Euphytica (2012) 187:289–301
DOI 10.1007/s10681-012-0770-5

Detecting epistatic effects associated with cotton traits by a modified MDR approach

Jixiang Wu · Johnie N. Jenkins · Jack C. McCarty · Karl Glover

Received: 4 April 2012 / Accepted: 13 July 2012 / Published online: 1 August 2012

Abstract Genetic expression of a trait is complicated and it is usually associated with many genes including their interactions (epistasis) and genotype-by-environment interactions. Genetic mapping currently focuses primarily on additive models or marginal genetic effects due to the complexity of epistatic effects. Thus, there exists a need to appropriately identify favorable epistatic effects for important biological traits. Several multifactor dimensionality reduction (MDR) based methods are important resources to identify high-order gene–gene interactions. These methods are mainly focused on human genetic studies. Many traits in plant systems are not only quantitatively inherited but also are often measured in repeated field plots under multiple environments. In this study, we proposed a mixed model based MDR approach, which is suitable for inclusion of various fixed and random effects. This approach was used to analyze a cotton data set that included eight agronomic and fiber traits and 20 DNA markers. The results revealed high order epistatic effects were detected for most of these traits using this modified MDR approach.

Keywords Epistasis · Multifactor dimensionality reduction · Mixed linear model · Cotton · DNA marker

J. Wu (✉) · K. Glover
Department of Plant Science, South Dakota State University, Brookings, SD 57007, USA
e-mail: jixiang.wu@sdstate.edu

J. N. Jenkins · J. C. McCarty
Crop Science Research Laboratory, USDA-ARS, Mississippi State, MS 39762, USA

Introduction

Epistatic effects, equivalent to gene-by-gene interaction effects, are common genetic components controlling quantitative traits in plant systems. The effects can be detected by appropriate genetic modeling and statistical approaches based on phenotypic data (Goldringer et al. 1997; Xu and Zhu 1999; Wu et al. 2006). For example, studies revealed that epistatic effects were associated with cotton agronomic and fiber traits (Xu and Zhu 1999; McCarty et al. 2004a, b; Saha et al. 2010) and wheat agronomic traits (Goldringer et al. 1997). Epistatic effects detected by the similar studies as mentioned above normally were cumulative from the whole genome.

A major goal of genetic mapping studies is to identify useful DNA markers that are associated with traits of importance and in turn to increase the probability of selecting superior genotypes (Stromberg et al. 1994; Knapp 1998; Liu et al. 2003). Once DNA markers that are highly associated with target traits are determined, selection process for one or more traits can

be conducted based on the identified DNA markers. Since DNA markers normally are neutral and genotyping is independent of environmental conditions or growing seasons, marker assisted selection (MAS) for a single trait or several traits can be conducted under laboratory conditions during off-seasons (Lande and Thompson 1990; Dudley 1993). The molecular selection process can be more efficient than the traditional breeding process. Clearly, the success of MAS for a trait is highly dependent on DNA markers being identified. Regression based approaches including expected maximum (EM) based or non-EM based are commonly used for quantitative trait loci (QTL) mapping or association mapping (Lander and Botstein 1989; Haley and Knott 1992; Jansen 1992, 1993; Zeng 1994; Broman 2001; Foulkes 2009). Mixed linear model approaches have also been employed in genetic mapping (Wang et al. 1999; Yu et al. 2006; Stich et al. 2008). However, due to the complexity of epistatic effects, many QTL and association mapping studies are focused on marginal effects of each marker, which are equivalent to additive models (Lander and Botstein 1989; Martin and Curnow 1992; Haley and Knott 1992; Jansen 1992, 1993; Zeng 1994). Using mixed linear model approaches, Wang et al. 1999, proposed a two-way search algorithm to determine QTL and QTL-by-QTL interaction effects for various experimental mapping populations. Epistatic effects among QTLs have been reported in cotton (Shen et al. 2006), wheat (Zhang et al. 2008), and rice (Cao et al. 2001; Xing et al. 2002; Yang et al. 2007). These studies were mainly focused on determination of interaction effects between two significant individual QTLs, suggesting that marginal effects need to be identified before epistatic effects are determined. In other words, if marginal effects for two QTLs are not significant, their interaction effects between these two QTLs would not be detected. It is well known that this assumption is not always appropriate.

It is very likely that more than two genes/QTLs control a trait in an interactive way. Recent emerging combinatorial approaches such as multifactor dimensionality reduction (MDR) method (Ritchie et al. 2001; Hahn et al. 2003; Moore et al. 2006), the combinatorial portioning method (CPM) (Nelson et al. 2001), restricted partition method (RPM) (Culverhouse et al. 2004), and the generalized multifactor dimensionality reduction (GMDR) method (Lou et al., 2007) are novel tools to reduce the dimensionality and yet to determine gene × gene or marker × marker interaction effects. Most of these approaches were proposed to determine epistatic effects associated with categorical traits such as human diseases. The CMP and RMP methods are able to detect epistatic effects associated with quantitative traits (Nelson et al. 2001; Culverhouse et al. 2004). The goals of these two methods are to determine distinct groups but covariates could not be added to the models to be investigated. The GMDR is a generalized MDR method, which allows covariates to be included into a model and can be used to analyze both categorical and quantitative traits (Lou et al. 2007). Unlike human disease, many traits in plants (e.g. yield and height) are often repeatedly measured in multiple environments. For example, environmental, genotype-by-environment interaction, and block effects should be included and they normally are random if a randomized complete block design is used. Therefore, a MDR approach with integration of the above mentioned random effects will be an important addition to the current statistical pool for detection of epistatic effects in plant association mapping studies.

Cotton, an important dual crop, producing both fibers for textile industries and seeds for oil and protein production, has been extensively explored in the past two decades at the molecular level (i.e. Meredith 1992; Cantrell and Davis 1993; Reinisch et al. 1994; Jiang et al. 1998, 2000; Kohel et al. 2001; Lacape et al. 2005; Gutierrez et al. 2002; Guo et al. 2003; Zhang et al. 2003; Mei et al. 2004; He et al. 2005; Lin et al. 2005; Ulloa et al. 2005; Shen et al. 2006; Wu et al. 2007, 2008). Like many studies in other plant systems, genetic mapping in cotton has generally focused on single markers or additive genetics models. Some studies showed the evidence of epistatic effects contributing to both agronomic and fiber traits (Xu and Zhu 1999; McCarty et al. 2004a, b, 2008; Shen et al. 2006; Wu et al. 2006; Saha et al. 2010; Lü et al. 2011).

Two major objectives are included in the present study. First, we proposed a mixed linear model based MDR (MMDR) approach that can be used to determine high-order epistatic effects among markers associated with quantitative traits. Second, we used this new approach to determining epistatic effects among 20 AFLP markers previously reported associated with cotton agronomic and fiber traits (Wu et al. 2007). The purpose of this study is to provide a new statistical framework to identify DNA markers that

can be used to improve MAS efficiency for crop improvement.

Materials and methods

Materials and experiment

The experimental materials, field experiments, and DNA marker genotyping were detailed in previous reports (Swindle 1993; Zhong 2001; Zhong et al. 2002; Wu et al. 2007). Four photoperiod sensitive accessions of cotton (*Gossypium hirsutum* L.) were crossed with the day-neutral flowering donor parent 'Deltapine 16' (DP16). The accessions were T78 (PI 549140) race latifolium, T174 (PI 163647) race latifolium, T326 (PI 165326), and T1149 (PI 529966). Twenty day-neutral (DN) populations (F_6, BC_1F_6, BC_2F_6, BC_3F_6, and BC_4F_6, for each of the four accessions) were derived from these four race accessions by backcrossing and selfing (McCarty et al. 1979; Swindle 1993). Four elite cultivars, 'DES 119', 'Deltapine 50' (DP50), 'Stoneville 453' (ST453), and 'Coker 315' (C315) were also included. Four elite cultivars and 20 DN populations were planted in two locations at Mississippi State, MS in 1989 (Swindle 1993). A randomized complete block (RCB) design with four replications was applied for each location. Planting dates were April 26 and 27 at the two locations, respectively. Plots were single rows, 12.3 m long with spacing of approximate 1 m between rows. Standard cultural and insect control practices were followed on these plots. Prior to machine harvest, a 50-boll sample was hand-picked from each plot for boll weight (BW) and lint percentage (LP) determinations. The lint samples were sent to a commercial laboratory (StarLab, Knoxville, TN) for determining 2.5 % fiber span length (SL), micronaire (MIC), fiber elongation (EL), and fiber strength (T1). Seed yield (YLD) and lint yield (LY) were converted from plot weight and expressed as kg ha^{-1}. Boll number ha^{-1} for each plot was converted using the formula used by Tang et al. (1996). Seeds from the 20 DN lines with their four original photoperiodic parental lines and the four cultivars were planted in the field in 1999 for DNA sample collection and AFLP analysis (Zhong 2001; Zhong et al. 2002). In a previous study, 119 AFLP markers were identified as polymorphic among 24 cotton lines. Twenty markers were selected to be associated with seven traits by stepwise regression analysis (Wu et al. 2007; Table 5). These 20 AFLP markers (with their codes: 1–20) for these 24 cotton genotypes were used for further analysis are listed in Table 7.

Methodology

Mixed model based MDR approach is detailed as following. Given a mapping population containing n inbred (homozygous) lines and m bi-allelic polymorphic markers, all possible allelic combinations among p markers are 2^p. For example, given $p = 3$, there are eight possible allelic combinations (000, 001,010, 011, 100, 101, 110, and 111). It is likely that total actual allelic combinations are smaller than 2^p for a small n. We treat each allelic combination among the p markers as a new genotype. Thus, p-dimension variables are converted to one-dimension new factor with up to 2^p levels (new genotypes). For $p = 3$, there are eight possible new genotypes among n inbred lines as mentioned above. We assume that genetic effects are different among different allelic combinations. If these n lines are only measured with no replications, a general linear model analysis can be applied, which are equivalent to the CMP and RMP methods (Nelson et al. 2001; Culverhouse et al. 2004). On the other hand, if these n lines are evaluated in several environments with a RCB design, then a mixed linear model including environmental effects, (new) genotypic effects, (new) genotype-by-environment (GE) interaction effects, and block effects within each environment can be employed (Wu et al. 2010). In this study, since there were 24 genotypes, we selected all two- to four- marker combinations among 20 markers for each trait.

For comparisons, various mixed linear models without including interaction effects among the AFLP markers being used for MMDR analysis were also developed. Each model included effects of environment, markers, marker-by-environment interaction, block, and residual. These mixed models can be considered as additive models with no epistasis.

Several mixed linear model approaches can be used to estimate variance components (Hartley and Rao 1967; Patterson and Thompson 1971; Rao 1971; Zhu 1989; Searle et al. 1992) and to predict genetic effects (Zhu 1993; Zhu and Weir 1994). Re-sampling approaches such as jackknife and bootstrapping can

be employed to standard errors for all parameters in the model being used and then an approximate-test or confidence intervals can be used to test significance of each parameter (Miller 1974; Davison and Hinkley 1998; Wu et al. 2008). Permutation test may also be used to obtain empirical probability value for each parameter (Manly 2006). This process needs to be repeated C_m^p times for choosing p markers among m markers to find the best marker combinations for the same dimensionality. Thus, this process can be very time consuming, however, if m is large.

Mixed linear model analysis with re-sampling process for all C_m^p marker combinations can increase computational burden dramatically. In order to reduce computational burden, the following procedure was employed alternatively.

Step 1: Obtain R^2 (an adjusted coefficient of determination) or V_e (variance for residual) for each of C_m^p marker combinations for each trait. Pick up the marker combination with the highest R^2 or the smallest V_e;
Step 2: Apply mixed linear model approaches to analyze the best combination only for each trait if necessary. Variance components can be estimated and genotypic effects for different allelic combinations can be predicted by different prediction approaches (Zhu 1993; Zhu and Weir 1994);
Step 3: Apply re-sampling techniques such as jackknife and bootstrapping to obtain standard error for each parameter (variance component or genotypic effect) and appropriate t tests can be conducted as mentioned above.

In this study, we used a MINQUE approach with all prior values equal to 1 (Zhu 1989) for variance component estimation and adjusted unbiased prediction (AUP) approach (Zhu 1993) to predict effects of all allelic combinations. A tenfold jackknife technique was applied to calculate standard deviation for each parameter as follows.

Step 1: Randomly divide the whole data into 10 groups;
Step 2: Delete one out of 10 groups each time and calculate each parameter each time; repeat 10 times;
Step 3: Calculate mean value for each parameter and its standard error and an approximate t test was used to test the significance of each parameter (Wu et al. 2008).

All data analyses were conducted by *R* programs developed by the authors of this study.

Results

Phenotypic data analysis

Phenotypic mean values for these 24 cotton lines were reported previously (Wu et al. 2007); however, variance components and genotypic effects for these eight agronomic and fiber traits were not reported. These results are listed in Tables 1 and 2, respectively.

Variance components

Estimated variance components expressed as proportions to the phenotypic variances for eight traits are summarized in Table 1. Genotypic effects were the primary genetic components contributed to all these traits, ranged from 44.6 % (fiber strength, T1) to 80.1 % (lint percentage, LP). Environmental effects were significant only for lint yield. Only lint percentage was significantly affected by genotype-by-environment interaction effects. However, the contribution to this trait was numerically small (3.9 %). Block effects were significant for lint yield, boll number, and lint percentage but were numerically small.

Genotypic effects

Genotypic effects of these 24 cotton lines were predicted by the AUP approach (Zhu 1993) and are listed in Table 2. In addition to the genotypic effects, the low- and upper-limits of 95 % confidence interval for each genotypic effect are presented for statistical test and multiple comparisons among different genotypes (Table 2). Four elite genotypes had higher lint yield than all DN lines derived from four race accessions. Two DN lines (7 and 16) had positive genotypic effects while nine of 20 DN lines had negative genotypic effects associated with lint yield. Four elite cultivars had positive genotypic effects and were greater than all DN lines for boll number. Among 20 DN lines, four had had positive genotypic effects and eight had negative genotypic effects for this trait. Four elite cultivars and five DN lines had positive genotypic effects for boll weight and nine DN lines had negative genotypic effects for this trait. Four elite

Table 1 Estimated variance components expressed as proportions to the total variances for eight traits in cotton

	LY		BN		BW		LP	
	Estimate	*P* value	Estimate	*P* value	Estimate	*P* value	Estimate	*P* value
$V_E/V_P^{\ddagger}$	0.021	0.045	0.038	0.055	0.000	1.000	0.002	0.836
V_G/V_P	0.650	0.000	0.421	0.000	0.567	0.000	0.801	0.000
V_{GE}/V_P	0.005	0.879	0.001	0.975	0.005	0.923	0.039	0.037
V_B/V_P	0.011	0.044	0.038	0.004	0.017	0.061	0.011	0.010
V_e/V_P	0.313	0.000	0.502	0.000	0.410	0.000	0.147	0.000
	MIC		SL		EL		T1	
	Estimate	*P* value	Estimate	*P* value	Estimate	*P* value	Estimate	*P* value
V_E/V_P	0.000	1.000	0.000	0.996	0.000	0.967	0.014	0.172
V_G/V_P	0.521	0.000	0.701	0.000	0.462	0.000	0.446	0.000
V_{GE}/V_P	0.028	0.354	0.034	0.080	0.013	0.852	0.036	0.342
V_B/V_P	0.073	0.000	0.064	0.000	0.001	0.953	0.003	0.865
V_e/V_P	0.377	0.000	0.202	0.000	0.524	0.000	0.501	0.000

LY lint yield, *BN* boll number/ha, *BW* boll weight, *LP* lint percentage, *MIC* fiber micronaire, *SL* 2.5 % fiber span length, *EL* elongation, and *T1* fiber strength

‡ V_E, V_G, V_{GE}, V_B, and V_e are the variance components for environmental effects, genotypic effects, genotype-by-environment interaction effects, block effects, and random errors, respectively. $V_P = (V_E + V_G + V_{GE} + V_B + V_e)$ is the phenotypic variance

cultivars and four DN lines had positive genotypic effects for lint percentage and 11 DN lines had negative genotypic effects. Three elite cultivars were significantly higher than the remaining cotton lines regarding genotypic effects for lint percentage. Unlike agronomic traits, genotypic effects showed no consistent pattern among elite cultivars and DN lines. One cultivar and nine DN lines had positive genotypic effects while two cultivars and six DN lines had negative genotypic effects for micronaire. One cultivar and nine DN lines had positive genotypic effects while two cultivars and six DN lines had negative genotypic effects for micronaire. All four cultivars and four DN lines had positive genotypic effects while 14 DN lines had negative genotypic effects for fiber span length. Three cultivars and six DN lines had positive genotypic effects while one cultivar and seven DN lines had negative genotypic effects for elongation. Two cultivars and six DN lines had positive genotypic effects while two cultivars and six DN lines had negative genotypic effects for fiber strength.

Model comparisons

Two approaches were used for the data analyses between eight cotton traits and 20 AFLP markers. For clarification, we briefly discuss these analyses. The first analysis was MDR based for two to four markers with GE interaction as detailed in "Materials and methods" section. The second approach was based on conventional model analysis, which is equivalent to additive effect models yet this model included marker-by-environmental interaction effects. Since the total number of genotypes was 24, the maximum number of p was 4. The adjusted coefficients of determination for both models for each of 2-, 3-, and 4- marker combinations were calculated. There were totals of 190, 1140, and 4845 marker combinations for two-, three-, and four-marker cases, respectively. The difference in adjusted coefficient of determination between two models for the same p markers indicates the interaction effect among these p markers. Table 3 lists the best marker combinations with the highest R^2 by the MDR method, and R^2 by the regular (additive) model for each of eight traits for the same two-, three-, and four-marker cases.

For the two-marker case, except lint percentage, the R^2 for two models were almost identical for the other seven traits, suggesting that the best two-marker combinations for these seven traits did not show interactions except markers 14 and 17 showing an interaction contributing to lint percentage (Table 3).

Table 2 Predicted genotypic effects of 24 cotton genotypes for eight traits

Line[††]	LY (kg/ha)			BN (10^3)			BW (g)			LP (%)		
	Effect	LL[‡]	UL	Effect	LL	UL	Effect	LL	UL	Effect	LL	UL
1	−249.26	−298.47	−200.06	−74.56	−104.00	−45.12	−0.25	−0.40	−0.10	−4.32	−4.89	−3.74
2	−268.15	−352.96	−183.33	−113.56	−161.97	−65.15	−0.11	−0.33	0.12	−3.37	−4.10	−2.65
3	−29.21	−89.65	31.24	35.36	−2.88	73.60	−0.21	−0.29	−0.13	−2.16	−2.44	−1.87
4	−365.79	−396.34	−335.24	−104.23	−124.21	−84.24	−0.91	−1.05	−0.78	−5.06	−5.53	−4.59
5	−249.70	−290.93	−208.46	−133.02	−158.48	−107.56	0.27	0.13	0.40	−1.14	−1.69	−0.59
6	−34.59	−109.42	40.24	−29.48	−70.31	11.35	0.39	0.24	0.54	−0.79	−1.99	0.42
7	83.74	38.95	128.53	50.96	24.98	76.95	0.07	−0.03	0.16	−0.31	−0.95	0.33
8	3.35	−34.47	41.17	12.63	−10.63	35.88	0.08	−0.03	0.20	−0.93	−1.50	−0.35
9	−12.02	−89.00	64.97	25.35	−15.60	66.31	−0.22	−0.31	−0.13	−0.33	−0.74	0.09
10	−127.21	−179.70	−74.72	−68.17	−90.89	−45.45	0.33	0.15	0.50	−1.61	−2.48	−0.73
11	−24.08	−86.16	37.99	73.98	22.57	125.39	−0.99	−1.24	−0.75	2.44	1.88	3.00
12	95.80	−14.69	206.28	29.91	−19.29	79.10	0.06	−0.10	0.23	1.57	0.35	2.79
13	−3.10	−64.89	58.70	38.24	4.70	71.78	−0.38	−0.55	−0.21	0.22	−0.20	0.65
14	−181.17	−239.28	−123.06	−57.30	−89.64	−24.96	−0.10	−0.21	0.00	−2.82	−3.35	−2.29
15	−165.03	−213.22	−116.85	−67.74	−90.75	−44.73	−0.21	−0.43	0.01	−0.56	−2.02	0.89
16	93.24	44.04	142.45	74.21	34.45	113.97	−0.28	−0.49	−0.08	1.25	0.63	1.86
17	−215.31	−258.52	−172.10	−81.90	−106.62	−57.18	−0.16	−0.30	−0.03	−1.96	−2.61	−1.32
18	−2.22	−35.00	30.56	5.22	−14.33	24.77	0.13	0.02	0.24	−0.85	−1.64	−0.06
19	−11.93	−46.27	22.41	−10.07	−29.20	9.06	0.06	−0.06	0.18	0.53	0.11	0.95
20	−5.87	−58.73	46.98	11.78	−15.73	39.29	0.11	0.06	0.16	−1.46	−1.97	−0.96
21	418.27	384.65	451.89	126.81	109.04	144.57	0.21	0.13	0.30	5.79	5.50	6.07
22	354.60	316.49	392.72	102.10	85.95	118.25	0.58	0.52	0.65	2.47	2.18	2.76
23	539.62	504.15	575.10	95.10	78.12	112.08	0.84	0.77	0.91	7.86	7.52	8.20
24	356.01	302.39	409.63	58.38	34.19	82.56	0.70	0.57	0.83	5.54	5.31	5.77

Line	MIC			SL 25 (mm)			EL (%)			T1 (kNm/kg)		
	Effect	LL	UL	Effect	LL	UL	Effect	LL	UL	Effect	LL	UL
1	−0.63	−0.72	−0.53	−0.20	−0.48	0.08	0.19	−0.06	0.44	−0.56	−5.95	4.84
2	−0.08	−0.23	0.06	−1.54	−2.01	−1.07	0.25	0.10	0.41	−7.98	−13.22	−2.74
3	0.13	0.06	0.21	−1.18	−1.50	−0.86	−0.18	−0.65	0.29	0.62	−5.06	6.29
4	−0.55	−0.65	−0.45	−1.57	−1.73	−1.42	0.51	0.04	0.98	−10.09	−12.65	−7.53
5	0.53	0.42	0.63	−1.33	−1.57	−1.09	0.29	0.15	0.42	−6.70	−10.02	−3.38
6	0.36	0.30	0.43	−0.35	−0.94	0.24	0.34	0.04	0.64	11.03	5.60	16.46
7	−0.23	−0.32	−0.14	0.07	−0.23	0.37	−0.29	−0.49	−0.09	−1.55	−6.11	3.01
8	−0.02	−0.11	0.07	−0.14	−0.45	0.16	−0.15	−0.34	0.04	−3.48	−8.19	1.23
9	0.24	0.12	0.35	0.65	0.50	0.80	−1.30	−1.61	−0.99	13.66	10.55	16.77
10	0.46	0.37	0.55	−0.94	−1.43	−0.45	−0.87	−1.17	−0.57	5.96	3.37	8.56
11	0.17	0.08	0.25	−0.54	−0.76	−0.32	0.25	−0.06	0.57	0.53	−2.87	3.93
12	−0.08	−0.16	0.00	0.28	0.03	0.54	−0.24	−0.45	−0.02	2.16	0.24	4.08
13	−0.13	−0.22	−0.03	0.84	0.63	1.04	−0.75	−0.98	−0.52	8.40	5.11	11.69
14	0.17	0.07	0.27	−0.35	−0.66	−0.05	−0.58	−0.87	−0.29	3.82	0.11	7.52
15	0.32	0.27	0.37	0.95	0.56	1.34	−0.32	−0.68	0.04	6.70	4.23	9.17
16	−0.10	−0.24	0.03	−1.00	−1.41	−0.59	0.17	−0.19	0.53	−3.75	−6.85	−0.65

Table 2 continued

Line	MIC			SL 25 (mm)			EL (%)			T1 (kNm/kg)		
	Effect	LL	UL	Effect	LL	UL	Effect	LL	UL	Effect	LL	UL
17	0.00	−0.07	0.07	−0.47	−0.83	−0.11	0.72	0.59	0.84	−1.12	−6.25	4.02
18	0.23	0.09	0.38	−0.84	−1.08	−0.60	−0.26	−0.56	0.03	1.89	−0.92	4.70
19	−0.14	−0.24	−0.03	0.10	−0.26	0.46	−0.01	−0.33	0.31	−8.80	−12.05	−5.54
20	−0.18	−0.30	−0.06	−0.84	−1.12	−0.56	1.18	0.90	1.45	−18.60	−21.91	−15.28
21	0.07	0.02	0.12	1.75	1.64	1.86	0.67	0.55	0.79	3.83	2.29	5.36
22	−0.15	−0.20	−0.10	1.95	1.82	2.09	0.85	0.73	0.97	−8.01	−9.67	−6.35
23	0.03	−0.02	0.08	1.43	1.31	1.55	0.35	0.20	0.49	−7.05	−9.13	−4.96
24	−0.42	−0.49	−0.35	3.27	3.09	3.45	−0.82	−0.95	−0.68	19.08	16.43	21.72

LY lint yield, *BN* boll number/ha, *BW* boll weight, *LP* lint percentage, *MIC* fiber micronaire, *SL* 2.5 % fiber span length, *EL* elongation, and *T1* fiber strength

‡ LL and UL are lower limit and upper limit of 95 % confidence interval, respectively

†† Lines 1–20 are the codes for 20 day-neutral cotton lines and lines 21:24 are the codes for four elite cotton cultivars

For the three-marker case, no interaction effects for boll number or fiber span length were detected while interaction effects for lint yield, boll weight, lint yield, micronaire, elongation, and fiber strength were detected. Interaction among the best four-marker combinations contributed to all traits. We also observed that the best four-marker combinations did not include all markers from the best three-marker or two marker combinations for most traits. The results suggested that (1) existence of high order interactions doesn't mean existence of low order interactions and (2) determination of best marker combinations for high order interaction effects cannot depend on the best marker combinations for low order interaction effects. Thus, this MMDR approach along with other combinatorial methods will need extensive computations when high order interaction effects need to be detected. Based on our MDR analysis, we decided the best four-marker combinations for these eight traits should be further analyzed by a MINQUE approach (Zhu 1989).

Mixed model based MDR analysis

It is interesting to apply the MMDR method for these best four marker combinations. The same procedure for phenotypic data was used for the MDR model. Both variance components were estimated and

Table 3 Adjusted coefficients of determination for the best 2- to 4- marker combinations (R^2_{MMDR}) and their corresponding adjusted coefficients of determination without interaction effects among markers (R^2_{LM}) for eight cotton traits

	Two markers				Three markers					Four markers					
	M1	M2	R^2_{MMDR}	R^2_{LM}	M1	M2	M3	R^2_{MMDR}	R^2_{LM}	M1	M2	M3	M4	R^2_{MMDR}	R^2_{LM}
LY	1	4	0.70	0.70	12	14	20	0.73	0.61	4	14	18	19	0.76	0.69
BN	3	4	0.43	0.43	3	4	13	0.45	0.45	11	12	19	20	0.49	0.38
BW	13	18	0.46	0.46	13	14	18	0.51	0.49	2	13	14	18	0.55	0.52
LP	14	17	0.75	0.66	11	14	17	0.78	0.74	3	11	14	17	0.81	0.78
MIC	8	11	0.21	0.21	4	8	11	0.32	0.22	8	11	12	19	0.41	0.29
SL25	1	8	0.67	0.67	1	5	8	0.73	0.73	5	8	15	18	0.76	0.66
EL	8	15	0.33	0.33	8	14	20	0.38	0.31	7	8	14	20	0.42	0.36
T1	8	15	0.37	0.37	3	8	14	0.42	0.40	8	13	14	16	0.46	0.42

LY lint yield, *BN* boll number/ha, *BW* boll weight, *LP* lint percentage, *MIC* fiber micronaire, *SL* 2.5 % fiber span length, *EL* elongation, and *T1* fiber strength

Table 4 Estimated proportional variance components of the best four-marker combination for each agronomic and fiber trait

	LY		BN		BW		LP	
	Estimate	*P* value	Estimate	*P* value	Estimate	*P* value	Estimate	*P* value
$V_E/V_P^{\ddagger}$	0.006	0.297	0.010	0.718	0.000	1.000	0.000	0.996
V_G/V_P	0.742	0.000	0.463	0.000	0.668	0.000	0.823	0.000
V_{GE}/V_P	0.014	0.154	0.012	0.805	0.003	0.960	0.002	0.845
V_B/V_P	0.008	0.066	0.034	0.003	0.010	0.184	0.005	0.133
V_e/V_P	0.230	0.000	0.480	0.000	0.320	0.000	0.170	0.000
	MIC		SL25		EL		T1	
	Estimate	*P* value	Estimate	*P* value	Estimate	*P* value	Estimate	*P* value
V_E/V_P	0.000	1.000	0.000	0.982	0.000	0.966	0.000	0.972
V_G/V_P	0.574	0.000	0.795	0.000	0.572	0.000	0.463	0.000
V_{GE}/V_P	0.047	0.170	0.027	0.155	0.001	0.978	0.010	0.837
V_B/V_P	0.046	0.000	0.035	0.000	0.000	0.981	0.001	0.949
V_e/V_P	0.334	0.000	0.142	0.000	0.426	0.000	0.526	0.000

LY lint yield, *BN* boll number/ha, *BW* boll weight, *LP* lint percentage, *MIC* fiber micronaire, *SL* 2.5 % fiber span length, *EL* elongation, and *T1* fiber strength

‡ V_E, V_G, V_{GE}, V_B, and V_e are the variance components for environmental effects, genotypic effects, genotype-by-environment interaction effects, block effects, and random errors, respectively. $V_P = (V_E + V_G + V_{GE} + V_B + V_e)$ is the phenotypic variance

genotypic effects for all existing allelic combinations for the best four-marker combinations for these traits were predicted (Zhu 1993). The estimated proportional variance components based on these four-marker combinations for these eight traits are summarized in Table 4. Comparing the results in Tables 1 and 4, the variance components were consistent for these traits, suggesting that these best four-marker combinations could be used to predict genotypic values for these cotton lines regarding these traits.

The variance components in Table 4 provided useful genetic information such as which markers can be used for marker assisted selection. More importantly, breeders are also interested in which marker allelic combinations can be used for MAS for a trait. The predicted genotypic effects for different allelic combinations were not listed but genotypic values for 24 cotton lines are summarized in Table 5. Based on the results in our previous report (Table 7), the genotypic effects for different allelic combinations can be obtained easily. The genotypic effects listed in Table 5 can be used to determine which allelic combinations can be used for marker assisted selection for a trait regarding a particular four-marker combination. The results showed that some genotypes had the same genotypic values for each trait because they shared the same allelic combinations for each four-marker combination. The correlation coefficients between the predicted genotypic effects from the phenotypic data and the predicted genotypic effects from the all existing allelic combinations for the best four-marker groups ranged from 0.81 (elongation) to 0.98 (lint yield) (the bottom line of Table 5), suggesting that these marker combinations can be used for marker assisted selection for these traits efficiently.

Comparing the results to our previous report (Wu et al. 2008), we observed that the markers and/or the numbers of markers that were associated with these traits were different. For example, in our previous report, four markers (1, 4, 19, and 20) were detected by stepwise linear regression analysis and contributed 93 % genotypic variation for lint yield, while four markers (4, 14, 18, 19) were selected by mixed MDR approach and contributed 96 % genotypic variation in this study. Four markers (1, 4, 8, and 12) were detected to contribute 85 % genotypic variation for 2.5 % fiber span length in the previous study, while in this study four markers (5, 8, 15, and 18) contributed 90 % genotypic variation for the same trait. Three markers (6, 7, and 17) were detected to contribute 55 % genotypic variation for fiber elongation in the previous study while four markers (7, 8, 14, and 20) contributed

Table 5 Predicted genotypic effects of 24 cotton genotypes based on the best 4- marker combinations and correlations with predicted genotypic effects (Table 2) from phenotypic data only

Line††	LY (kg/ha)	BN (10^3)	BW (g)	LP (%)	MIC	SL 25 (mm)	EL (%)	T1 (kNm/kg)
1	−228.56	−39.84	−0.17	−5.25	−0.54	−0.69	0.49	−7.93
2	−228.56	−78.46	−0.17	−5.25	0.00	−1.60	0.49	−7.93
3	16.51	69.95	−0.17	−5.25	0.21	−1.60	0.49	−7.93
4	−323.05	−68.56	−0.88	−5.25	−0.46	−1.60	0.49	−7.93
5	−209.80	−97.93	0.29	−2.43	0.60	−1.40	0.49	−7.93
6	62.94	66.03	0.02	−2.33	0.08	0.05	−0.13	6.71
7	62.94	66.03	0.02	−2.33	0.08	0.05	−0.13	−3.68
8	62.94	66.03	0.02	−2.33	0.08	0.05	−0.13	−3.68
9	62.94	66.03	0.02	−2.33	0.08	0.05	−1.18	6.71
10	−87.90	−16.78	0.52	−2.33	0.30	−1.02	−0.13	6.71
11	62.94	66.03	−0.68	0.69	0.08	0.05	−0.13	−1.91
12	62.94	66.03	0.02	0.69	0.08	0.05	−0.13	−3.68
13	62.94	66.03	−0.68	−2.33	0.08	0.05	−0.13	6.05
14	−139.56	−16.78	0.02	−2.33	0.30	0.05	−0.13	6.71
15	−139.56	−16.78	0.02	−2.33	0.30	0.05	−0.13	−3.68
16	62.94	77.97	0.02	−0.07	0.13	−0.69	−0.13	−3.68
17	−75.77	−46.82	0.03	−2.43	0.13	−0.69	0.58	−3.55
18	62.94	77.97	0.02	−2.33	0.13	−0.69	−0.13	−3.68
19	16.51	−16.78	0.03	−2.43	0.30	−0.69	0.49	−12.32
20	−75.77	66.03	0.03	−2.43	0.08	−0.69	0.49	−12.32
21	415.36	134.45	0.52	3.13	0.08	1.58	0.92	−3.68
22	415.36	134.45	0.52	3.13	0.08	1.58	0.92	−3.68
23	569.83	134.45	0.86	6.25	0.08	1.58	0.58	−12.32
24	415.36	134.45	0.52	3.13	−0.33	3.07	−0.68	16.63
$r^{\ddagger}$	0.98	0.94	0.91	0.95	0.83	0.95	0.81	0.87

LY lint yield, *BN* boll number/ha, *BW* boll weight, *LP* lint percentage, *MIC* fiber micronaire, *SL* 2.5 % fiber span length, *EL* elongation, and *T1* fiber strength

‡ Correlation coefficient between two predicted genotypic effects (based on phenotypic data only and based on the best-four marker) for each trait

†† Lines 1–20 are the codes for 20 day-neutral cotton lines and lines 21:24 are the codes for four elite cotton cultivars

66 % for this trait. Similar results can be observed when comparing these two studies. Thus, the results in this study clearly show that the mixed model based MDR method had better prediction than linear regression analysis (Wu et al. 2007) and additive models mentioned in this study.

Discussion

The original MDR methods were developed to determine gene–gene or gene–environment interaction effects contributing to human diseases data (Ritchie et al. 2001; Hahn et al. 2003; Nelson et al. 2001; Culverhouse et al. 2004; Lou et al. 2007, 2008). One major limitation is that these methods are not able to analyze a model including random effects such as in many plant association mapping studies. With the MMDR method, we analyzed the data subject to a model including both environmental, genotypic, and their interaction effects. Thus, this method can be considered as an extension of the MDR method proposed by Lou et al. (2007). In addition, unlike previously reported MDR methods which classify the whole data into high risk and low risk groups for categorical variables or several groups for quantitative

Table 6 Genotypic effects and their 95 % confidence intervals of 10 allelic combinations of four AFLP markers for lint yield

Marker combination				Effect (kg/ha)	LL‡	UL
4†	14†	18†	19†			
1	1	1	1	−228.56	−284.64	−172.47
0	1	1	0	16.51	−18.62	51.64
1	1	1	0	−323.05	−353.90	−292.19
1	1	0	0	−209.80	−252.07	−167.53
0	0	1	1	62.94	42.74	83.13
0	0	0	0	−87.90	−142.99	−32.82
0	0	1	0	−139.56	−176.84	−102.28
0	1	1	1	−75.77	−117.24	−34.31
0	0	1	1	415.36	386.92	443.79
0	1	0	1	569.83	535.55	604.12

† Marker code for each of four AFLP markers (Table 7)

‡ LL and UL are lower limit and upper limit of 95 % confidence interval, respectively

traits (Ritchie et al. 2001; Hahn et al. 2003; Nelson et al. 2001; Culverhouse et al. 2004), the modified MDR method treated all allelic combinations as different new genotypes with different genotypic effects. Thus, these predicted effects for different allelic combinations can be used for MAS for a trait. On the other hand, based on the 95 % confidence interval for each predicted genotypic effect, multiple comparisons among different allelic combinations can be conducted accordingly. For example, two allelic combinations had genotypic effects (which were also significantly different) for lint yield that were greater than the remaining allelic combinations (Table 6). These two allelic combinations from these four markers can be used for MAS for lint yield.

By using the MMDR method in this cotton data set, two-way interaction effects were detected for lint percentage but not for the other seven traits. Interaction effects among three and four AFLP markers were more common for these traits. On the other hand, the best four-marker combinations did not include the best two- or three-marker combinations for the same traits. The results suggested that high intensive computations may be expected when more markers are used for detecting epistatic effects. Compared to multiple linear regression analysis (Wu et al. 2007) and additive models in this study, the MMDR method showed better prediction for genotypic values for these agronomic and fiber traits. Thus, MMDR method is recommended when epistatic effects exist.

It is very likely that some marker combinations had similar contribution to a target trait. On one hand, the data suggests that different marker groups can be used equally well for MAS for a trait. On the other hand, it indicates that different gene path ways may be involved in controlling a trait. If EST and/or SNP markers, which are related to potential gene functions, are used, then additional genetic information related to these markers can be further investigated via bioinformatics tools.

Many previous methods to identify the best marker groups from available DNA markers are combinatorial (Ritchie et al. 2001; Hahn et al. 2003; Lou et al. 2007, 2008). In the same way, the combinatorial approach was integrated into the MMDR method in this study. Since these methods are global search algorithm, they can be very time consuming. For example, to identify the best five-marker group among 50 DNA markers requires comparing the results from a total of 2,118,760 combinations. In many studies, the number of markers can range from a few hundred to several thousands. Obviously, globally search is computational prohibited. Alternative approaches such as ensemble approach, RPM method, and genetic algorithm can be used to identify desirable marker groups (Culverhouse et al. 2004). It will be worthwhile to incorporate these alternative algorithms into the mixed model based MDR method so that determining high order epistatic effects will be computationally feasible when a large number of DNA markers are used.

Data collected for association mapping from human genetics normally are not repeated. Permutation test or k-fold cross-validation are used to test the best model (Martin et al. 2006; Lou et al. 2008). In this study, we used a k-fold jackknife procedure, which was addressed in a separate manuscript. In brief, the data set was randomly divided into 10 groups with nearly equal sizes. At each time, one of 10 groups data set was removed from the original data set, then a non-pseudo value based *t* test was applied (Wu et al. 2008). From our various simulation studies, the tenfold jackknife method appears robust to control Type I error yet holds a desirable test power.

Though AFLP marker data were used for this study, the modified MDR approach in this study can be applied to other types of DNA markers and various genetic mapping studies including QTL mapping. For

Table 7 Distribution of 20 AFLP markers associated with seven cotton traits (Wu et al. 2007)

Entry	Marker code																			
	1 E1M1-106	2 E1M2-39	3 E1M3-168	4 E1M4-153	5 E1M4-203	6 E1M4-335	7 E1M4-421	8 E3M5-137	9 E3M5-183	10 E3M5-405	11 E3M6-219	12 E3M6-236	13 E3M6-288	14 E3M8-201	15 E5M6-260	16 E6M1-314	17 E6M3-266	18 E7M6-179	19 E8M5-231	20 E8M8-229
1	0[†]	1	1	1	0	0	0	0	1	1	1	0	1	1	0	0	0	1	1	1
2	0	1	1	1	1	0	0	0	1	1	1	1	1	1	0	0	0	1	1	1
3	0	1	1	0	1	0	0	0	1	1	1	0	1	1	0	0	0	1	0	1
4	0	0	1	1	1	0	0	0	1	1	1	1	1	1	0	0	0	1	0	1
5	0	1	1	1	1	0	0	0	1	1	0	1	1	1	0	0	0	0	0	1
6	0	1	1	0	0	0	0	0	1	1	0	0	0	0	1	1	0	1	1	1
7	0	1	1	0	0	0	0	0	1	1	0	0	0	0	1	0	0	1	1	1
8	0	1	1	0	0	0	0	0	1	1	0	0	0	0	1	0	0	1	1	1
9	0	1	1	0	0	0	1	0	1	1	0	0	0	0	1	1	0	1	1	1
10	0	1	1	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	1
11	0	1	1	0	0	0	0	0	0	1	0	0	1	0	1	1	1	1	1	1
12	0	1	1	0	0	0	0	0	1	1	0	0	0	0	1	0	1	1	1	1
13	0	1	1	0	0	0	0	0	1	1	0	0	1	0	1	0	0	1	1	1
14	0	1	1	0	0	0	0	0	1	1	0	0	0	0	1	1	0	1	0	1
15	0	1	1	0	0	0	0	0	1	1	0	0	0	0	1	0	0	1	0	1
16	0	1	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	1	1	1
17	0	1	1	0	0	0	0	0	1	1	0	1	0	1	0	1	0	1	1	0
18	0	1	1	0	0	0	0	0	1	1	0	1	0	0	0	0	0	1	1	1
19	0	1	1	0	0	0	0	0	1	1	0	0	0	1	0	0	0	1	0	1
20	0	1	1	0	0	1	0	0	1	1	0	0	0	1	0	0	0	1	1	1
21	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	1	0
22	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	1	0
23	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0
24	1	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	1	0	1	0

[†] 0 indicates marker allele absent: 1 indicates marker allele present

example, epistatic effects among QTLs for both mapping populations such as F_2, DH, RI, and BC populations can be determined. It can also be applied to these mapping studies for epistasis-by-environment interaction effects when studies are conducted in several environments. These types of studies are worthy of further investigation in the future.

References

Broman KW (2001) Review of statistical methods for QTL mapping in experimental crosses. Lab Animal 30(7):44–52

Cao G, Zhu J, He C, Gao Y, Yan J, Wu P (2001) Impact of epistasis and QTL × environment interaction on the developmental behavior of plant height in rice (*Oryza sativa* L.). Theor Appl Genet 103:153–160

Culverhouse R, Klein T, Shannon W (2004) Detecting epistatic interactions contributing to quantitative traits. Genet Epidemiol 27:141–152

Davison AC, Hinkley DV (1998) Bootstrap methods and their application. Cambridge University Press. New York, NY

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

Foulkes AS (2009) Applied statistical genetics with R. Springer, New York, NY

Goldringer I, Brabant P, Gallais A (1997) Estimation of additive and epistatic genetic variances for agronomic traits in a population of doubled-haploid lines of wheat. Heredity 79:60–71

Guo W, Zhang T, Shen X, Yu J, Kohel RJ (2003) Development of SCAR marker linked to a major QTL for high fiber strength and its usage in molecular-marker assisted selection in upland cotton. Crop Sci 43:2252–2256

Gutierrez OA, Basu S, Saha S, Jenkins JN, Shoemaker DB, Cheatham CL, McCarty JC (2002) Genetic distance among selected cotton genotypes and its relationship with F_2 performance. Crop Sci 42:1841–1847

Hahn LW, Ritchie MD, Moore JH (2003) Multifactor dimensionality reduction software for detecting gene–gene and gene–environment interactions. Bioinformatics 19:376–382

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Hartley HO, Rao JNK (1967) Maximum-likelihood estimation for the mixed analysis of variance model. Biometrika 54:93–108

He D, Lin Z, Zhang X, Nie Y, Guo X, Feng C, Stewart JM (2005) Mapping QTLs of traits contributing to yield and analysis of genetic effects in tetraploid cotton. Euphytica 144:141–149

Jansen RC (1992) A general mixture model for mapping quantitative trait loci by using molecular markers. Theor Appl Genet 85:252–260

Jansen RC (1993) Interval mapping of multiple quantitative trait loci. Genetics 135:205–211

Jiang C, Wright RJ, El-Zik KM, Paterson AH (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). Proc Natl Acad Sci USA 95: 4419–4424

Jiang C, Wright RJ, Woo SS, DelMonte TA, Paterson AH (2000) QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton). Theor Appl Genet 100:409–418

Knapp SJ (1998) Marker-assisted selection as a strategy for increasing the probability of selecting superior genotypes. Crop Sci 38:1164–1174

Kohel RJ, Yu J, Park YH, Lazo GR (2001) Molecular mapping and characterization of traits controlling fiber quality in cotton. Euphytica 121:163–172

Lacape JM, Nguyen TB, Courtois B, Belot JL, Giband M, Gourlot JP, Gawryziak G, Roques S, Hau B (2005) QTL analysis of cotton fiber quality using multiple *Gossypium hirsutum* × *Gossypium barbadense* backcross generations. Crop Sci 45:123–140

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756

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

Lin Z, He D, Zhang X, Nie Y, Guo X, Feng C, Stewart JM (2005) Linkage map construction and mapping QTL for cotton fibre quality using SRAP, SSR and RAPD. Plant Breed 124:180–187

Liu P, Zhu J, Lou XY, Lu Y (2003) A method for marker-assisted selection based on QTLs with epistatic effects. Genetica 119:75–86

Lou XY, Chen GB, Yan L, Ma JZ, Zhu J, Elston RC, Li MD (2007) A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence. Am J Human Genet 80:1125–1137

Lou XY, Chen GB, Yan L, Ma JZ, Mangold JE, Zhu J, Elston RC, Li MD (2008) A combinatorial approach to detecting gene–gene and gene–environment interactions in family studies. Am J Human Genet 83:457–467

Lü HY, Liu XF, Wei SP, Zhang YM (2011) Epistatic association mapping in homozygous crop cultivars. PLoS One 6(3): e17773

Manly BFJ (2006) Randomization, bootstrap and Monte Carlo methods in biology, 3rd edn. Chapman and Hall/CRC, Boca Raton

Martin ER, Ritchie MD, Hahn L, Kang S, Moore JH (2006) A novel method to identify gene-gene effects in nuclear families: the MDR-PDT. Genet Epidemiol 2:111–123

Martin EZO, Curnow RN (1992) Estimation the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor Appl Genet 85:480–488

McCarty JC, Jenkins JN, Parrott WL, Creech RG (1979) The conversion of photoperiodic primitive race stocks of cotton to day-neutral stocks. Miss Agric For Exp Stn Res Rep 4(19):4

McCarty JC, Jenkins JN, Wu J (2004a) Primitive accession germplasm by cultivar crosses as sources for cotton improvement I: phenotypic values and variance components. Crop Sci 44:1226–1230

McCarty JC, Jenkins JN, Wu J (2004b) Primitive accession germplasm by cultivar crosses as sources for cotton

improvement II: genetic effects and genotype values. Crop Sci 44:1231–1235

McCarty JC, Wu J, Jenkins JN (2008) Genetic associations of cotton yield with its component traits in derived primitive accessions crossed by elite upland cultivars using the conditional ADAA genetic model. Euphytica 161:337–352

Mei M, Syed NH, Gao W, Thaxton PM, Smith CW, Stelly DM, Chen Z (2004) Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). Theor Appl Genet 108:280–291

Miller RG (1974) The jackknife: a review. Biometrika 61:1–15

Moore JH, Gilbert JC, Tsai CT, Chiang FT, Holden T, Barney N, White BC (2006) A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. J Theor Biol 241:252–261

Nelson MR, Kardia SL, Ferrell RE, Sing CF (2001) A combinatorial partitioning method to identify multilocus genotypic partitions that predict quantitative trait variation. Genome Res 11:458–470

Patterson HD, Thompson R (1971) Recovery of inter-block information when block size are unequal. Biometrika 58:545–554

Rao CR (1971) Estimation of variance and covariance components MINQUE theory. J Multivar Anal 1:257–275

Reinisch AJ, Dong J, Brubaker CL, Stelly DM, Wendel JF, Paterson AH (1994) A detailed RFLP map of cotton, *Gossypium hirsutum* × *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploidy genome. Genetics 138:829–847

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH (2001) Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Human Genet 69:138–147

Saha S, Wu J, Jenkins JN, McCarty JC, Hayes RW, Stelly DM (2010) Genetic dissection of chromosome substitution lines of cotton to discover novel *Gossypium barbadense* L. alleles for improvement of agronomic traits. Theor Appl Genet 120:1193–1205

Searle SR, Casella G, McCulloch CE (1992) Variance components. Wiley, New York

Shen S, Zhang T, Guo W, Zhu X, Zhang X (2006) Mapping fiber and yield QTLs with main, epistatic, and QTL × environment Interaction effects in recombinant inbred lines of upland cotton. Crop Sci 46:61–66

Stich B, Möhring J, Piepho HP, Heckenberger M, Buckler ES, Melchinger AE (2008) Comparison of mixed-model approaches for association mapping. Genetics 178:1745–1754

Stromberg LD, Dudley JW, Rufener GK (1994) Comparing conventional early generation selection with molecular marker assisted selection in maize. Crop Sci 34:1221–1225

Swindle MG (1993) Performance of F_2 hybrids in cotton from crosses of exotic germplasm and cultivars. M.S. thesis, Mississippi State University, Mississippi State, MS

Tang B, Jenkins JN, Watson CE, McCarty JC, G. Creech RG (1996) Evaluation of genetic variances, heritabilities, and correlations for yield and fiber traits among cotton F2 hybrid populations. Euphytica 91:315–322

Ulloa M, Saha S, Jenkins JN, Meredith WR, McCarty JC, Stelly DM (2005) Chromosomal assignment of RFLP linkage groups harboring important QTLs on an intraspecific cotton (*Gossypium hirsutum* L.) join map. J Hered 96:132–144

Wang DL, Zhu J, Li ZK, Paterson AH (1999) Mapping QTLs with epistatic effects and QTL × environment interactions by mixed linear model approaches. Theor Appl Genet 99:1255–1264

Wu J, Jenkins JN, McCarty JC (2011) A Generalized approach and computer tool for quantitative genetics study. Proceedings of Applied Statistics in Agriculture April 25–27, 2010, Manhattan, KS, USA, pp 85–106

Wu J, Jenkins JC, McCarty JC, Wu D (2006) Variance component estimation using the ADAA model when genotypes vary across environments. Crop Sci 46:174–179

Wu J, Jenkins JN, McCarty JC, Zhong M, Swindle M (2007) AFLP marker associations with agronomic and fiber traits in cotton. Euphytica 153:153–163

Wu J, Jenkins JN, McCarty JC (2008) Testing variance components by two jackknife techniques. In: Proceedings of applied statistics in agriculture, Manhattan, pp 1–17

Xing YZ, Tan YF, Hua JP, Sun XL, Xu CG, Zhang Q (2002) Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. Theor Appl Genet 105:248–257

Xu ZC, Zhu J (1999) A new approach for predicting heterosis based on an additive, dominance and additive × additive model with environment interaction. J Hered 82:510–517

Yang J, Zhu J, Williams RW (2007) Mapping the genetic architecture of complex traits in experimental populations. Bioinformatics 23:1527–1536

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38:203–208

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

Zhang T, Yuan Y, Yu J, Guo W, Kohel RJ (2003) Molecular tagging of major QTL for fiber strength in Upland cotton and its marker-assisted selection. Theor Appl Genet 106:262–268

Zhang K, Tian J, Zhao L, Wang S (2008) Mapping QTLs with epistatic effects and QTL × environment interactions for plant height using a doubled haploid population in cultivated wheat. J Genetics Genomics 35:119–127

Zhong M (2001) Comparative analyses of backcrossed derived lines from exotic germplasm for day-neutral genes using AFLP markers. M.S. thesis, Mississippi State University, Mississippi State, MS

Zhong M, McCarty JC, Jenkins JN, Saha S (2002) Assessment of day-neutral backcross populations of cotton using AFLP markers. J Cotton Sci 6:97–103

Zhu J (1989) Estimation of genetic variance components in the general mixed model. Ph.D. dissertation, North Carolina State University, Raleigh, NC

Zhu J (1993) Methods of predicting genotype value and heterosis for offspring of hybrids. J Biomath 8:32–44

Zhu J, Weir BS (1994) Analysis of cytoplasmic and maternal effects: I. A genetic model for diploid plant seeds and animals. Theor Appl Genet 89:153–159