was found that many of the markers selected by the DA and other methods pointed to the region of QTL for the different traits. However, DA selected more markers that were associated with QTL positions than BSA or MR. In this analysis, the top four selected markers for classification of lines into groups of low and high trait values were considered. Other markers were selected which mapped to different chromosomal locations on the map. However, the optimum number of markers should be considered for practical purposes. The DA method therefore proved to be more efficient than the other three methods in the selection of markers associated with these traits in the DH population. The percent correct classification was also highest for the DA procedure.

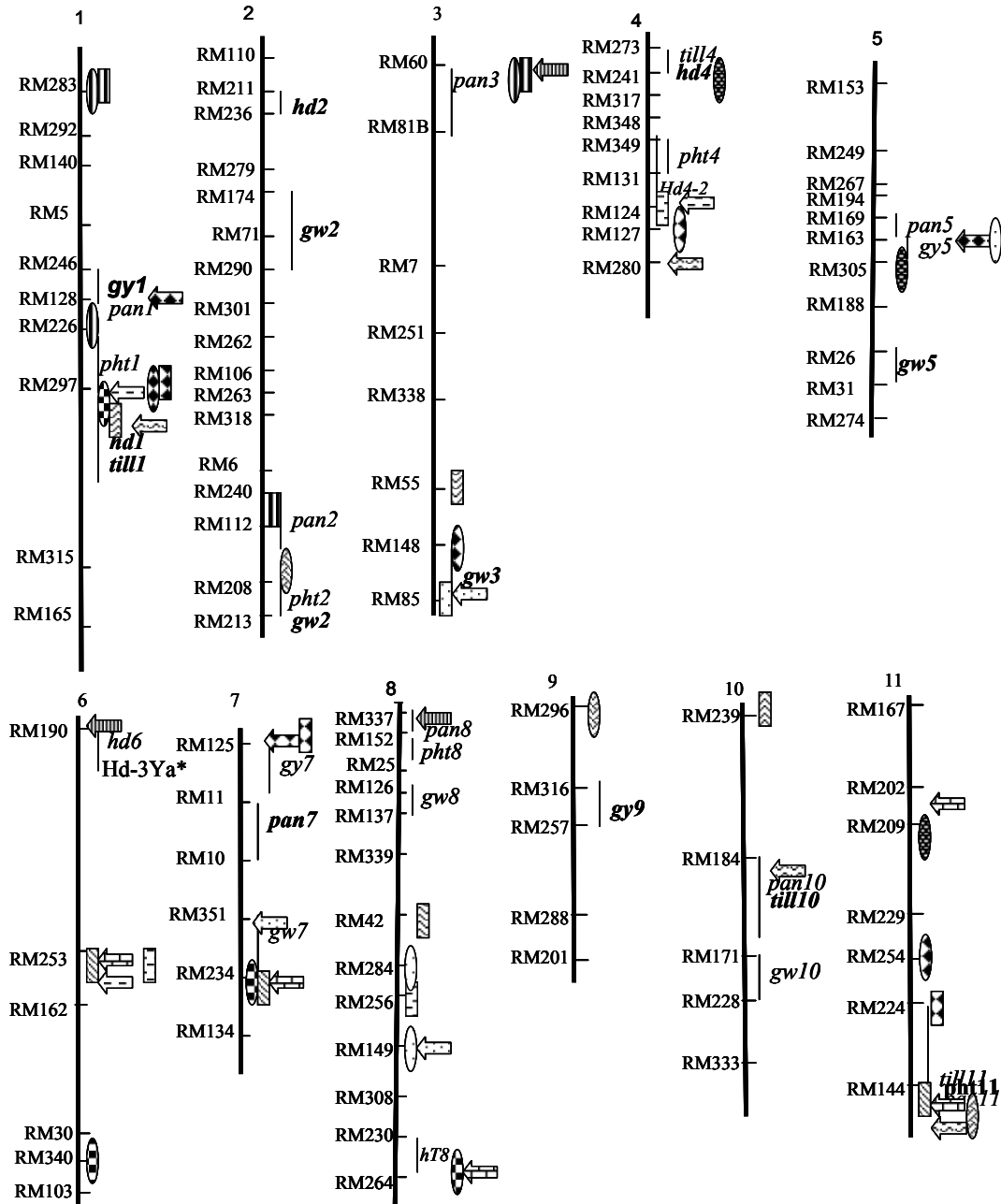
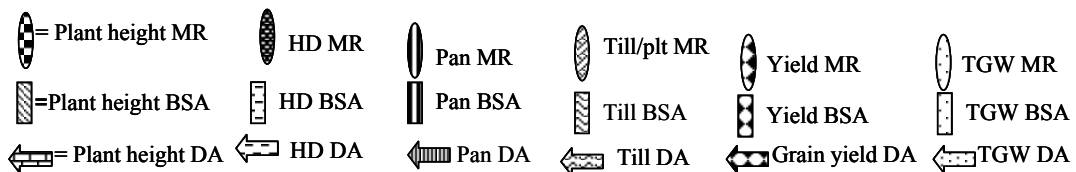


Fig. 3.4. Chromosomal locations of markers selected by MR, BSA, DA and Interval Analysis among 312 DH lines derived from *O. sativa* (Caiapo) x *O. glaberrima* (AC# IRGC103544), from Crowley. *ht*: Plant height ; *dh*: Days to heading; *till*: Tillers per plant; *pan*: Panicle length; *gy*: Grain yield; *gw*: 1000-grain weight.



Size of icons does not represent confidence interval

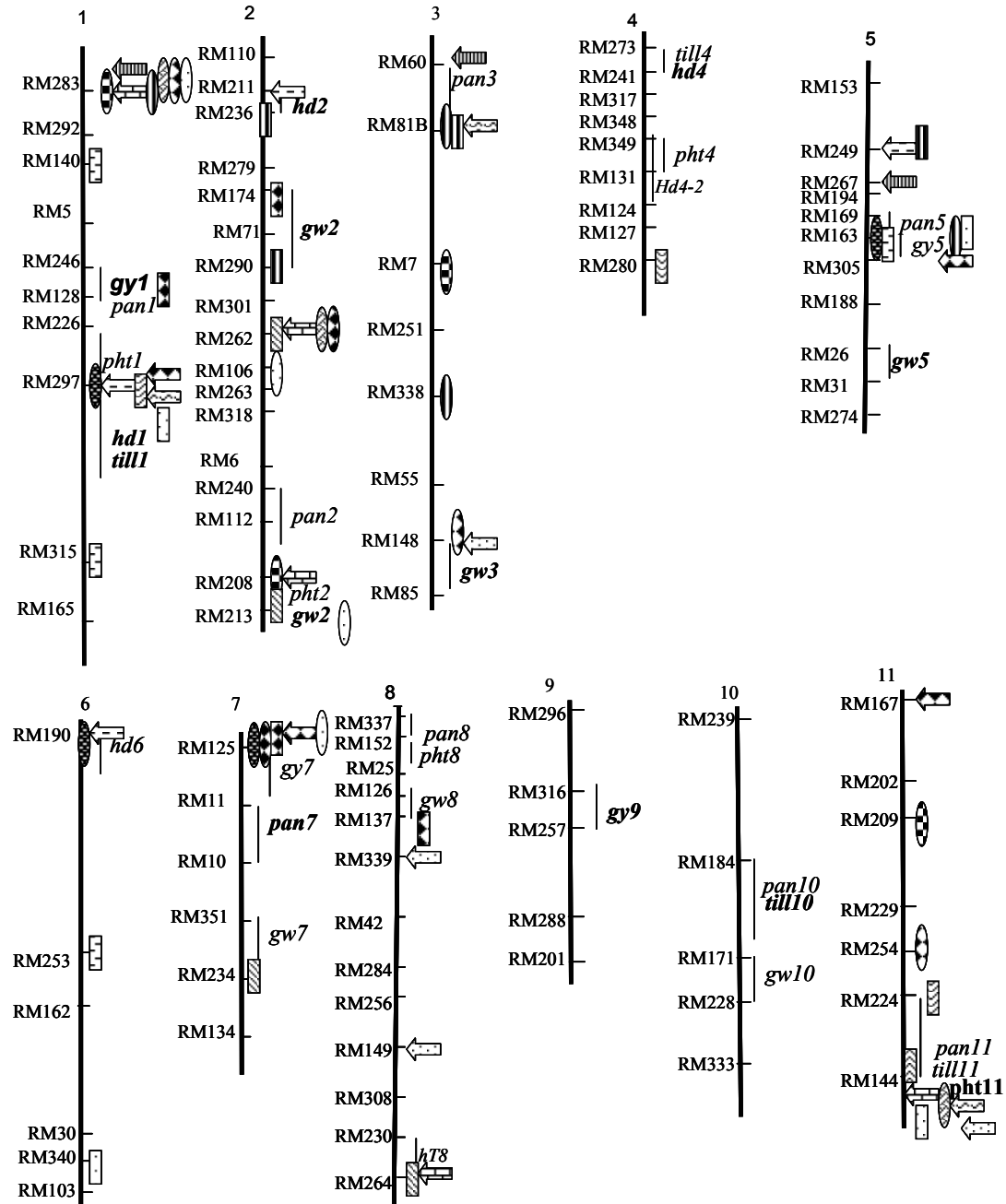
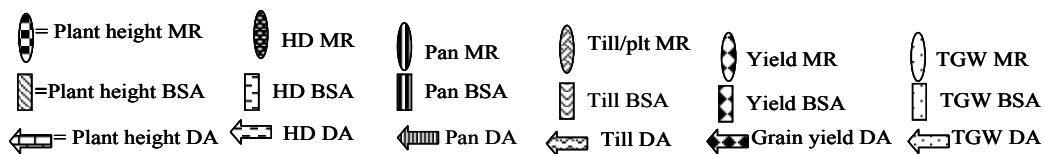


Fig. 3.5. Chromosomal locations of markers selected by MR, BSA, DA and Interval Analysis among 312 DH lines derived from *O. sativa* (Caiapo) x *O. glaberrima* (AC# IRGC103544), from Colombia. *ht*: Plant height ; *dh*: Days to heading; *till*: Tillers per plant; *pan*: Panicle length; *gy*: Grain yield; *gw*: 1000-grain weight.



Size of icons does not represent confidence interval

In this study, chromosomal regions for tillers per plant and plant height were detected by all four methods (MR, BSA, IM and DA) on chromosome 11. Other researchers (Brondani et al. 2002, Yu et al. 2002) have reported QTL for these traits in the same region. It is also worthy to note that DA identified the same chromosomal region reported by Zhuang et al. (1997) for grain yield. The location of QTL for days to heading reported by Yano et al. (1997) on chromosome 6 coincided with that identified by DA in this study. The four methods formed more clusters in Colombia than at Crowley. However, there were cases of common clusters. For example on chromosome 11 clustering for tiller number and plant height were common to both Colombia and Crowley. On chromosome 3, clustering of all four methods for panicle length in Crowley was a little bit modified in Colombia. Except for DA, the other methods identified a marker 14.5 cM from RM60.

In this study, a new QTL, *gy9*, detected on chromosome 9 was found to be associated with grain yield. It is interesting to note that alleles from the *O. glaberrima* parent, IRGC 103544, were associated with increase in grain yield at this locus. The introgression of yield enhancing alleles from wild relatives of rice, *O. rufipogon* (Moncada et al. 2001) and *O. glumaepatula* (Brondani et al. 2002) and *O. glaberrima* in this study is an interesting development in rice breeding. Marker-assisted selection would be very useful to speed up the process of introgressions from these species (Brondani et al. 2002).

To develop a practical marker-assisted classification procedure for selection of lines based on markers, one critical consideration should be the minimum number of markers required to achieve a high percentage of correct classification into high and low groups. From the results obtained in this study, when using highly differentiated groups, five to 10 markers would be an optimum number of markers required to achieve the highest accuracy of 100 percent. The fact

that DA markers pointed to the same direction as markers associated with QTL by interval analysis indicated that the selection had a genetic basis. Therefore this procedure can be used as a complementary tool to QTL in the association of markers with traits for breeding purposes.

3.5 References

- Anderson JA, Churchill GA, Sutriquet JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. *Genome* 36: 181-186
- Balzarini M, Aluko GK, Capdevielle FM, Oard JH (2000) Towards allocation of rice lines into predefined groups: discriminant analysis using molecular marker information. In: Proceedings of the 28th Rice Technical Working Group, Biloxi, MS p.51
- Barrella W, Petrere M (2003) Fish community alterations due to pollution and damming in Tiete and Paranapanema rivers (Brazil). *River Res Appl* 17 (1): 59-76
- Basten CJ, Weir BS, Zeng ZB (1994) Zmap- a QTL cartographer *Computing Strategies and Software Proc 5th Congr on Genetics Applied to Livestock Production: Guelph, Ontario*
- Basten CJ, Weir BS, Zeng ZB (1994) Zmap- a QTL cartographer: a reference manual and tutorial for QTL mapping Department of Statistics, North Carolina State University, Raleigh, North Carolina
- Bennett ST, Lucassen AM, Gough SGL, Powell EE, Undlien DE, Pritchard LE, Merriman ME, (1995) Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nat Genet* 9: 284-292
- Bishop, D. T., Cannings, C., Scholnick, M. & Williamson, J. (1983) in *Statistical Analysis of DNA Sequencing Data*, ed. Weir, B. S. (Dekker, New York), pp. 181-200
- Blum K, Sheridan PJ, Montgomery A, Ritchie T, Jagadeeswaran P, Nogami H, Briggs AH, Cohn JB (1990) Allelic association of human dopamine-D2 receptor gene in alcoholism. *JAMA-Journal of American medical association* 263:2055-2060
- Boehnke M (1994) Limits of resolution of genetic linkage studies: implications for positional cloning of human disease genes: *Am J Hum Genet* 55: 279-290
- Briscoe D, Stephens JC, O'Brien SJ (1994) Linkage disequilibrium in admixed populations – Applications in gene mapping. *Journal of Heredity* 85:59-63
- Brondani C, PHN Rangel RPV Brondani (2002) QTL mapping and introgression of yield-related traits from *Oryza blumaepatula* to cultivated rice (*Oryza sativa*) using microsatellite markers. *Theor Appl Genet* 104:1192-1203

Bylund D, Samskog J, Markides KE, Jacobsson SP (2003) Classification of lactate dehydrogenase of different origin by liquid chromatography-mass spectrometry and multivariate analysis: Journal of the American society for mass spectrometry 14 (3): 236-240

Cao D, Oard JH (1997) Pedigree and RAPD-based DNA analysis of commercial U.S. rice cultivars. Crop Sci. 37:1630-1635

Capdevielle FM, Aluko GK, Balzarini M, Oard JH (2000) Application of molecular markers and discriminant analysis to identify rice lines with contrasting phenotypes for agronomic traits. In Fourth International Rice Genetics Symposium, International Rice Research Institute, Philippines, p.216

Causse MA, Fulton TM, Cho YG, Ahn SN, Chunwongse 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 interspecific backcross population. Genetics 138:1251–1274

Charkraborty R and Smouse PE (1988) Recombination of haplotypes leads to biased estimates of admixture proportions in human populations. Proc Natl Acad Sci USA, 85, 3071-4

Chagnon YC, Perusse L, Bouchard C (1998) The human obesity gene map: the 1997 update. Obes Res 5:76-92

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

Cottrell, J. E., Forrest, G. I. & White, I.M.S. (1996) The use of random amplified polymorphic DNA markers to identify and estimate the relatedness of clones belonging to the genus *Populus*. Bot. J. of Scotland. Cited in Bishop 1983.

Dingkuhn, M, David E Johnson, Monty P Jones, and A Sow (1996) The Physiological basis of developing low- management upland rice plant types Interspecific Hybridization: Progress and prospects 81-102

Dong NV, Subudhi PK, Luong PN (2000) Molecular mapping of a rice gene conditioning thermo sensitive genic male sterility using AFLP, RFLP, and SSR techniques. Theor Appl Genet 100 (5): 727-734

Ebdon JS, Petrovic AM and Schwager SJ (1998) Evaluation of discriminant analysis in identification of low-and-high-water use Kentucky bluegrass cultivars. Crop Sci 38:152-157

Eisman JA (1995) Vitamin-D-receptor gene alleles and Osteoporosis – An affirmative view. Journal of bone and mineral research 10 (9):1289-1293

Excoffier L, Smouse PE and Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial DNA restriction data. *Genetics* 131:479-491

Fahima T, Roder MS, Wendehake K, Kirzhner VM, Nevo E (2002) Microsatellite polymorphism in natural populations of wild emmer wheat, *Triticum dicoccoides*, in Israel. *Theor Appl Genet* 104:17-29

Falconer, DS and Mackay TFC (1996) *Introduction to Quantitative Genetics*, 4th edition Longman, Harlow, UK

Francisco O, Ciavatta C, Montecchico D, Sanchez-Cortes S, Gessa C (2003) Quantitative estimation of peat, brown coal and lignite humic acids using chemical parameters, H-1-NMR and DTA analyses: *bioresource technology* 88 (3): 189-195

Giovannoni JJ, Wing RA, Ganai MW, Tanksley SD (1991) Isolation of molecular markers from specific chromosomal interval using DNA pools from existing mapping population. *Nucleic acids Research* 19 (23): 6553-6558

Gnanadesikan R (1987) Report: Panel on discriminant analysis, classification and clustering Published by the Committee on Applied and Theoretical Statistics, Washington

Hand DJ (1997) *Construction and assessment of classification* John Wiley and Sons, Chichester

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiyama H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T (1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics* 148:479-494

Hua JP, Xing YZ, Xu CG, Sun LX, Yu SB, Zhang Qifa (2002) Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantages for performance. *Genetics* 162:1885-1895

Huff, DR, Peakall, R Smouse, PE (1993) RAPD variation within and among natural populations of outcrossing buffalograss *Buchloe dactyloides* (Nutt.) Engelm. *Theor Appl Genet* 86: 927-934

Hwang SI, Powers SE (2003) Using particles-size distribution models to estimate soil hydraulic properties. *Soil Science Soc Amer j* 69 (4): 1103-1112

Hwang D, Alevizos I, Schmitt WA, Misra J, Ohshima H, Todd R, Mahadevappa M, Warrington JA, Stephanopoulos G, Wong DT, Stephanopoulos G (2003) Genomic dissection for characterization of cancerous oral epithelium tissues using transcription profiling: *oral oncology* 39 (3): 259-268

Johnson, RA and Wichern, DW (1992) *Applied multivariate statistical analysis* 3rd ed Prentice-Hall Englewood Cliffs, NJ

Jonathan K Pritchard, Matthew Stephens and Peter Donnelly (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959

Lachenbruch, PA and MR Mickey (1968) Estimation of error rates in discriminant analysis. *Technometrics* 10(1): 1-11

Lander ES, Green P, Abrahamson J, Barlow A, Daley M (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174-181

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

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. *Science* 265:2037-2048

Lebowitz, RL, Scoller, M and Beckmann, JS (1987) Trait-based analysis for the detection of linkage between loci and quantitative trait loci in cross between inbred lines. *Theor Appl Genet* 73:556-562

Lentini Z, P Reyes, CP Martinez, WM Roca(1995) Androgenesis of highly recalcitrant rice genotypes with maltose and silver nitrate *Plant Science* 110:127-138

Lheureux F, Carreel F, Jenny C, et al (2003) Identification of genetic markers linked to banana streak disease expression in inter-specific *Musa* hybrids. *Theor Appl Genet* 106 (4): 594-598

Lincoln S, Daly M, Lander E (1992) Constructing genetic linkage maps with MAPMAKER/EXP Whitehead Institute Technical Report [http://www-genomewimite.edu/ftp/distribution/software](http://www.genomewimite.edu/ftp/distribution/software)

Liu CJ, Witcombe JR, Pittaway TS, Nash CT, Busso CS, Gale MD (1994) An RFLP-based genetic map of pearl millet (*Pennisetum glaucum*). *Theor Appl Genet* 89:481-487

Liu BH (1997) *Statistical genomes: Linkage mapping and QTL analysis*. CRC press, Boca Raton, Florida

Liu G, Lu G, Zeng L, et al (2002) Two broad-spectrum blast resistance genes, Pi9(t) and Pi2(t), are physically linked on rice chromosome 6. *Mol Genet Genomics* 267 (4): 472-480

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

Lynch, M and Walsh, B (1998) *Genetics and Analysis of Quantitative Traits* Sinauer Associates, Sunderland, Massachusetts, pp 980

- Mammadov JA, Zwonitzer JC, Biyashev RM (2003) Molecular mapping of leaf rust resistance gene Rph5 in barley. *Crop Sci.* 43 (1): 335-346
- Mazur B, Krebblers E, Tingey S (1999) Gene discovery and product development for grain quality traits. *Science* 285 (5426): 372-375
- McCouch SR, Cho YG, Yano M, Paul, Blinstrub M (1997) Report on QTL nomenclature. *Rice Genet Newsl* 14: 11-13
- McCouch SR, Chen XL, Panaud O, Temnykh S, Xu YB, Cho YG, Huang N, Ishii T, Blair M (1997) Microsatellite marker development, mapping and applications in rice genetics and breeding. *Plant Molecular Biology* 35 (1-2): 89-99
- Mei HW, Luo LJ, YingCS, Wang YO (2003) Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred population and two testcross populations. *Theor Appl Genet* 107:89-101
- Michelmore RW, Paran I, Kesseli RV(1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis-a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the national academy of sciences of the united states of America* 88 (21): 9828-9832
- Moncada P, CP Martinez, J Borrero, M Chatel, H Gauch Jr, E Guimaraes, J Tohme, SR McCouch (2001) Quantitative trait loci for yield and yield components in an *Oryza sativa* x *Oryza rufipogon* BC₂F₂ population evaluated in an upland environment. *Theor Appl Genet* (2001) 102:41-52
- Morgante M Olivieri AM: PCR-amplified microsatellite as markers in plant genetics. *Plant J* 1: 175-182 (1993)
- Morrison NA, Qi JC, Tokita A, Kelly PJ, Crofts L, Nguyen TV, Sambrook PN, Eisman JA (1994) Prediction of bone-density from vitamin-D alleles. *Nature* 367 (6460):284-287
- Neter J, Kutner MH, Nachtsheim CJ, Wasserman W (1996) *Applied Linear Statistical Models* Richard D Irwin, Inc Burr Ridge, Illinois 1408 pp
- O'Brien, SJ (Editor) (1993) *Genetic Maps*, pp 661-679 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Panaud, O, 1992 Mise en oeuvre d'une methode de marquage nonradioactif de l'AND pour l'etude des RFLP chez le Riz: cartographie du genome et suivi des introgressions entre *Oryza sativa* et *O branchyantha* PhD Thesis, Universite de Paris, SUD, Centre, D'Orsay
- Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335:721-726

Paterson, AH, TH Lan, KP Reischmann, C Chang, and YR Lin (1996) Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence. *Nat. Genet.* 14:380-382

Pato CN, Macciardi F, Pato MT, Verga M, Kennedy JL, (1993) Review of the putative association of Dopamine D2 receptor and alcoholism- A meta analysis. *American Journal of Medical Genetics* 48:78-82

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

Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945-959

Quarrie SA, Lazic-Jancic V, Kovacevic D, Steed A, Pekic S (1999) Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. *Journal of Experimental Botany* 50 (337): 1299-1306

Rasmuson 1933 A contribution to the theory of quantitative character inheritance. *Hereditas* 18: 245-261

SAS Institute (2000) SAS/SAT User's Guide, Version 8 SAS publishing, Cary, NC

Schneider, S, Kueffer, JM, Roesslie, D, and Excoffier, L (2000) Arlequin ver 2,0: A software for population genetic data analysis *Genetics and Biometry Laboratory, University of Geneva, Geneva*

Semagn K, Bjornstad A, Stedje B and Bekele E (2000) Comparison of multivariate methods for the analysis of genetic resources and adaptation in *Phytolacca dodecandra* using RAPD. *Theor Appl Genet* 101:1145-1154

Spielman RS, and Ewens WJ (1996) The TDT and other family-based tests for linkage disequilibrium and association. *American Journal of Human Genetics* 59 (5): 983-989

Stuber, CW and Edwards, MD (1986) Genotypic selection for improvement of quantitative traits in maize using molecular marker loci In: *Proceedings of the 41st Annual Maize and Sorghum Research Conference American Seed Trade Association, Chicago, Illinois, pp 40-83*

Talbot M (1997) Resource allocation for selection systems In *Statistical methods for plant variety evaluation*, Kempton RA and Fox PN Eds, Chapman Hill, London

Tanksley SD, Bernachi D, Beck-Bunn T, et al (1998) Yield and quality evaluations on a pair of processing tomato lines nearly isogenic for the Tm2(a) gene for resistance to the tobacco mosaic virus. *Euphytica* 99 (22): 77-83

- Thoday, JM (1961) Location of polygenes Nature (London) 191:368-370
- Thornsberry, JM, Goodman, MM, Doebley, J, Kresovich, S, Nielson, D, and Buckler, ES, IV (2001). *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 28:286-289
- Van Gestel S, Houwing-Duistermaat JJ, Adolfsson R, van Duijn CM, Van Broeckhoven C (2000) Power of selective genotyping in genetic association analyses of quantitative traits. Behav Genet 30(2):141-6
- Virk PS, Ford-Lloyd BV, Jackson MT, and Newbury HJ (1995) Use of RAPD for the study of diversity within plant germplasm collections. Heredity 74:170-179
- Virk, PS, Ford-Lloyd BV, Jackson MT, Pooni HS, Clemeno TP, Newbury, HJ (1996) Predicting quantitative variation within rice germplasm using molecular markers. Heredity 76:296-304
- Virk PS, Newbury HJ, Jackson MT and Ford-Lloyd BV (2000) Are mapped markers more useful for assessing genetic diversity? Theoretical and Applied Genetics 100:607-613
- Vos P, Hogers R, Bleeker M, Reijans M, Vandeleer T, Hornes M, Fritjers A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP – A new technique for fingerprinting. Nucleic Acids Res 23:4407-4414
- Wang GL, Paterson AH (1994) Assessment of DNA pooling strategies for mapping of QTLs. Theor Appl Genet 88 (3-4): 355-361
- Wang GL, Mackill DJ, Bonman JM, McCouch SR, Champoux MC, Nelson RJ (1994) RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar. Genetics 136: 1421-1434
- Williams, JGK, Kubelik, AR, Livak, KJ, Rafalski, JA, and Tingey, SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18: 6531-6535
- 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
- Yu SB, Li JX Xu CG Tan YF Li XH Qifa Zhang (2002) Identification of quantitative trait loci and epistatic interactions for plant height and heading date in rice. Theor Appl Genet 104: 619-625
- Zeng ZB (1993) Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. Proc Natl Acad Sci USA 90:10972-10976
- Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468

CHAPTER 4: SUMMARY AND CONCLUSIONS

A non-parametric multivariate procedure called Discriminant Analysis was evaluated as a potential tool for marker assisted selection in rice breeding. This procedure was compared with the standard QTL, multiple regression, and bulk segregant analyses as tools for predicting marker trait associations in a population of 312 doubled haploid lines. The objectives of this study were: (1) to identify QTLs linked to economically important agronomic traits among doubled haploid lines (DH) derived from the interspecific cross *Oryza sativa* (L.) and the African rice *O. glaberrima* (Steud.), (2) identify potential molecular markers from *O. glaberrima* that contribute to agronomic traits measured, (3) determine the effect of population structure on power and precision to detect markers associated with the measured agronomic traits, (4) evaluate DA for correct allocation of DH lines into pre-defined groups based on their microsatellite profiles and (5) compare genetic map locations of DA-selected markers with those markers detected by traditional QTL, BSA and multiple regression approaches. An additional objective was to map the QTLs for milling, eating and cooking qualities of rice using this population of DH lines derived from the interspecific cross between African (*O. glaberrima*, IRGC 103544) and Asian rice (*O. sativa*, cv. Caiapo). Finally, detection of *O. glaberrima*-derived QTLs and their comparison with previous studies will shed light on the potential of African rice as new sources for enhanced grain quality traits.

Extensive segregation distortion was observed among the molecular markers as 39 of the 100 microsatellite markers showed aberrant segregation. In spite of this extensive segregation distortion, 28 QTLs were detected for 10 grain quality traits. Significant digenic interactions were detected in percent head rice, amylose content, grain width, milled rice and rice bran percent. Alleles from IRGC103544 were found that increased percent rice bran, alkali spreading score,

protein content, grain width and grain size measured by the length/width ratio. Despite the phenotypic inferiority of *O. glaberrima* and low yield potential, transgressive segregants were produced for most of the grain quality traits. Twenty three putative QTLs were detected in Colombia in 2001 while 20 were detected at Crowley, LA in 2002 for six agronomic traits (plant height, heading date, number of tillers per plant, panicle length, and grain yield and 1000-grain weight). IRGC 103544 contributed at least one allele for five of the six traits evaluated in both experiments.

In the evaluation of discriminant analysis as a classificatory procedure, phenotypic differences between the high and low groups directly affected the percent correct classification. When the lines were highly differentiated, as few as 5-10 markers were selected by this procedure which gave 100 percent correct classification of the DH lines into groups of contrasting phenotypes. Adjusting for population structure improved the performance of DA as a predictive tool for classification. Not only were few markers selected that differentiated clearly between groups, classification accuracy increased remarkably. For the highly differentiated groups, it took 5 fold less number of markers to achieve the same degree of accuracy when structure was not assumed compared to when it was ignored. Markers selected through DA were either the same as those found to be associated with QTLs or pointing to the same regions where QTLs have been detected. This indicated that the DA procedure had genetic basis. When compared with other procedures (bulk segregant and multiple regression), DA markers were more accurate in classifying lines into predefined groups of high and low trait values. The findings of this research have interesting implications for rice breeding. The DA procedure could be applied to germplasm improvement for the selection of parental lines for crossing in a breeding program.

The DA procedure proposed in this study was able to classify DH lines into predefined groups of agronomic interest with a high level of accuracy. These results indicate the suitability of DA as a complementary procedure to traditional MR, BSA and Interval Analyses in the detection of markers associated with agronomic traits of interest.

We do not suggest that traditional methods should be discarded, since each method has advantages and limitations. Future research should focus on predictive discriminant analysis whereby an unrelated rice line can be classified, based on its molecular profile, into a group of agronomic interest using the DA model developed in this study. Selection can then be made as early as possible during crop growth. This will be of immense advantage to rice breeders, since population development would not be a prerequisite for this selection process. Moreover, substantial resources and time would be saved.

VITA

Gabriel Kayode Aluko was born at Ikole Ekiti, Ekiti State of Nigeria, on July 17, 1954. He attended MacJob Grammar School Abeokuta, Ogun State of Nigeria, from 1967 to 1971. After graduating from high school, he attended the University of Ibadan from 1975 to 1978 for his bachelor's degree in crop science. He worked in the Ondo State Ministry of Agriculture from 1980 to 1981 after which he won the Federal Government scholarship to study for his Master of Science in agronomy in 1981. After completing his master's degree in agronomy in 1982 he returned to the Ministry of Agriculture, and two years later he won the Commonwealth scholarship to pursue Master of Agricultural studies in Queensland University, Brisbane, Australia, in 1984.

Kayode returned from Australia to continue his work in the Ondo State ministry of agriculture where he served as Zonal coordinator of agricultural activities in three local government areas of Ondo State, Nigeria. He has worked in two international research institutes namely, International institute of tropical agriculture in Ibadan, Nigeria, from 1989 to 1994 as research associate to the cowpea breeder, and also at the West Africa Rice Development Association in Bouake, Cote d'Ivoire, as assistant to the upland rice breeder from 1994 to 1999. Kayode is a recipient of the Gerald Mott award for meritorious graduate students. He has three publications in three different journals, the Australian Journal of Agricultural Research, Euphytica and Breeding Science respectively, four conference proceedings and has presented posters at both local and international conferences. He was awarded a Rockefeller Foundation Fellowship to obtain his doctoral degree in molecular plant breeding in the Department of Agronomy, Louisiana State University, in August 1999 which he completed in the Fall of 2003.