

# QTL, Multivariate and Population Structure Analyses of Agronomic Traits in the Interspecific Cross *Oryza sativa* X *Oryza glaberrima*

G. Aluko<sup>1</sup>, C. Martinez<sup>2</sup>, J. Tohme<sup>2</sup>, C. Castano<sup>2</sup>, C. Bergman<sup>3</sup>, J.H. Oard<sup>1</sup>

<sup>1</sup>Dept. Agronomy, LSU AgCenter, Louisiana State University, Baton Rouge LA 70803

<sup>2</sup>Centro Internacional de Agricultura Tropical, Apdo Aereo 6713, Cali, Colombia

<sup>3</sup>USDA-ARS Rice Research Unit, 1509 Aggie Drive, Beaumont, TX 77713, USA



## Abstract

Wild relatives of cultivated varieties offer new genetic sources for enhancing economic value, but traditional interval mapping techniques have not gained widespread support among applied U.S. rice researchers for introgression of traits with complex inheritance. The objective of this research was to compare interval, multiple regression, bulked segregant, and Discriminant Analysis procedures for genetic mapping of economically important traits among 312 doubled-haploid lines derived from the interspecific cross *O. sativa* x *O. glaberrima*. Genetic material was planted in 2001 in replicated field plots in Colombia, and the same lines were evaluated in Louisiana in 2002. New markers derived from *O. glaberrima* were detected for percent rice bran, panicle length, and grain yield. Seven QTL for panicle length, tillers/plant, heading date, and 1000 grain weight were detected in both locations. High levels of percent correct classification were obtained for markers identified by the Discriminant Analysis (DA) procedure vs. the other methods. Adjustment for population structure in this controlled cross enhanced classification and improved mapping results. When compared to previous traditional QTL mapping experiments for agronomic traits, markers detected by the DA procedure pointed to the same and different regions on the rice genetic map. Results from this study indicated that use of non-parametric, multivariate methods such as DA and adjustment for population structure can improve mapping of economically important traits vs. traditional genetic approaches.

## Introduction

The African rice *O. glaberrima* exhibits several desirable characteristics that could be introgressed into U.S. germplasm via marker-assisted selection. Traditional genetic mapping techniques (interval mapping, multiple regression, bulked-segregant analysis) have identified putative QTLs for complex agronomic traits in rice that could benefit marker-assisted selection, candidate gene studies, and map-based cloning. However, the majority of these studies have rarely been repeated for verification by other research groups. Population structure, originally described in human studies, is a condition that arises from unequal allelic frequencies among subgroups of a population that leads to spurious associations between genetic marker and phenotype. We show in this study that population structure can also exist within controlled crosses in plants that results in reduced mapping efficiency. Statistical methods were employed in this study to account for population structure detected in the mapping data and to improve quality of the results. Non-parametric approaches such as Discriminant Analysis (DA) offer a heuristic, robust alternative to standard mapping techniques in controlled crosses and even unrelated inbred lines. In addition to markers selected for various agronomic traits, results for our study indicate that adjustment for population structure and use of DA can enhance overall mapping efforts. Non-parametric procedures should therefore be considered as an effective complement to standard interval, multiple regression, and bulked-segregant mapping methods.

## Materials and Methods

**Plant Material:** IRGC 103544 (*O. glaberrima*), native African accession; Caiapo (*O. sativa*), upland indica variety from Brazil, recurrent parent.

**Population Development:** 312 doubled-haploid (DH) lines derived from 97 BC<sub>1</sub>F<sub>2</sub> plants

**Field Experiments:** 2001, Cali, Colombia, 3 reps / DH line  
2002, Crowley, LA, 2 reps / DH line

**Traits:** date of flowering, plant height, days to heading, tiller number, panicle sterility, grain yield, 1000-grain weight, % brown rice, % head rice, % rice bran, % milled rice, amylose content, alkal spreading score, percent protein, grain length, grain width, grain length/width ratio. Inverse or log transformations were used to achieve normal distribution of each trait.

**Statistical Markers:** 100 polymorphic SSR markers – every 10.5 cM on genetic map produced in this study (see Figs 1, 2).

### Statistical Analysis, Map Construction, Marker Detection

**Framework map:** MAPMAKER 3.1, minimum LOD score = 3.0, Kosambi mapping function; "MapDisto" software program (Lorieux 2000) used to assign marker order.

**Marker detection, genetic and molecular analyses:** Interval mapping (IM) (Liu 1997); Composite interval mapping (CIM) (QTL cartographer, Zeng 1994); Multiple regression (MR)(SAS Institute, 2000); Bulked segregant analysis (BSA)(Wang and Paterson 1994; Mitchelmore et al. 1991). Discriminant analysis (SAS Institute, ver. 9.0, 2003); Population structure (Pritchard et al. 2000); Percent correct classification/cross validation (SAS Institute, ver. 9.0, 2003); Analysis of Molecular Variance (AMOVA; Excoffier et al. 1992; Arlequin software, Schneider et al. 2000)

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

Trait	Chrom	MR		IM/CIM		BSA		DA		
		% correct	Chrom	% correct	Chrom	% correct	Chrom	% correct	Chrom	
PHIT	1	RM2136	8	RM2136	8	RM2136	8	RM2136	8	
	6	RM2149	8	RM2149	8	RM2149	8	RM2149	8	
	8	RM2164	10	RM2164	10	RM2164	10	RM2164	10	
	11	RM2182	11	RM2182	11	RM2182	11	RM2182	11	
	8	RM2164	10	RM2164	10	RM2164	10	RM2164	10	
	2	RM2188	11	RM2188	11	RM2188	11	RM2188	11	
	1	RM2197	1	RM2197	1	RM2197	1	RM2197	1	
	2	RM2111	85	RM2197	87	3	RM2185	100	RM2188	100
	11	RM2189	4	RM2149	1	RM2197	4	RM2124	100	
	4	RM2141	4	RM2124	4	RM2124	4	RM2124	100	
HID	5	RM2185	4	RM2185	4	RM2185	4	RM2185	4	
	3	RM2148	6	RM2125	6	RM2125	6	RM2125	6	
	2	RM2148	6	RM2125	6	RM2125	6	RM2125	6	
	8	RM2148	6	RM2125	6	RM2125	6	RM2125	6	
	8	RM2148	6	RM2125	6	RM2125	6	RM2125	6	
	2	RM2148	6	RM2125	6	RM2125	6	RM2125	6	
	8	RM2148	6	RM2125	6	RM2125	6	RM2125	6	
	8	RM2148	6	RM2125	6	RM2125	6	RM2125	6	
	8	RM2148	6	RM2125	6	RM2125	6	RM2125	6	
	8	RM2148	6	RM2125	6	RM2125	6	RM2125	6	
PAN	1	RM2183	2	RM2140	1	RM2183	8	RM2137	8	
	3	RM2166	2	RM2166	2	RM2166	2	RM2166	2	
	1	RM2126	3	RM2166	3	RM2166	3	RM2166	3	
	1	RM2145	3	RM2188	3	RM2188	3	RM2188	3	
	2	RM2163	3	RM2188	3	RM2188	3	RM2188	3	
	3	RM2151	5	RM2163	6	RM2190	6	RM2190	6	
	4	RM2149	8	RM2163	6	RM2190	6	RM2190	6	
	7	RM211	7	RM2190	7	RM2190	7	RM2190	7	
	2	RM2124	8	RM2167	11	RM2167	11	RM2167	11	
	11	RM2129	85	8	RM2137	84	2	RM2171	87	

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



Table 2. Chromosomal location of SSR markers and % correct classification of DH lines using MR, IM, BSA and DA procedures, Crowley, 2002, subpopulation 1 (cont'd)

Trait	Chrom	MR		IM/CIM		BSA		DA	
		% correct	Chrom						
TILL	1	RM2144	11	RM2144	11	RM2144	11	RM2144	11
	1	RM2144	11	RM2144	11	RM2144	11	RM2144	11
	3	RM2148	4	RM2124	4	RM2124	4	RM2124	4
	3	RM2148	4	RM2124	4	RM2124	4	RM2124	4
	11	RM209	11	RM214	11	RM214	11	RM214	11
	8	RM2148	4	RM2124	4	RM2124	4	RM2124	4
	11	RM209	11	RM214	11	RM214	11	RM214	11
	8	RM2148	4	RM2124	4	RM2124	4	RM2124	4
	11	RM209	11	RM214	11	RM214	11	RM214	11
	8	RM2148	4	RM2124	4	RM2124	4	RM2124	4
VLD	11	RM214	11	RM214	11	RM214	11	RM214	11
	11	RM214	11	RM214	11	RM214	11	RM214	11
	11	RM214	11	RM214	11	RM214	11	RM214	11
	11	RM214	11	RM214	11	RM214	11	RM214	11
	11	RM214	11	RM214	11	RM214	11	RM214	11
	11	RM214	11	RM214	11	RM214	11	RM214	11
	11	RM214	11	RM214	11	RM214	11	RM214	11
	11	RM214	11	RM214	11	RM214	11	RM214	11
	11	RM214	11	RM214	11	RM214	11	RM214	11
	11	RM214	11	RM214	11	RM214	11	RM214	11
TCW	8	RM214	2	RM218	2	RM218	2	RM218	2
	8	RM214	2	RM218	2	RM218	2	RM218	2
	8	RM214	2	RM218	2	RM218	2	RM218	2
	8	RM214	2	RM218	2	RM218	2	RM218	2
	8	RM214	2	RM218	2	RM218	2	RM218	2
	8	RM214	2	RM218	2	RM218	2	RM218	2
	8	RM214	2	RM218	2	RM218	2	RM218	2
	8	RM214	2	RM218	2	RM218	2	RM218	2
	8	RM214	2	RM218	2	RM218	2	RM218	2
	8	RM214	2	RM218	2	RM218	2	RM218	2

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

Table 3. Chromosomal location of SSR markers and % correct classification of DH lines using MR, IM, BSA and DA procedures, Columbia 2001, subpopulation 1

Trait	Chrom	MR		IM/CIM		BSA		DA	
		% correct	Chrom	% correct	Chrom	% correct	Chrom	% correct	Chrom
PHIT	1	RM2189	8	RM2189	8	RM2189	8	RM2189	8
	1	RM2189	8	RM2189	8	RM2189	8	RM2189	8
	1	RM2189	8	RM2189	8	RM2189	8	RM2189	8
	1	RM2189	8	RM2189	8	RM2189	8	RM2189	8
	1	RM2189	8	RM2189	8	RM2189	8	RM2189	8
	1	RM2189	8	RM2189	8	RM2189	8	RM2189	8
	1	RM2189	8	RM2189	8	RM2189	8	RM2189	8
	1	RM2189	8	RM2189	8	RM2189	8	RM2189	8
	1	RM2189	8	RM2189	8	RM2189	8	RM2189	8
	1	RM2189	8	RM2189	8	RM2189	8	RM2189	8
HID	5	RM2185	4	RM2185	4	RM2185	4	RM2185	4
	11	RM2182	11	RM2182	11	RM2182	11	RM2182	11
	11	RM2182	11	RM2182	11	RM2182	11	RM2182	11
	11	RM2182	11	RM2182	11	RM2182	11	RM2182	11
	11	RM2182	11	RM2182	11	RM2182	11	RM2182	11
	11	RM2182	11	RM2182	11	RM2182	11	RM2182	11
	11	RM2182	11	RM2182	11	RM2182	11	RM2182	11
	11	RM2182	11	RM2182	11	RM2182	11	RM2182	11
	11	RM2182	11	RM2182	11	RM2182	11	RM2182	11
	11	RM2182	11	RM2182	11	RM2182	11	RM2182	11
PAN	1	RM2183	2	RM2140	1	RM2183	8	RM2137	8
	3	RM2166	2	RM2166	2	RM2166	2	RM2166	2
	1	RM2126	3	RM2166	3	RM2166	3	RM2166	3
	1	RM2145	3	RM2188	3	RM2188	3	RM2188	3
	2	RM2163	3	RM2188	3	RM2188	3	RM2188	3
	3	RM2151	5	RM2163	6	RM2190	6	RM2190	6
	4	RM2149	8	RM2163	6	RM2190	6	RM2190	6
	7	RM211	7	RM2190	7	RM2190	7	RM2190	7
	2	RM2124	8	RM2167	11	RM2167	11	RM2167	11
	11	RM2129	85	8	RM2137	84	2	RM2171	87

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

Table 4. Chromosomal location of SSR markers and % correct classification of DH lines using MR, IM, BSA and DA procedures, Columbia 2001, subpopulation 1 (cont'd)

Trait	Chrom	MR		IM/CIM		BSA		DA	
		% correct	Chrom						
TILL	1	RM2144	11	RM2144	11	RM2144	11	RM2144	11
	1	RM2144	11	RM2144	11	RM2144	11	RM2144	11
	3	RM2148	4	RM2124	4	RM2124	4	RM2124	4
	3	RM2148	4	RM2124	4	RM2124	4	RM2124	4
	11	RM209	11	RM214	11	RM214	11	RM214	11
	8	RM2148	4	RM2124	4	RM2124	4	RM2124	4
	11	RM209	11	RM214	11	RM214	11	RM214	11
	8	RM2148	4	RM2124	4	RM2124	4	RM2124	4
	11	RM209	11	RM214	11	RM214	11	RM214	11
	8	RM2148	4	RM2124	4	RM2124	4	RM2124	4
VLD	11	RM214	11	RM214	11	RM214	11	RM214	11
	11	RM214	11	RM214	11	RM214	11	RM214	11
	11	RM214							