

Joint tests for quantitative trait loci in experimental crosses

T. Mark BEASLEY^{a*}, Dongyan YANG^a, Nengjun YI^a,
Daniel C. BULLARD^b, Elizabeth L. TRAVIS^c, Christopher I. AMOS^d,
Shizhong XU^e, David B. ALLISON^{a,f}

^a Department of Biostatistics, Section on Statistical Genetics, University of Alabama
at Birmingham, Birmingham, AL, USA

^b Department of Genomics and Pathobiology, University of Alabama at Birmingham,
Birmingham, AL, USA

^c Department of Experimental Radiation Oncology, University of Texas, M.D. Anderson
Cancer Center, Houston, TX, USA

^d Department of Epidemiology, University of Texas, M.D. Anderson Cancer Center Houston,
TX, USA

^e University of California, Riverside, CA, USA

^f Clinical Nutrition Research Center, University of Alabama at Birmingham,
Birmingham, AL, USA

(Received 16 February 2004; accepted 24 May 2004)

Abstract – Selective genotyping is common because it can increase the expected correlation between QTL genotype and phenotype and thus increase the statistical power of linkage tests (*i.e.*, *regression-based* tests). Linkage can also be tested by assessing whether the marginal genotypic distribution conforms to its expectation, a *marginal-based* test. We developed a class of joint tests that, by constraining intercepts in regression-based analyses, capitalize on the information available in both *regression-based* and *marginal-based* tests. We simulated data corresponding to the null hypothesis of no QTL effect and the alternative of some QTL effect at the locus for a backcross and an F2 intercross between inbred strains. Regression-based and marginal-based tests were compared to corresponding joint tests. We studied the effects of random sampling, selective sampling from a single tail of the phenotypic distribution, and selective sampling from both tails of the phenotypic distribution. Joint tests were nearly as powerful as all competing alternatives for random sampling and two-tailed selection under both backcross and F2 intercross situations. Joint tests were generally more powerful for one-tailed selection under both backcross and F2 intercross situations. However, joint tests cannot be recommended for one-tailed selective genotyping if segregation distortion is suspected.

joint tests / quantitative trait loci / linkage / F2 cross / backcross

* Corresponding author: MBeasley@UAB.edu

1. INTRODUCTION

Selective genotyping is a common approach used to enhance the efficiency of quantitative trait loci (QTL) mapping studies [13, 25], which employs an extreme threshold (ET) design and entails analyzing only a subset of individuals with extreme scores. In an ET2 design, individuals are sampled from both tails of the phenotypic distribution (*i.e.*, cases with unusually high and low values of the phenotype). ET2 designs have been shown to decrease uncertainty about the underlying QTL genotypes, yield valid false positive rates, and increase the statistical power per genotyped individual [1, 7, 13] because the expected correlation between genotype and phenotype generally increases [3]. For example, Allison [2] showed that the ET2 design increased the power of his TDT_{Q5}. However, there is a trade-off between (a) increasing the correlation through extreme sampling and (b) reducing the overall statistical power due to the reduction in sample size. The association between genotype and phenotype has been the focus of tests in QTL mapping, including studies of experimental crosses. We refer to tests that evaluate whether the distribution of the phenotype (Y) is dependent on some function of the genotype (G) as *regression-based tests*.

It is also common for genetics researchers to use an ET1 design and sample from only one tail of the phenotypic distribution. The ET1 design is similar in concept to “case-only” designs often used in human studies [15]. However, ET1 designs decrease the power of regression-based tests due to a restriction of range [16]. It is important to note that, when (and only when) the null hypothesis is false, extreme sampling can also affect the marginal distribution of genotypes. That is, under the null hypothesis of no linkage, the marginal distribution of genotypes has the same expected frequencies regardless of the phenotypic value. For example, in an experimental $BB \times BD$ backcross, all offspring would be either BB or BD and these two genotypes would be equally likely, assuming no segregation distortion. Under the null hypothesis of no linkage, Y is not related to the genotype (G). Likewise, G is not related to Y and the probability of sampling a case with either BB or BD genotype should be equal regardless of Y , $P(BB|Y) = P(BD|Y) = 1/2$, assuming no segregation distortion. Recognizing this, one can construct tests of linkage when ET designs are used by testing for departures from the genotypic distribution that would be expected under the null hypothesis. We refer to such tests as *marginal-based tests*. Lander and Botstein [13] have provided considerable detail on increasing the power of QTL mapping by selective genotyping of progeny with extreme phenotypes in backcross designs. Similar discussions that include F2 intercross designs appear in [5] and [20]. Nevertheless, marginal-based tests have been

underutilized in the development of QTL mapping procedures for experimental crosses.

In this paper, we develop methods that capitalize on the information available in both regression-based and marginal-based tests of linkage for experimental crosses. We show that these tests are rarely less powerful and are usually more powerful than regression-based or marginal-based tests alone. Moreover, the tests we have developed are easily implemented in standard software, should be robust to non-normality, can be applied to either backcross or F2 intercross designs, and allow for extreme sampling with either ET1 or ET2 sampling. In developing these tests, we assume that there is no segregation distortion. However, we note that the marginal-based and joint tests rely crucially on this assumption, especially in ET1 designs. Therefore, we examine the statistical properties of these tests when segregation distortion is present. We also discuss how the tests herein should be used if segregation distortion is suspected.

2. INDIVIDUAL TESTS OF LINKAGE

Before proceeding further, it will be useful to define the specific tests of linkage that we employed (see Tab. I). We considered two types of experimental crosses: A backcross and an F2 intercross. Let the two parental strains be denoted BB and DD . Assume that the backcross utilized is one between the BB strain and the $BB \times DD$ F1. Then, at each locus, progeny in a backcross can have either BB or BD genotypes. Scoring these by the number of D alleles, the corresponding genotypic values would be $G = 0$ and 1, respectively. For the F2 intercross, BB , BD , and DD genotypes would be scored $G = 0$, 1, and 2, respectively.

2.1. Regression-based tests

The first two regression-based tests involve ordinary least squares (OLS) regression in which phenotype (Y) is regressed on genotype (G):

$$E[Y|G] = \beta_0 + \beta_1 G. \quad (1)$$

R_1 refers to treating G as a continuous variable with a 1 degree-of-freedom (df) test and testing the null hypothesis that the slope (β_1) equals zero. R_2 refers to treating G as a categorical variable with a 2 df test of the null hypothesis that both slopes (β_1 and β_2) equal zero to allow for departures from additivity:

$$E[Y|A, D] = \beta_0 + \beta_1 A + \beta_2 D, \quad (2)$$

where A and D are linear and quadratic polynomial contrast variables, respectively. We note that R_2 cannot be applied to backcross designs because there are only two genotypes and thus 1 *df*. For F2 intercrosses, however, both R_1 and R_2 can be applied. Although OLS regression procedures can be used to estimate linkage parameters with selective genotyping, the estimates are expected to be biased, and thus, a maximum likelihood procedure for obtaining unbiased estimates has been suggested [13]. For F2 intercrosses, we define R_3 and R_4 as the maximum likelihood procedure of Xu and Vogl [25] applied to the linear models (1) and (2), respectively. Briefly, this technique is a simple modification of the EM algorithm for assessing linkage for selective genotyping using only the phenotypic values of genotyped individuals.

The fifth and sixth tests depend on whether the experiment involves a backcross or F2 intercross. For backcross designs, R_5 is calculated by regressing the genotype (G) on phenotype (Y) using logistic regression [11] and testing the null hypothesis that the slope (β_1) equals zero:

$$\ln \left[\frac{P(G = 1)}{P(G = 0)} \right] = \ln \left[\frac{P(BD)}{P(BB)} \right] = \beta_0 + \beta_1 Y. \quad (3)$$

This method was proposed for binary variables and thus is not generally applicable to F2 intercross designs. In the case of an F2 intercross, we define R_6 as multinomial regression with three categories for the response variable, which requires estimating two slopes and two intercepts:

$$\begin{aligned} \ln \left[\frac{P(G = 1)}{P(G = 0)} \right] &= \ln \left[\frac{P(BD)}{P(BB)} \right] = \beta_0 + \beta_1 Y \\ \ln \left[\frac{P(G = 2)}{P(G = 0)} \right] &= \ln \left[\frac{P(DD)}{P(BB)} \right] = \gamma_0 + \gamma_1 Y. \end{aligned} \quad (4)$$

Thus, R_6 is a 2 *df* test of whether both slopes (β_1 and γ_1) are equal to zero.

2.2. Marginal-based tests

Under the null hypothesis of no linkage $P(G = 0) = P(G = 1) = 1/2$ in the backcross and $P(G = 0) = P(G = 2) = 1/4$ and $P(G = 1) = 1/2$ for the F2 intercross, assuming no segregation distortion, and thus, the expected genotypic mean in the backcross is $E[G|Y] = \mu_G = 1/2$ and the expected genotypic mean in the F2 intercross is $\mu_G = 1$, regardless of the value of the phenotype (Y).

We defined six marginal-based tests, three for each ET sampling design. For ET1 designs, M_7 is defined as a single-sample t -test of whether the mean of G is different from its null expectation (μ_G). Specifically, $\mu_G = 1/2$ in a backcross and $\mu_G = 1$ in an F2 intercross. As alternatives, we utilize chi-square goodness of fit tests. For the backcross, we define M_8 as a 1 $df \chi^2$ test of G versus expected frequencies of $P(G = 0) = P(G = 1) = 1/2$. For F2 crosses, we define M_9 as a 2 $df \chi^2$ test of whether the sample frequencies for G departs from the null expectation of $P(G = 0) = 1/4$, $P(G = 1) = 1/2$, and $P(G = 2) = 1/4$.

We note that these marginal tests rely heavily on the assumption of random segregation in the ET1 design; however, for an ET2 design, this is not necessarily the case. In family-based studies, test statistics that incorporate information from both affected and unaffected siblings are used to control for segregation distortion [22]. Likewise, for QTL studies the use of information from both ends of the distribution will control for segregation distortion [2]. Under the null hypothesis of no linkage, the marginal distribution of genotypes has the same expected frequencies regardless of the phenotypic value. Therefore, the upper and lower tails will have the same expected values of G (same genotype frequencies) under the null hypothesis regardless of whether or not there is segregation distortion. There are standard statistical tests that can be applied as marginal-based tests for an ET2 design. For either an F2 or backcross design, we define M_{10} as an independent samples t -test to assess whether the mean of G is equal for the upper and lower tails. As alternatives, we utilize chi-square tests of independence. For a backcross design, we define M_{11} as a 2×2 (e.g., BB vs. BD by Upper vs. Lower) chi-square test with $df = 1$. For an F2 intercross, we define M_{12} as a 3×2 (e.g., BB vs. BD vs. DD by Upper vs. Lower) chi-square test with $df = 2$.

3. JOINT TESTS OF LINKAGE

In the context of human IBD-based QTL mapping in sib-pair studies, Forrest and Feingold [8] provide proof that under the null hypothesis of no linkage, regression-based tests and marginal-based tests are independent. Therefore, one way to construct composite tests that capitalize on the information from regression-based and marginal-based test statistics is simply to sum them up and treat them as χ^2 with df equal to the sum of the df of the two tests being combined. We introduced *joint tests* that do not require the asymptotic independence of the tests, which we found to be more powerful than composite tests in preliminary studies.

We modified the Henshall and Goddard [11] approach, which reverses the position of dependent and independent variables in a regression model (*i.e.*, regressing genotype on phenotype). Our modification involves constraining the intercept to have a pre-specified value based on expectations from the marginal distribution of the genotype given the experimental cross. Large test statistics reflect deviations from the null hypothesis of no association between G and Y and deviations from the genotype frequencies expected under the null hypothesis of no linkage. Thus, these methods provide *joint tests* of the null hypotheses for the regression-based and marginal-based tests. Sham *et al.* [21] present a similar approach in the context of human linkage studies.

To employ OLS regression, prior to regressing the genotype on the phenotype, we transform G to $G^* = G - E[G]$, where $E[G] = 1/2$ in a BB \times BD backcross, $1^{1/2}$ in a DD \times BD backcross, and 1 in an F2 intercross. One can then center Y , $Y^* = Y - \bar{Y}$, and regress G^* on the Y^* and force the regression through the origin:

$$E[G^*|Y^*] = \beta_0 + \beta_1 Y^*, \quad (5)$$

with $\beta_0 \equiv 0$. This offers a single *df* test that will be sensitive to departures from both the null expectation of $G^* = 0$ and the null covariance between G^* and the phenotype. We denote this OLS-based joint test as J_{13} .

Although OLS should be robust to the non-normality of residuals that will occur when G^* is used as the dependent variable given the sample sizes typically used in QTL mapping, logistic regression offers an alternative that models the categorical nature of the genotypes and avoids the normality assumption. In the case of a BB \times BD backcross we can simply regress G on Y^* as in model (3), except that we constrain the estimate of $\beta_0 \equiv 0$. This is because under the null hypothesis, $\beta_1 = 0$, and thus, $\ln [P(BD)/P(BB)] = \beta_0$. Also, under the null hypothesis, $P(BD) = P(BB) = 1/2$, which implies that $\beta_0 = \ln [P(BD)/P(BB)] = 0$. Thus, we define J_{14} as the 1 *df* test that $\beta_1 = 0$ while restricting β_0 to be 0.

In the case of an F2 intercross, we can replace binary logistic regression with multinomial regression and regress G on Y^* as in model (4). However, in this context, treating BB as the “reference” genotype, we can constrain $\beta_0 \equiv \ln [2]$ because under the null hypothesis $\beta_1 = 0$, $P(BD) = 1/2$, and $P(BB) = 1/4$, which implies that $\beta_0 = \ln [P(BD)/P(BB)] = \ln [2]$. Likewise, we constrain $\gamma_0 \equiv 0$ because under the null hypothesis $\gamma_0 = 0$ and $P(DD) = P(BB) = 1/4$, which implies that $\gamma_0 = \ln [P(DD)/P(BB)] = 0$. This allows the logistic

regression approach to be extended to the F2 intercross design and also to accommodate marked nonadditivity in the genotype-phenotype relationship. We denote the joint tests involving multinomial regression with constrained intercepts as J_{15} .

4. SIMULATION STUDIES

To demonstrate the validity of our joint tests with respect to Type 1 error rates and to evaluate their power relative to the marginal-based and regression-based tests, we conducted a variety of simulations. Table I provides a summary of the tests compared in these simulations. To evaluate Type 1 error rates, simulations were conducted under the null hypothesis of no linkage. To evaluate Type 2 error rates (*i.e.*, statistical power), the basic model used in the simulations was that of a quantitative trait with a single major QTL. For the non-null situations, the proportion of phenotypic variance explained by the QTL was fixed at $h^2 = 3\%$, 5% , 8% , and 11% of the total phenotypic variance in two separate sets of simulations for backcross and F2 intercross designs. Additive and non-additive (dominant) models were simulated. The residual within genotype distribution was normal with a mean of zero and unit variance.

Type 1 and Type 2 errors were evaluated at a significance level of $\alpha = 0.0001$. For simulations under the null model, 100 000 simulated datasets were used for each situation to ensure reasonable precision for an alpha level as small as 0.0001. For simulations under the alternative hypothesis, 10 000 simulated datasets were used for each situation. A total sample size of $N = 500$ progeny was used in all the simulations.

Three sampling schemes were considered: (1) Random sampling. All 500 progeny were analyzed; (2) Selection from both tails of the phenotypic distribution (ET2 design). The 500 progeny were ranked with respect to their phenotypic values and the top and bottom 125 (50%) or 50 (20%) progeny were selected for genotyping and analysis; and (3) selection from one tail of the phenotypic distribution (ET1 design). The 500 progeny were ranked with respect to their phenotypic values and the top 250 (50%) or 100 (20%) progeny were selected for genotyping and analysis.

Because segregation distortion is often seen in crosses between inbred lines of both plants and animals, two conditions of allelic segregation were imposed. One condition is random segregation (no segregation distortion) where the probability of the offspring receiving the D allele during meiosis is 0.5. The second condition simulates segregation distortion where the probability of the offspring receiving the D allele during meiosis is 0.7.

Table I. Summary of individual tests considered.

Tests		Description	Applicable crosses	Sampling designs	Dominant variance?	Referent distribution
Regression based tests	R ₁	1 <i>df</i> OLS regression (Eq. 1)	Backcross F2	ET1 ET2	No	$F(1, N-2)$
	R ₂	2 <i>df</i> OLS regression (Eq. 2)	F2	ET1 ET2	Yes	$F(2, N-3)$
	R ₃	1 <i>df</i> ML regression (Eq. 1) (Xu & Vogl, 2000)	Backcross F2	ET1 ET2	No	$F(1, N-2)$
	R ₄	2 <i>df</i> ML regression (Eq. 2) (Xu & Vogl, 2000)	F2	ET1 ET2	Yes	$F(2, N-3)$
	R ₅	Logistic regression (Eq. 3) (Henshall & Goddard, 1999)	Backcross	ET1 ET2	No	$\chi^2(1)$
	R ₆	Multinomial regression (Eq. 4)	F2	ET1 ET2	Yes	$\chi^2(2)$
Marginal based test	M ₇	Single-sample <i>t</i> -test on <i>G</i>	Backcross F2	ET1	No	$t(N-1)$
	M ₈	1 <i>df</i> χ^2 Goodness of fit	Backcross	ET1	No	$\chi^2(1)$
	M ₉	2 <i>df</i> χ^2 Goodness of fit	F2	ET1	Yes	$\chi^2(2)$
	M ₁₀	Independent-sample <i>t</i> -test	Backcross F2	ET2	Yes	$t(N-2)$
	M ₁₁	$2 \times 2\chi^2$ Test of independence	Backcross	ET2	No	$\chi^2(1)$
	M ₁₂	$3 \times 2\chi^2$ Test of independence	F2	ET2	Yes	$\chi^2(2)$
Joint tests	J ₁₃	1 <i>df</i> OLS regression (Eq. 6) $\beta_0 \equiv 0$	Backcross F2	ET1 ET2	No	$F(1, N-1)$
	J ₁₄	1 <i>df</i> Logistic regression (Eq. 3) $\beta_0 \equiv 0$	Backcross	ET1 ET2	No	$\chi^2(1)$
	J ₁₅	2 <i>df</i> Multinomial regression (Eq. 4) $\beta_0 \equiv \ln [2]$ $\gamma_0 \equiv 0$	F2	ET1 ET2	Yes	$\chi^2(2)$

5. RESULTS

5.1. Type 1 error rate

Tables II and III show the Type 1 error rates of all tests at $\alpha = 0.0001$ for the backcross and F2 intercross designs, respectively. These values serve as an evaluation of the conformity of the test statistics to their asymptotic distribution for relatively small sample sizes. Lander and Botstein [13] suggest that linear regression cannot be used when only extreme progeny have been genotyped because genotypic effects will be grossly overestimated because of the biased selection; however, this does not imply that the Type 1 error rate will be inflated. Our results confirmed this. For all tests considered, the empirical Type 1 error rates are very close to the nominal alpha indicating excellent conformity to the asymptotic distribution of the test statistics, when there was no segregation distortion.

When segregation distortion ($P = 0.7$) was simulated, the Type 1 error rates for the regression-based tests were basically unaffected. By contrast, the Type 1 error rates for the marginal-based tests were severely inflated when either random sampling or an ET1 design was employed (see Tabs. II and III). For the joint tests developed for a backcross design, the Type 1 error rates were inflated when there was segregation distortion ($P = 0.7$) and one-tailed (ET1) sampling (see Tab. II). Similarly for the joint tests developed for an F2 design, the Type 1 error rates were inflated when there was segregation distortion ($P = 0.7$) and ET1 sampling (see Tab. II), but there was also some inflation in the false positive rate under a Random and ET2 sampling for the joint test involving multinomial regression with fixed intercepts (M_{15}). The results for selective sampling of $N = 250$ were very similar and for a brevity that was not displayed.

5.2. Statistical power

In some cases, the empirical power rates reached the maximum of unity; however, the tests demonstrated low to moderate statistical power in many other cases. We note that the power curves for the maximum likelihood regression tests (R_3 and R_4) were so similar to their OLS counterparts that for graphic clarity we did not plot their results.

Table II. Empirical Type 1 error rates under the null hypothesis (Model 1) with $\alpha = 0.0001$ for backcross design. (100 000 simulations per row).

		Random	$N = 500$	ET1	$N = 100$	ET2	$N = 100$
Tests		$P = .5$	$P = .7$	$P = .5$	$P = .7$	$P = .5$	$P = .7$
Regression based tests	R ₁	.00010	.00006	.00008	.00010	.00013	.00012
	R ₃	.00011	.00010	.00010	.00012	.00011	.00013
	R ₅	.00007	.00004	0	.00001	.00005	.00001
Marginal based tests	M ₇	.00011	1	.00019	.6315	–	–
	M ₈	.00011	1	.00008	.5472	–	–
	M ₁₀	–	–	–	–	.00018	.00014
	M ₁₁	–	–	–	–	.00008	.00010
Joint tests	J ₁₃	.00010	.00014	.00012	.5141	.00013	.00004
	J ₁₄	.00008	.00010	.00003	.3902	.00005	0

Table III. Empirical Type 1 error rates under the null hypothesis (Model 1) with $\alpha = 0.0001$ for F2 intercross design. (100 000 simulations per row).

		Random	$N = 500$	ET1	$N = 100$	ET2	$N = 100$
Tests		$P = .5$	$P = .7$	$P = .5$	$P = .7$	$P = .5$	$P = .7$
Regression based tests	R ₁	.00007	.00014	.00008	.00009	.00018	.00012
	R ₂	.00008	.00009	.00012	.00027	.00016	.00017
	R ₃	.00009	.00010	.00011	.00013	.00009	.00011
	R ₄	.00008	.00009	.00024	.00029	.00007	.00009
	R ₆	.00003	.00007	0	.00001	.00002	.00002
Marginal based tests	M ₇	.00008	1	.00012	.9713	–	–
	M ₉	.00005	1	.00015	.9462	–	–
	M ₁₀	–	–	–	–	.00015	.00012
	M ₁₂	–	–	–	–	.00006	.00004
Joint tests	J ₁₃	.00006	.00015	.00015	.9345	.00016	.00002
	J ₁₅	.00005	.00095	.00002	.8455	.00006	.00026

5.2.1. Backcross designs

Figure 1 shows that when there was no segregation distortion the regression-based and the joint tests had virtually identical power; whereas, the marginal-based test had virtually no statistical power. When segregation distortion was present, the joint tests showed a slight power advantage over the regression-based tests. Figure 2 shows that with an ET1 design the joint tests demonstrated a considerable power advantage over the marginal-based tests, while the regression-based tests had minimal power due to restriction of range. However, this power advantage dissipated with the reduction of the sample size from $N = 250$ to 100. When segregation distortion was present only the

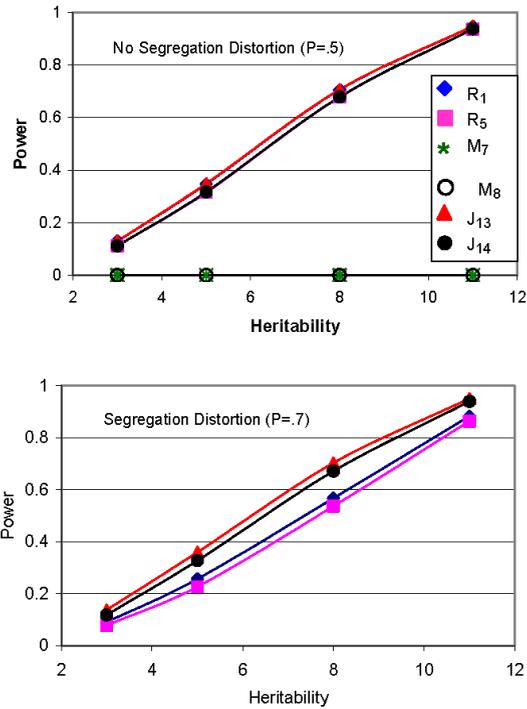


Figure 1. Power results for backcross design with additive mode of inheritance and random sampling of $N = 500$. [Marginal tests (M_7 and M_8) did not have acceptable (*i.e.*, inflated) Type 1 error rates under segregation distortion and thus were excluded from the second panel].

regression-based tests demonstrated acceptable Type 1 error rates under ET1 sampling and therefore were the only valid tests under these circumstances. For ET2 designs, the joint tests had very similar power curves as both the regression and marginal tests. For ET2 designs, the results indicate that most procedures have similar power curves when segregation distortions were present, especially the regression-based and joint tests (results not displayed).

5.2.2. F2 Intercross designs

For an F2 intercross with no segregation distortion and random sampling of $N = 500$ progeny, the regression-based and the joint tests had virtually identical power; whereas, the marginal-based test had virtually no statistical power. The OLS single-*df* tests (R_1 and J_{13}) demonstrated more power when there was an additive model. By contrast, when there was a dominant

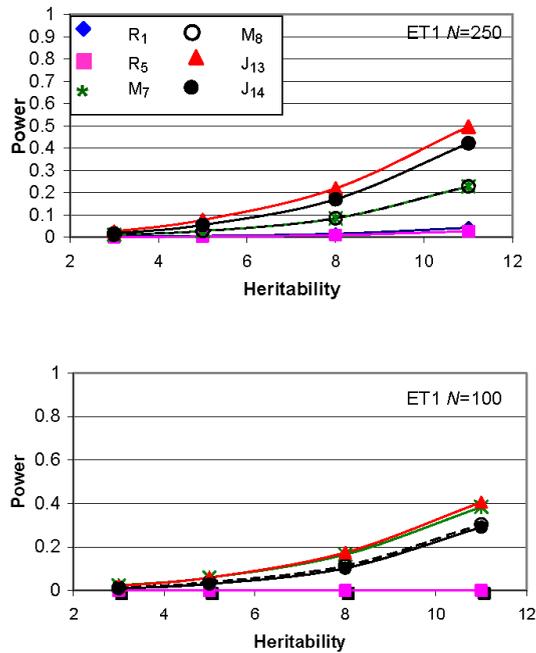


Figure 2. Power results for backcross design with no segregation distortion for one-tailed (ET1) selective genotyping.

mode of inheritance, the OLS two-*df* tests (R_2) and the multinomial regression tests (R_6 and J_{15}) demonstrated more power. For the tests that maintained valid Type 1 error rates under segregation distortion, the OLS joint test (J_{13}) demonstrated considerably more statistical power than the other regression-based tests (results not displayed).

Figure 3 displays the power curves for each test for an F2 intercross with an additive mode of inheritance and no segregation distortion. As was the case for the backcross, under an ET1 design, the joint tests demonstrated a considerable power advantage over the marginal-based tests, while the regression-based tests had minimal power. However, this power advantage dissipated with the reduction of sample size from $N = 250$ to 100, especially for the multinomial joint test (J_{15}). For ET2 designs, the joint tests had very similar power curves as both the regression and marginal tests.

For a dominant mode of inheritance under an ET2 design, the joint tests had very similar power curves as both the regression and marginal tests with R_2 and J_{15} demonstrating more power. Under one-tailed sampling, J_{13} demonstrated similar power to the marginal-based tests, whereas R_2 and J_{15} had virtually

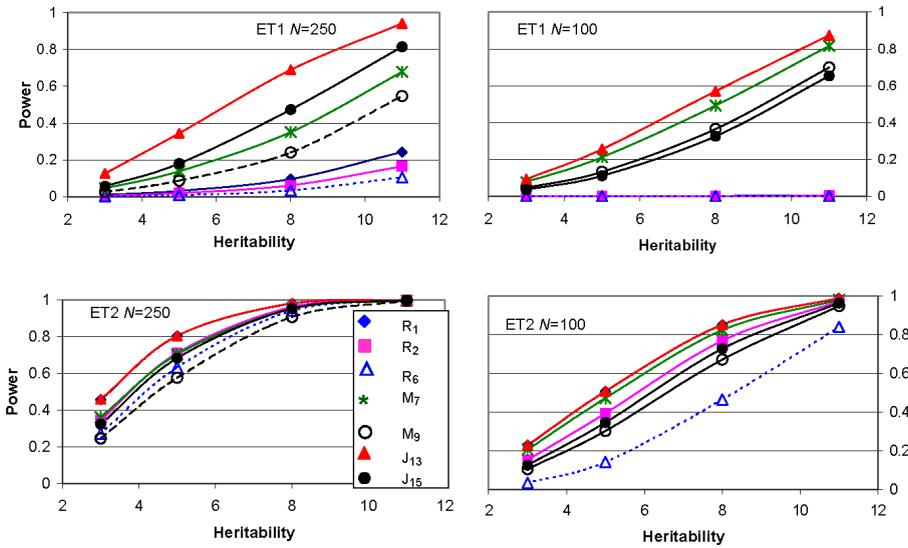


Figure 3. Power results for F2 design with additive mode of inheritance and no segregation distortion for one-tailed (ET1) and two-tailed (ET2) selective genotyping.

no power (results not displayed). This result can be attributed to the fact that with one-tailed sampling we chose the upper tail of the phenotypic distribution. Under a dominant mode of inheritance, the heterozygote (BD) and the homozygote (DD) are expected to have the same average phenotype that is greater than the average phenotype of the other homozygote (BB). By selecting the upper tail of the phenotypic distribution one may end up comparing two genotypes with the same expected value.

For an F2 intercross with additive and dominant modes of inheritance and segregation distortion, J_{13} had similar power curves to other procedures and had more power than the other procedures under an ET2 design and a dominant mode of inheritance. Again, this power advantage dissipated when the sample was reduced to $N = 100$ (results not displayed).

6. DISCUSSION

All tests held Type 1 error rate reasonably near the nominal α under random segregation. But the marginal-based and joint Tests inflated the Type 1 error rate under one-tailed selection (ET1) and segregation distortion. However, the marginal-based and joint tests had valid Type 1 error rates under two-tailed selection (ET2) even when there was segregation distortion. Although the fact

that Mendelian inheritance is nearly universal, segregation distortion is considered to be a potent evolutionary force [19]. Yet, the prevalence and importance of segregation distortion is widely debated. Although some researchers contend that segregation distortion is a rare curiosity with little evolutionary importance, it is well known that it occurs more frequently among inbred strains of plants and animals [23]. For example, Xu *et al.* [26] observed 7% to 32% of markers in inbred rice strains to demonstrate segregation distortion. Also, Liu *et al.* [14] reported that 29.4% of the 238 loci mapped in inbred soybean strains were found with segregation distortion. Regardless of the evolutionary importance of segregation distortion, we demonstrated it to be a statistical problem for the marginal and joint tests under ET1 designs. We also show that under situations of random sampling and symmetric selective sampling from both tails of the phenotypic distribution (ET2) in an F2 intercross, these tests have roughly equivalent power compared to corresponding alternative tests. In cases of selective sampling from one tail of the phenotypic distribution (ET1), these joint tests are generally more powerful than corresponding alternatives assuming no segregation distortion.

When segregation distortion was present only the regression-based approaches yield valid tests under one-tailed selective genotyping. Thus, the joint tests can only be validly employed with two-tailed selective genotyping if segregation distortion is suspected. The joint tests showed a distinct power advantage over the regression-based tests with random sampling and ET2 designs for additive models with 50% ($N = 250$) sampling. Thus, for backcross and F2 intercross designs, joint tests are recommended for analyzing data, especially if there is an additive mode of inheritance; however, joint tests are not generally recommended for non-additive modes of inheritance. Furthermore, when segregation distortion is present and an ET1 design is used, the joint tests cannot be recommended. Thus, we developed joint tests that capitalize on information available in both the marginal distribution of genotype and a genotype/phenotype association and are valid in a variety of situation, however, they should be used cautiously if segregation distortion is suspected. Therefore, we also recommend that if segregation distortion is likely, then genetic researchers should use ET2 designs when possible, in which case joint tests may have more statistical power. However, if an ET1 design is employed, the researcher should follow up significant results with more complete genotyping to investigate the possibility of segregation distortion.

Of course it is important to concede that our results only definitively apply to the conditions that were simulated. Our simulations assumed that, within genotype, phenotypic errors were normally distributed and that the error variance

was constant across genotypes. Under situations in which these conditions are not met, the relative power of the different tests may not be exactly as reported herein. However, it is noteworthy that the joint tests as constructed should be relatively robust. This is due to features of the backcross and F2 intercross designs and that the tests are based on OLS or logistic regression procedures.

The OLS-based tests assume error distributions that are Gaussian with a constant variance across the genotypes but have been shown to be relatively robust to many forms of non-normality in samples of modest size. Importantly, however, violations of the normality and homogeneous variances are more likely to reduce the statistical power of the OLS-based procedures for detecting linkage. The logistic and multinomial-based procedures are used to predict the genotype membership as a function of the phenotype. In regression terminology, the phenotype is a fixed effect, and thus, there are no distributional assumptions for the phenotype. In this regard, the tests developed herein, particularly those based on logistic (or multinomial) regression, may have some advantages over other tests that assume normality and use maximum likelihood estimation. Fan and Wang [6] have demonstrated that unequal variances and unequal sample sizes do not drastically affect the error rates of logistic regression in the two-group problem (*i.e.*, backcross). Barón [4] demonstrated that in the three-group problem (*i.e.*, F2 intercross), multinomial regression models were preferable with non-normal data, and were comparable to OLS-based procedures with normal data. It is important to note that violations of standard linear model assumptions (*i.e.*, normality; homoscedasticity) are commonplace in data from many agriculture disciplines (*e.g.*, livestock breeding). Thus, researchers in these fields have learned not to trust distribution-based *P*-values and resampling-based tests (*e.g.*, permutation; bootstrap) are generally applied. Thus, future work should examine the statistical properties of these joint tests when resampling based methods are applied.

There are several other strengths of the methods developed that should be considered. First, the extension of these methods to multivariate testing is extremely simple because one simply needs to add more phenotypes on the predictor side of the logistic or multinomial regression equations. This allows one to consider designs in which researchers are mapping genes for a binary (disease) trait and some quantitative phenotypes are also measured on all organisms. In contrast, extension of the joint tests herein to multilocus models will be somewhat more challenging, though certainly not impossible. Such tests would require putting multiple variables on the dependent side of the equation. This might best be done through the multinomial regression approach by extending it to allow for modeling of multilocus genotype contingency tables.

However, using this type of joint test for multilocus models may require more biological assumptions (*e.g.*, Hardy-Weinberg equilibrium; random segregation) or the use of haplotypes.

Similar to the Haley and Knott [9] method, the joint tests can be extended to test any locus within a marker interval in order to approximate interval mapping. First, the probabilities of genotypes can be calculated using the multi-point method of Jiang and Zeng [12]. Then these values can be used in place of marker genotypes. However, developing an exact interval mapping method (similar to the conventional interval mapping in line crosses) will require an additional effort based on the EM algorithm, which is beyond the scope of this paper. Yet, this presents an interesting topic for future study. There are several other future research directions for these joint tests. In the field of experimental crosses, when modeling a disease trait, it may be a useful alternative to threshold models [17]; however, studies have found threshold models to be less powerful for QTL detection than simpler linear models [18]. Future research comparing these approaches may be warranted.

7. EXAMPLE

For demonstration purpose, we took selected data from a project designed to locate the gene on the mouse chromosome linked with the quantitative trait, the percentage of lung fibrosis that was induced by bleomycin, a therapy for treatment of cancer. Significant positions have been detected on chromosome 17 marker *D17mit16* and chromosome 11 marker *D11mit272* [10].

For brevity, we took $N = 165$ mice that had been genotyped on *D11mit272* (see Tab. IV). Using tests R_1 and R_2 , the results for all 165 mice show that there are significant mean differences in fibrosis among the three genotypes of *D11mit272* [R_2 : $F_{(2,172)} = 7.58$, $P = 0.00070$, $\eta^2 = 0.081$]. The additive component of this genetic effect was statistically significant [R_1 : $F_{(1,172)} = 14.09$, $P = 0.00024$, $\eta^2 = 0.075$]. The difference in these two models indicates that the non-additive effect was not statistically significant [$F_{(1,172)} = 1.08$, $P > 0.05$]. The multinomial regression (R_6) results in a likelihood ratio chi-square statistic [$\chi^2 = 14.46$, $df = 2$, $P = 0.00072$, Cox-Snell $R^2 = 0.079$].

Suppose that the researchers employed the ET2 design and genotyped the 88 phenotypically extreme mice. These mice included the top 25% (*i.e.*, mice with fibrosis $\geq 2.50\%$ of the lung $n_U = 44$) and $n_L = 44$ mice with zero fibrosis. The results of the OLS regression-based tests, show the significant mean differences in fibrosis among the three genotypes of *D11mit272* [R_2 : $F_{(2,85)} = 9.46$, $P = 0.00020$, $\eta^2 = 0.182$]. The additive component of this

Table IV. Results of the marginal-based tests performed on ET2 data.

<i>D17mit272</i>				
	B6-B6 (G=0)	B6-C3 (G = 1)	C3-C3 (G = 2)	Mean of G
Full sample	$n_0 = 43$	$n_1 = 89$	$n_2 = 43$	(SD)
Mean fibrosis	4.28	1.94	0.90	1.00
(SD)	(5.90)	(4.02)	(1.68)	(0.70)
ET2 sample	$n_0 = 22$	$n_1 = 44$	$n_2 = 22$	
Upper 25% ($n_U = 44$)	17	21	6	0.75
fibrosis > 2%				(SD = 0.69)
Lower 25% ($n_L = 44$)	5	23	16	1.25
fibrosis = 0%				(SD = 0.65)
Mean fibrosis	7.75	3.57	1.21	
(SD)	(6.58)	(5.23)	(2.22)	

genetic effect was statistically significant [$R_1: F_{(1,85)} = 18.21, P = 0.00005, \eta^2 = 0.175$]. The difference in these two models indicates that the non-additive effect was of marginal significance [$F_{(1,208)} = 0.71, P > 0.05$]. The multinomial regression (R_6) results in a likelihood ratio chi-square statistic [$\chi^2 = 17.92, df = 2, P = 0.00013, \text{Cox-Snell } R^2 = 0.184$]. By comparing the η^2 (or R^2) values, which provide an effect size that is independent of sample size, one can see that selecting the extreme ends of the phenotypic distribution can increase the power per genotyped organism.

For the marginal based tests, the results are shown descriptively in Table IV. The results of the marginal tests were statistically significant. The independent samples t -test comparing the mean genotype across the two extremes of the phenotypic distribution value was [$M_{10}: t_{(107)} = 3.51, P = 0.00073, \eta^2 = 0.125$]. The 2×3 contingency table analysis comparing the proportions of each genotype across the two extremes of the phenotypic distribution resulted in $\chi^2 = 11.18, df = 2, P = 0.00373, \phi^2 = 0.127$. For the OLS based joint test, J_{13} , the results were statistically significant [$F_{(1,85)} = 18.48, P = 0.00004$]. The multinomial regression joint test (J_{15}) results in a likelihood ratio chi-square statistic [$\chi^2 = 16.67, df = 2, P = 0.00024$]. Since distribution-based P -values are often questioned we also computed P -values *via* resampling. In this situation with $N = 88$ the number of permutations was extremely large; therefore we randomly permuted the data 20 000 times and took the test statistics percentile rank in the distribution of permuted test statistics as a permutation-based P -value [24]. The permutation-based P -values were $P = 0.00002$ for the OLS-based J_{13} and $P = 0.00006$ for the multinomial regression-based J_{15} .

ACKNOWLEDGEMENTS

This research was supported in part by National Institutes of Health grants P30DK56336, R01RR017009, R01ES09912, R01DK056366, and R01CA64193.

REFERENCES

- [1] Alf E.F., Abrahams N.M., The use of extreme groups in assessing relationships, *Psychometrika* 40 (1975) 563–572.
- [2] Allison D.B., Transmission-disequilibrium tests for quantitative traits, *Am. J. Hum. Genet.* 60 (1997) 676–690.
- [3] Allison D.B., Heo M., Schork N.J., Wong S.L., Elston R.C., Extreme selection strategies in gene mapping studies of oligogenic quantitative traits do not always increase power, *Hum. Hered.* 48 (1998) 97–107.
- [4] Barón A.E., Misclassification among methods used for multiple group discrimination – The effects of distributional properties, *Stat. Med.* 10 (1991) 757–766.
- [5] Dupuis J., Siegmund D., Statistical methods for mapping quantitative trait loci from a dense set of markers, *Genetics* 151 (1999) 373–386.
- [6] Fan X., Wang L., Comparing linear discriminant function with logistic regression for the two-group classification problem, *J. Exper. Educ.* 67 (1999) 265–286.
- [7] Feldt L.S., The use of extreme groups to test for the presence of a relationship, *Psychometrika* 26 (1961) 307–316.
- [8] Forrest W.F., Feingold E., Composite statistics for QTL mapping with moderately discordant sibling pairs, *Am. J. Hum. Genet.* 66 (2000) 1642–1660.
- [9] Haley C.S., Knott S.A., A simple method for mapping quantitative trait loci in line crosses using flanking markers, *Heredity* 69 (1992) 315–324.
- [10] Haston C.K., Wang M., Dejournett R.E., Zhou X., Ni D., Gu X., King T.M., Weil M.M., Newman R.A., Amos C.I., Travis E.L., Bleomycin hydrolase and a genetic locus within the MHC affect risk for pulmonary fibrosis in mice, *Hum. Mol. Genet.* 11 (2002) 1855–1863.
- [11] Henshall J.M., Goddard M.E., Multiple-trait mapping of quantitative trait loci after selective genotyping using logistic regression, *Genetics* 151 (1999) 885–894.
- [12] Jiang C., Zeng Z.B., Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines, *Genetica* 101 (1997) 47–58.
- [13] Lander E.S., Botstein D., Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps, *Genetics* 121 (1989) 185–199.
- [14] Liu F., Wu X.L., Chen S.Y., Segregation distortion of molecular markers in recombinant inbred populations in soybean (*G. max*). *Acta Genetica Sinica* 27 (2000) 883–887.
- [15] Piegorsch W.W., Weinberg C.R., Taylor J.A., Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies, *Stat. Med.* 13 (1994) 153–162.

- [16] Raju N.S., Brand P.A., Determining the significance of correlations corrected for unreliability and range restriction, *App. Psych. Meas.* 27 (2003) 52–71.
- [17] Rao S., Xia L., Strategies for genetic mapping of categorical traits, *Genetica* 109 (2000) 183–197.
- [18] Rebai A., Goffinet B., Mangin B., Comparing power of different methods for QTL detection, *Biometrics* 51 (1995) 87–99.
- [19] Sandler L., Novitski E., Meiotic drive as an evolutionary force, *Am. Nat.* 91 (1957) 105–110.
- [20] Sen S., Churchill G.A., A statistical framework for quantitative trait mapping, *Genetics* 159 (2001) 371–387.
- [21] Sham P.C., Purcell S., Cherny S.S., Abecasis G.R., Powerful regression-based quantitative-trait linkage analysis of general pedigrees, *Am. J. Hum. Genet.* 71 (2002) 238–253.
- [22] Spielman R.S., McGinnis R.E., Ewens W.J., Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM), *Am. J. Hum. Genet.* 52 (1993) 506–516.
- [23] Taylor D.R., Ingvarsson P.K., Common features of segregation distortion in plants and animals, *Genetica* 117 (2003) 27–35.
- [24] Westfall P.H., Young S.S., *Resampling-based multiple testing: Examples and methods for P -value adjustment*, Wiley & Sons, New York, 1993.
- [25] Xu S., Vogl C., Maximum likelihood analysis of quantitative trait loci under selective genotyping, *Heredity* 84 (2000) 525–537.
- [26] Xu Y., Zhu L., Xiao J., Huang N., McCouch S.R., Chromosomal regions associated with segregation distortion of molecular markers in F₂, backcross, double haploid, and recombinant inbred populations in rice (*Oryza sativa* L.), *Mol. Gen. Genet.* 253 (1997) 535–545.